OECD Series on Testing and Assessment

Revised Guidance Document 150 on Standardised Test Guidelines for Evaluating Chemicals for Endocrine Disruption



OECD

OECD Series on Testing and Assessment

Revised Guidance Document 150 on Standardised Test Guidelines for Evaluating Chemicals for Endocrine Disruption

OECD
BETTER POLICIES FOR BETTER LIVES

This work is published under the responsibility of the Secretary-General of the OECD. The opinions expressed and arguments employed herein do not necessarily reflect the official views of OECD member countries.

This document, as well as any data and any map included herein, are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

Please cite this publication as:
OECD (2018), *Revised Guidance Document 150 on Standardised Test Guidelines for Evaluating Chemicals for Endocrine Disruption*, OECD Series on Testing and Assessment, OECD Publishing, Paris.
https://doi.org/10.1787/9789264304741-en

ISBN 978-92-64-30474-1 (PDF)

Series: OECD Series on Testing and Assessment
ISSN 2077-7876 (online)

Photo credits: Cover © Sebastian Kaulitzk/Shutterstock.com; © Looker_Studio/Shutterstock.com.

Corrigenda to OECD publications may be found on line at: *www.oecd.org/about/publishing/corrigenda.htm*.
© OECD 2018

You can copy, download or print OECD content for your own use, and you can include excerpts from OECD publications, databases and multimedia products in your own documents, presentations, blogs, websites and teaching materials, provided that suitable acknowledgment of the source and copyright owner(s) is given. All requests for public or commercial use and translation rights should be submitted to *rights@oecd.org*. Requests for permission to photocopy portions of this material for public or commercial use shall be addressed directly to the Copyright Clearance Center (CCC) at *info@copyright.com* or the Centre francais d'exploitation du droit de copie (CFC) at *contact@cfcopies.com*.

Foreword

The Environmental Health and Safety (EHS) Programme of the OECD has been working to help countries design environmental policies and address chemical safety issues since 1971, with the purpose of sharing information among member countries in order to act jointly to reduce risks. One of the focuses of the programme is to enable the exchange of data that can be used in the safety evaluation of chemicals, including nanomaterials, pesticides and biocides.

To better address issues relevant to large numbers of chemicals, the OECD began to develop harmonised tools that countries could use to evaluate the safety of chemicals. Agreement on experimental approaches for measuring toxicity and a rigorous environment in which experiments were conducted led to the Mutual Acceptance of Data (MAD) system among OECD countries, a crucial step towards sharing data and reducing the burden of chemical safety evaluations by industry and governments.

One of the concerns regarding chemical safety is the potential for environmental chemicals to act as endocrine disruptors, mimicking the effects of naturally occurring hormones and disturbing the body's endocrine system. In this way, endocrine disrupters can alter the synthesis, release, action or elimination of natural hormones, and result in impaired development, reproduction, neurological function and immune system responses. Because the endocrine systems of vertebrates are evolutionarily conserved, environment endocrine disrupting chemicals can have similar effects in humans and wildlife.

In the 1990s, the public became aware of issues regarding endocrine disrupting chemicals following reports of decreased sperm quality and increasing trends in endocrine-related cancers that linked these health concerns to human chemical exposures. In 1996, an Advisory Group on Endocrine Disrupters Testing and Assessment (EDTA) was established at the OECD to develop new and update existing test guidelines in order to identify endocrine disrupting chemicals. The EDTA is a multi-disciplinary group comprised of representatives from member countries. It includes research and regulatory scientists, as well as representatives from the Business and Industry Advisory Committee (BIAC), environmental non-governmental organisations, and the International Council on Animal Protection (ICAPO) in OECD Programmes provide input. Collectively, the EDTA oversees the development of harmonised test guidelines for measuring endocrine disrupting effects of chemicals and guidance documents (including this one) for interpreting data resulting from the test guidelines.

The first edition of the OECD Guidance Document No. 150, was published in 2012 and was intended to provide guidance for interpreting the outcome of individual tests and compiling evidence on whether or not a substance may be an endocrine disrupter. The guidance document also provides a general description of each standardised test guideline, and tabular presentations of the endpoints measured in each test and the endocrine pathway affected. The document also describes a Conceptual Framework for Testing and Assessment of Endocrine Disrupters that helps organise available test methods at different

levels of biological organisation to determine additional testing needs or conclude about the potential endocrine disrupter action of a chemical. This document was the first comprehensive, international guide on the identification of endocrine disrupting chemicals and included step-by-step guidance for analysing results from standardised tests, weighing evidence for an endocrine mode of action, and identifying adverse effects in whole organisms. This guidance document is a reference for regulators and industry scientists and is a snapshot of the current test guidelines used for evaluating chemicals as possible endocrine disrupters. The guidance is not intended to be an exhaustive list of all methods available for evaluating chemicals, nor is it intended to be a rigid or prescriptive strategy for screening and testing chemicals.

At the time of the publication of the first edition, the expectation was that Guidance Document (GD) No. 150 would be updated to reflect new standardised test methods and new scientific understanding of how endocrine disrupters produce their effects. This revised edition includes new and updated test guidelines that have been validated, or are currently in the validation process. Updates to the test guidelines concern all levels in the OECD Conceptual Framework for Testing and Assessment of Endocrine Disrupting Chemicals, and the Conceptual Framework itself has been updated. The revised GD includes various cross-cutting issues and a summary of some experiences gained from using the guidance in the first edition. Finally, this revised edition is nearly double the length of the first, so electronic navigation around the document has been introduced to assist users. The revised edition was approved by Working Group of National Coordinators of the Test Guidelines Programme at its meeting in April 2018, and it is published under the responsibility of the Joint Meeting of the Chemicals Committee and the Working Party on Chemicals, Pesticides, and Biocides.

All OECD guidance documents, scoping documents and test guidelines mentioned herein are available individually free of charge on the OECD website for testing and assessment and OECD Guidelines for the Testing of Chemicals.

Acknowledgements

The following people contributed to writing this Guidance Document in 2012: Sneha Satya (Australia); Birgit Mertens (Belgium); Mike Wade (Canada); Ulla Hass, Henrik Holbech, Magnus Løfstedt, Anne Marie Vinggaard, Henrik Tyle, Pia Juul Nielsen, Marie Louise Holmer and Sofie Christiansen (Denmark); Jukka Ahtiainen and Petteri Talasniemi (Finland); Jean-Marc Porcher, Rija Samsera and Claire Beausoleil (France); Anja Kehrer, Bernd Schaefer, Frauke Stock, Tanja Juffernholz, Marike Kolossa-Gehring, Karen-ildico Hirsch-Ernst, Vera Ritz and Susanne Walter-Rohde (Germany); Zoltan Marcsek (Hungary); Albertio Mantovani and Cinzia La Rocca (Italy); Taisen Iguchi, Shigeki Miyachi, Kunihiko Yamazaki, Chisumi Eto, Akihiro Hirose, Atsushi Ono, Jun Kanno and Norihisa Tatarazako (Japan); Jong-Ryool Kim and Jisung Ryu (Korea); Betty Hakkert, ZhiChao Dang, Theo Traas, Theo Vermeire, Wim Mennes, Aldert Piersma and Emiel Rorije (Netherlands); Boguslaw Baranski (Poland); Yvonne Adersson, Agneta Ohlsson and Ing-Marie Olsson (Sweden); José Maria Navas (Spain); Walter Lichtensteiger (Switzerland); Philippa Edwards, Mike Roberts, Ovnair Sepai, Susy Brescia, Ian Indans, Christina Lye, Daniel Merckel and David Lovell (United Kingdom); Vicki Dellarco, Leslie Touart and Gary Timm (United States); Aude Kienzler, Peter Korytar, Sharon Munn, François Busquet, Elise Grignard, Sander van der Linden and Miriam Jacobs (European Union); Yasuyoshi Okuno, Remi Bars, Nina Hallmark, Masatoshi Kumamoto, Gernot Klotz, Rick Becker and James Wheeler (Business and Industry Advisory Committee [BIAC]); Gwynne Lyons (environmental NGO); Kate Willett and Kristy Sullivan (International Council on Animal Protection in OECD Programmes [ICAPO]); Jenny Odum and Peter Matthiessen (consultants).

The following people formed the Steering Committee that guided and contributed to the revision in 2017: Sofie Christiansen, Henrik Holbech, Henrik Tyle and Marie Louise Holmer (Denmark); Elise Grignard, Kimmo Louekari, Sharon Munn, Juan Riego Sintes and Gabriele Schoening (European Union); Alberto Mantovani (Italy); Akihiko Hirose and Taisen Iguchi (Japan); HeeSeok Lee (Korea); Zhichao Dang and Betty Hakkert (Netherlands); Anne-Laure Demierre and Petra Kunz (Switzerland); Miriam Jacobs (United Kingdom); Warren Casey, Mark Miller, Seema Schappelle and Greg Akerman (United States); Richard Green and James Wheeler (BIAC); Catherine Willett (ICAPO); Jenny Odum and Peter Matthiessen (consultants).

Table of contents

Acronyms and abbreviations **13**

Executive summary **17**

A. Introduction **19**

A.1. Background 19
A.2. The OECD Conceptual Framework for Testing and Assessment of Endocrine Disrupting Chemicals 19
A.3. Definitions and terms used 22
A.4. Objectives 24
A.5. General approach 24
A.6. Scope and limitations 25
Notes 36
References 36

B. General guidance on endocrine assessment, assays and endpoints **41**

B.1. Considerations on the assays addressed 42
B.2. Endpoints in the various assays of the Conceptual Framework 54
B.3. Cross-species extrapolations 72
B.4. Considering potential for multiple modes of endocrine action 73
B.5. Use of weight of evidence and adverse outcome approaches 74
B.6. Regulatory experience of endocrine assessment 77
Notes 81
References 81

C. Specific guidance for the test guidelines addressed **93**

Introduction to specific guidelines 95
OECD *in vitro* screens (Conceptual Framework Level 2) 99
C.1.1. Performance-based Test Guideline for Human Recombinant Estrogen Receptor (hrER) *In Vitro* Assays to Detect Chemicals with ER Binding Affinity (OECD TG 493) 100
C.1.2. Performance-Based Test Guideline for Stably Transfected Transactivation *In Vitro* Assays to Detect Estrogen Receptor Agonists and Antagonists (ER STTA) (OECD TG 455) 114
C.1.3. Stably Transfected Human Androgen Receptor Transcriptional Activation Assay for Detection of Androgenic Agonist and Antagonist Activity of Chemicals (AR STTA) (OECD TG 458) 130
C.1.4. H295R Steroidogenesis Assay (OECD TG 456) 144
Non-OECD *in vitro* screens (Conceptual Framework Level 2) 159
C.1.5. Androgen Receptor Binding Assay (US EPA OPPTS 890.1150) 160
C.1.6. Aromatase Assay (US EPA OPPTS 890.1200) 173

OECD non-mammalian screens and tests (Conceptual Framework Levels 3-5) 187
C.2.1. Fish Short-Term Reproduction Assay (FSTRA) (OECD TG 229) 189
C.2.2. 21-Day Fish Assay (OECD TG 230) 201
C.2.3. Amphibian Metamorphosis Assay (AMA) (OECD TG 231) 211
C.2.4 *Potamopyrgus antipodarum* Reproduction Test (OECD TG 242) 221
C.2.5 *Lymnaea stagnalis* Reproduction Test (OECD TG 243) 231
C.2.6 Chironomid Toxicity Test Using Spiked Sediment (OECD TG 218) or Spiked Water (OECD TG 219) 241
C.2.7. *Daphnia magna* Reproduction Test (OECD TG 211) 253
C.2.8. Fish, Early-Life Stage (FELS) Toxicity Test (OECD TG 210) 265
C.2.9. Fish Sexual Development Test (FSDT) (OECD TG 234) 275
C.2.10. Larval Amphibian Growth and Development Assay (LAGDA) (OECD TG 241) 289
C.2.11. Avian Reproduction Test (OECD TG 206) 301
C.2.12. Sediment-Water Chironomid Life-Cycle Toxicity Test (OECD TG 233) 311
C.2.13. Medaka Extended One-Generation Reproduction Test (MEOGRT) (OECD TG 240) 321
Non-OECD non-mammalian screens and tests (Conceptual Framework Levels 3-5) 333
C.2.14. Short-Term Juvenile Hormone Activity Screening Assay using *Daphnia magna* (SJHASA) (draft OECD TG) 335
C.2.15. Androgenised Female Stickleback Screen (AFSS) (GD 148) (variant of OECD TG 230) 345
C.2.16. EASZY Assay: Detection of Substances Acting through Estrogen Receptors using Transge nic cyp19a1b GFP Zebrafish Embryos (draft OECD TG) 355
C.2.17. JMASA: Juvenile Medaka Anti-Androgen Screening Assay (draft OECD GD) 367
C.2.18. *Xenopus* Embryonic Thyroid Signalling Assay (XETA) (draft OECD TG) 377
C.2.19. Harpacticoid Copepod Development and Reproduction Test with *Amphiascus* (OECD GD 201) 387
C.2.20. *Daphnia* Multigeneration Test for Assessment of Endocrine-Active Chemicals (DMGT) (proposed OECD TG) 397
C.2.21. Fish Life Cycle Toxicity Test (FLCTT) (US EPA OPPTS 850.1500) 407
C.2.22. Zebrafish Extended One-Generation Reproduction Test (ZEOGRT) (draft OECD TG 240) 415
C.2.23. Avian Two-Generation Toxicity Test in the Japanese Quail (ATGT) (US EPA OCSPP 890.2100/740-C-15-003) 427
C.2.24. RADAR: Rapid Androgen Disruption Adverse Outcome Reporter Assay (draft OECD TG) 437
OECD mammalian screens and tests (Conceptual Framework Levels 3-5) 447
C.3.1. Uterotrophic Bioassay in Rodents (UT assay) (OECD TG 440) (including OECD GD 71 on the procedure to test for anti-estrogenicity) 449
C.3.2. Hershberger Bioassay in Rats (H assay) (OECD TG 441) (including OECD GD 115 on the Weanling Hershberger Bioassay) 463
C.3.3. Repeated Dose 28-Day Oral Toxicity Study in Rodents (OECD TG 407) 477
C.3.4. Repeated Dose 90-Day Oral Toxicity Study in Rodents (OECD TG 408) 491
C.3.5. Combined Chronic Toxicity/Carcinogenicity Studies (OECD TG 451-3) 503
C.3.6. Reproduction/Developmental Toxicity Screening Test (OECD TG 421) and Combined Repeated Dose Toxicity Study with the Reproduction/Developmental Toxicity Screening Test (OECD TG 422) 515
C.3.7. Prenatal Developmental Toxicity Study (OECD TG 414) 531
C.3.8. Developmental Neurotoxicity Study (OECD TG 426) 545
C.3.9. Repeated Dose Dermal Toxicity: 21/28-Day Study (OECD TG 410) 557
C.3.10. Subchronic Dermal Toxicity: 90-Day Study (OECD TG 411) 569

C.3.11. 28-Day (Subacute) Inhalation Toxicity Study (OECD TG 412) 581
C.3.12. Subchronic Inhalation Toxicity: 90-Day Study (OECD TG 413) 593
C.3.13. Repeated Dose 90-Day Oral Toxicity Study in Non-Rodents (OECD TG 409) 605
C.3.14. Two-Generation Reproduction Toxicity Study (OECD TG 416) 617
C.3.15. Extended One-Generation Reproductive Toxicity Study (EOGRTS) (OECD TG 443) 631
Non-OECD mammalian screens and tests (Conceptual Framework Levels 3-5) 649
C.3.16. Pubertal Development and Thyroid Function Assay in Peripubertal Male Rats (male PP assay) (US EPA OPPTS 890.1500) 650
C.3.17. Pubertal Development and Thyroid Function Assay in Peripubertal Female Rats (female PP assay) (US EPA OPPTS 890.1450) 666

Glossary **683**

Tables

Table A.1. Screens and tests for which guidance is provided in this document 31
Table A.2. List of assays for which guidance has been developed in GD 150, showing their responsiveness to various endocrine modalities 35
Table B.1. Endpoints relevant for endocrine disruption modalities in test guidelines and other EAS-sensitive assays (in the Conceptual Framework) for which guidance with interpretation of data has been developed 57
Table B.2. A selection of evidence approaches for assessment of endocrine effects (in order of publication date) 76
Table C.1.1. Performance-Based Test Guideline for Human Recombinant Estrogen Receptor (hrER) *In Vitro* Assays to Detect Chemicals with ER Binding Affinity (OECD TG 493): Guidance for scenarios of combinations of results with existing data 107
Table C.1.2. Performance-Based Test Guideline for Stably Transfected Transactivation *In Vitro* Assays to Detect Estrogen Receptor Agonists and Antagonists (ER STTA) (OECD TG 455): Guidance for scenarios of combinations of results with existing data 123
Table C.1.3. Stably Transfected Human Androgen Receptor Transcriptional Activation Assay for Detection of Androgenic Agonist and Antagonist Activity of Chemicals (AR STTA) (OECD TG 458): Guidance for scenarios of combinations of results with existing data 137
Table C.1.4. H295R Steroidogenesis Assay (OECD TG 456): Guidance for scenarios of combinations of results with existing data 151
Table C.1.5. Androgen Receptor Binding Assay (US EPA OPPTS 890.1150): Guidance for scenarios of combinations of results with existing data 166
Table C.1.6. Aromatase Assay (US EPA OPPTS 890.1200): Guidance for scenarios of combinations of results with existing data 180
Table C.2.1. Fish Short-Term Reproduction Assay (FSTRA) (OECD TG 229): Guidance for scenarios of combinations of results with existing data 194
Table C.2.2. 21-Day Fish Assay (OECD TG 230): Guidance for scenarios of combinations of results with existing data 205
Table C.2.3. Amphibian Metamorphosis Assay (AMA) (OECD TG 231): Guidance for scenarios of combinations of results with existing data 216
Table C.2.4. *Potamopyrgus antipodarum* Reproduction Test (OECD TG 242): Guidance for scenarios of combinations of results with existing data 225
Table C.2.5. *Lymnaea stagnalis* Reproduction Test (OECD TG 243): Guidance for scenarios of combinations of results with existing data 235

Table C.2.6. Chironomid Toxicity Test Using Spiked Sediment (OECD TG 218) or Spiked Water (OECD TG 219): Guidance for scenarios of combinations of results with existing data 246
Table C.2.7. *Daphnia magna* Reproduction Test (OECD TG 211): Guidance for scenarios of combinations of results with existing data 258
Table C.2.8. Fish, Early-Life Stage (FELS) Toxicity Test (OECD TG 210): Guidance for scenarios of combinations of results with existing data 270
Table C.2.9. Fish Sexual Development Test (FSDT) (OECD TG 234): Guidance for scenarios of combinations of results with existing data 281
Table C.2.10. Larval Amphibian Growth and Development Assay (LAGDA) (OECD TG 241): Guidance for scenarios of combinations of results with existing data 294
Table C.2.11. Avian Reproduction Test (OECD TG 206): Guidance for scenarios of combinations of results with existing data 305
Table C.2.12. Sediment-Water Chironomid Life-Cycle Toxicity Test (OECD TG 233): Guidance for scenarios of combinations of results with existing data 315
Table C.2.13. Medaka Extended One Generation Reproduction Test (MEOGRT) (OECD TG 240): Guidance for scenarios of combinations of results with existing data 326
Table C.2.14. Short-Term Juvenile Hormone Activity Screening Assay using *Daphnia magna* (SJHASA) (draft OECD TG): Guidance for scenarios of combinations of results with existing data 339
Table C.2.15. Androgenised Female Stickleback Screen (AFSS) (OECD GD 148) (variant of OECD TG 230): Guidance for scenarios of combinations of results with existing data 349
Table C.2.16. EASZY Assay: Detection of Substances Acting through Estrogen Receptors using Transgenic cyp19a1b GFP Zebrafish Embryos (draft OECD TG): Guidance for scenarios of combinations of results with existing data 360
Table C.2.17. JMASA: Juvenile Medaka Anti-Androgen Screening Assay (draft OECD GD): Guidance for scenarios of combinations of results with existing data 371
Table C.2.18. *Xenopus* Embryonic Thyroid Signalling Assay (XETA) (draft OECD TG): Guidance for scenarios of combinations of results with existing data 382
Table C.2.19. Harpacticoid Copepod Development and Reproduction Test with *Amphiascus* (OECD GD 201): Guidance for scenarios of combinations of results with existing data 391
Table C.2.20. *Daphnia* Multigeneration Test for Assessment of Endocrine-Active Chemicals (DMGT) (proposed OECD TG): Guidance for scenarios of combinations of results with existing data 401
Table C.2.21. Fish Life Cycle Toxicity Test (FLCTT) (US EPA OPPTS 850.1500): Guidance for scenarios of combinations of results with existing data 411
Table C.2.22. Zebrafish Extended One-Generation Reproduction Test (ZEOGRT) (draft OECD TG): Guidance for scenarios of combinations of results with existing data 420
Table C.2.23. Avian Two-Generation Toxicity Test in the Japanese Quail (ATGT) (US EPA OCSPP 890.2100/740-C-15-003): Guidance for scenarios of combinations of results with existing data 431
Table C.2.24. RADAR: Rapid Androgen Disruption Adverse Outcome Reporter Assay (draft OECD TG): Guidance for scenarios of combinations of results with existing data 441

Table C.3.1. Uterotrophic Bioassay in Rodents (UT assay) (OECD TG 440) (including OECD GD 71 on the Procedure to Test for Anti-estrogenicity): Guidance for scenarios of combinations of results with existing data 457
Table C.3.2. Hershberger Bioassay (H assay) (OECD TG 441) (including OECD GD 115 on the Weanling Hershberger Bioassay): Guidance for scenarios of combinations of results with existing data 471
Table C.3.3. Repeated Dose 28-Day Oral Toxicity Study in Rodents (OECD TG 407): Guidance for scenarios of combinations of results with existing data 484
Table C.3.4. Repeated Dose 90-Day Oral Toxicity Study in Rodents (OECD TG 408): Guidance for scenarios of combinations of results with existing data 497
Table C.3.5. Combined Chronic Toxicity/Carcinogenicity Studies (OECD TG 451-3): Guidance for scenarios of combinations of results with existing data 509
Table C.3.6. Reproduction/Developmental Toxicity Screening Test (OECD TG 421) and Combined Repeated Dose Toxicity Study with the Reproduction/ Developmental Toxicity Screening Test (OECD TG 422): Guidance for scenarios of combinations of results with existing data 523
Table C.3.7. Prenatal Developmental Toxicity Study (OECD TG 414): Guidance for scenarios of combinations of results with existing data 538
Table C.3.8. Developmental Neurotoxicity Study (OECD TG 426): Guidance for scenarios of combinations of results with existing data 551
Table C.3.9. Repeated Dose Dermal Toxicity: 21/28-Day Study (OECD TG 410): Guidance for scenarios of combinations of results with existing data 563
Table C.3.10. Subchronic Dermal Toxicity: 90-Day Study (OECD TG 411): Guidance for scenarios of combinations of results with existing data 575
Table C.3.11. 28-Day (Subacute) Inhalation Toxicity Study (OECD TG 412): Guidance for scenarios of combinations of results with existing data 587
Table C.3.12. Subchronic Inhalation Toxicity: 90-Day Study (OECD TG 413): Guidance for scenarios of combinations of results with existing data 599
Table C.3.13. Repeated Dose 90-Day Oral Toxicity Study in Non-Rodents (OECD TG 409): Guidance for scenarios of combinations of results with existing data 611
Table C.3.14. Two-Generation Reproduction Toxicity Study (OECD TG 416): Guidance for scenarios of combinations of results with existing data 624
Table C.3.15. Extended One-Generation Reproductive Toxicity Study (EOGRTS) (OECD TG 443): Guidance for scenarios of combinations of results with existing data 640
Table C.3.16. Pubertal Development and Thyroid Function Assay in Peripubertal Male Rats (male PP assay) (OPPTS 890.1500): Guidance for scenarios of combinations of results with existing data 658
Table C.3.17. Pubertal Development and Thyroid Function Assay in Peripubertal Female Rats (female PP assay) (US EPA OPPTS 890.1450): Guidance for scenarios of combinations of results with existing data 674

Acronyms and abbreviations

A	Androgen
ADME	Absorption, distribution, metabolism and excretion
AFSS	Androgenised Female Stickleback Screen (OECD GD 140)
AGD	Anogenital distance
AMA	Amphibian Metamorphosis Assay (OECD TG 231)
AOP	Adverse outcome pathway
AR	Androgen receptor
AR STTA	Stably Transfected Human Andogen Receptor Transactivation Assay for Detection of Androgenic (ant)agonist-Activity of Chemicals
ATGT	Avian two-generation test
BCF	Bioconcentration factor
CF	Conceptual Framework
DNT	Developmental neurotoxicity
E	Estrogen
EAS	Endocrine active substance
EASZY	Detection of Substances Acting through Estrogen Receptors using Transgenic cyp19a1b GFP Zebrafish Embryos (draft OECD TG)
E,A,T,S	Estrogen/androgen/thyroid/steroidogenesis
ECx	x% effect concentration
ED	Endocrine disrupter
EDSP	Endocrine Disruptor Screening Program (US EPA)
EDTA AG	Endocrine Disrupters Testing and Assessment Advisory Group
EOGRTS	Extended One-Generation Reproductive Toxicity Study (OECD TG 443)
ER	Estrogen receptor
ER STTA	Stably Transfected Human Estrogen Receptor-alpha Transactivation Assay for Detection of Estrogenic Agonist-Activity of Chemicals (OECD TG 455)
ERTA	Estrogen Receptor Transactivation Assay

FELS	Fish Early Life Stage Toxicity Test
FLCTT	Fish Life Cycle Toxicity Test
FSDT	Fish Sexual Development Test (draft OECD TG 234)
FSH	Follicle stimulating hormone
FSTRA	Fish Short-Term Reproduction Assay (OECD TG 229)
GD	Guidance document
GFP	Green fluorescent protein
GIVIMP	Good *in vitro* method practices
GnRH	Gonadotropin releasing hormone
GR	Glucocorticoid receptor
H assay	Hershberger Bioassay
HPG axis	Hypothalamic/pituitary/gonadal axis
HPT axis	Hypothalamic/pituitary/thyroid axis
HTS	High throughput screening
IATA	Integrated approaches to testing and assessment
ICCVAM	Interagency Coordinating Committee on the Validation of Alternative Methods
ICE	Integrated chemical environment
JMASA	Juvenile Medaka Anti-androgen Screening Assay
LABC	Levator ani plus bulbocavernosus muscle complex
LAGDA	Larval Amphibian Growth and Development Assay
LH	Luteinising hormone
LOEC	Lowest-observed-effect-concentration
MEOGRT	Medaka Extended One-Generation Reproduction Test
MOA	Mode of action
NOAEL	No observed adverse effect level
NOEC	No-observed-effect-concentration
OECD	Organisation for Economic Co-operation and Development
PBTG	Performance-based test guideline
PEC/PNEC	Predicted environmental concentration/predicted no-effect concentration
PND	Postnatal day
POP	Persistant organic pollutant
PP assay	Peripubertal Assay (male or female)

PPAR	Peroxisome proliferator-activated receptor
PPS	Preputial separation
QSAR	Quantitative structure activity relationship
RADAR	Rapid Androgen Disruption Adverse Outcome Reporter Assay
REACH	Registration, evaluation, authorisation and restriction of chemicals
RXR	Retinoid X receptor
S	Steroidogenesis
STTA	Stably Transfected Transactivation Assay
T	Thyroid
T3	Tri-iodothyronine (thyroid hormone)
T4	Thyroxine (thyroid hormone)
TG	Test guideline
TPO	Thyroperoxidase
TR	Thyroid hormone receptor
TRH	Thyrotropin releasing hormone
TSH	Thyroid stimulating hormone
US EPA	United States Environmental Protection Agency
UT assay	Uterotrophic Bioassay
UVCB	Unknown or variable composition substance, complex reaction product or biological material
VO	Vaginal opening (or patency)
VTG	Vitellogenin
WHO	World Health Organization
WOE	Weight of evidence
XETA	*Xenopus* Embryonic Thyroid Signalling Assay
YAS	Yeast androgen screen
YES	Yeast estrogen screen
ZEOGRT	Zebrafish Extended One-Generation Reproduction Test

Executive summary

Guidance Document on Standardised Test Guidelines for Evaluating Chemicals for Endocrine Disruption (Guidance Document [GD] 150) was originally published in 2012 to provide guidance for evaluating chemicals in a regulatory context. Development of GD 150 was overseen by the Advisory Group for Endocrine Disrupters Testing and Assessment (EDTA) established in 1996. The group includes experts from OECD member countries, as well as representatives from animal welfare and environmental non-governmental organisations, and the chemical industry. The charge to the EDTA was to develop and update existing test guidelines to identify endocrine disrupting chemicals and provide guidance on interpreting the results from these test guidelines. Test method development followed the regulatory needs of OECD member countries for the screening and testing of endocrine disrupters, initially focusing on chemicals that interfered with estrogen, androgen and thyroid pathway signaling, and chemicals that altered steroidogenesis in mammals and vertebrate wildlife. Over the subsequent 20 years since the EDTA was established, the OECD has validated and published more than 35 test guidelines with endpoints specific for endocrine disruption and many more guidelines with "informative endpoints", such as reproductive organ weights or histopathology.

When originally published in 2012, GD 150 was the first comprehensive, international guide for identifying endocrine disrupting chemicals. The general approach taken in the document is primarily to provide guidance on how test results might be interpreted based on the outcome of standardised assays. Key questions addressed in the document concern likely mechanisms of endocrine action and any resulting apical effects that can be attributed to such action. The document is not prescriptive but provides suggestions for possible next steps in testing (if any) which might be appropriate for a regulatory authority to take, given the various data scenarios. Specific objectives included providing guidance for analysing results from individual tests, the endocrine endpoints measured in each test, the endocrine mode of action covered by each test, support for regulatory decisions on whether a substance may be an endocrine disrupter, and recommendations for a next testing step if a conclusion cannot be made. In addition, GD 150 describes the Conceptual Framework for endocrine testing and assessment. The Conceptual Framework is an approach to organising data into five levels of increasingly biologically complex information that help to evaluate the overall strength of the evidence that a chemical may be acting as an endocrine disrupter. Application of the Conceptual Framework and GD 150 are demonstrated in the "Guidance document on standardised test guidelines for evaluating chemicals for endocrine disruption: Case studies using example chemicals" (Series on Testing and Assessment, No. 181).

This guidance document is organised into three main sections. Section A provides background on endocrine testing in the OECD context, the Conceptual Framework, a description of definitions and terms used, objectives of the document, the assays and endocrine modalities covered, and a list of OECD test guidelines with endpoints specific for endocrine active substances and test guidelines with endpoints with that maybe informative but that are not specific to endocrine active substances. Section B provides

guidance on endocrine assessment, assays and endpoints. Section B also provides more specific detail for each assay and distinguishes non-test and test methods for evaluating endocrine disrupters by the levels of the Conceptual Framework, and provides a discussion of the use of weight of evidence approaches for integrating information from multiple assays and provides regulatory experience using the document for evaluating chemicals for potential endocrine activity. Section C provides a thorough description of each of the assays included in the Conceptual Framework, including background of the assay, under what circumstances the assay may be used, and provides a variety of example data scenarios and suggestions for a single next testing step if a conclusion cannot be reached with data illustrated in the scenario.

In the 2018 edition of Guidance Document 150, all sections have been updated to include new and revised OECD internationally harmonised test guidelines, assays validated at the national level and assays that are currently in the OECD validation process. In addition, the 2018 edition discusses the application of new approaches to toxicity testing, such as integrated approaches to testing and assessment, use of adverse outcome pathways for evaluating endocrine disruption, extrapolating assay results across mammalian and non-mammalian vertebrate species, and approaches for evaluating chemicals with multiple modes of action.

A. Introduction

A.1. Background

1. The OECD initiated a high-priority activity in 1998 to revise existing, and develop new, test guidelines (TGs) for the screening and testing of endocrine disrupting chemicals. Since then a number of potential assays have been developed into test guidelines and others are in development. The screens and tests are contained within the OECD Conceptual Framework for the Screening and Testing of Endocrine Disrupting Chemicals (CF) which was developed in 2002 by the Endocrine Disrupters Testing and Assessment Advisory Group (EDTA AG), modified and updated in 2012 and again in 2017. The 2017 revised version of the CF is shown in Section A.2. A workshop on "OECD Countries' Activities Regarding Testing, Assessment and Management of Endocrine Disrupters" was held in Copenhagen on 22-24 September 2009 (OECD, 2010b). One output from this workshop was a recommendation that a guidance document (GD) on the assessment of chemicals for endocrine disruption should be developed by the EDTA AG. This was supported by the EDTA AG at its meeting on 17-18 May 2010. The objectives and scope of the GD were defined such that the document would be a tool to support regulatory authorities by helping to interpret assay results and suggesting possible additional studies for reducing uncertainty. The guidance should not prejudge or constrain what regulatory actions may be taken by a member country and should not suggest a testing strategy. The guidance should also support, but not duplicate, other GDs (e.g. guidance on hazard assessment). It should be noted that the use of many of these tests for determination of toxicity due to endocrine disruption (hazard identification/characterisation) for mammals and non-mammals was still rather new, and therefore the guidance given was considered to be subject to changes based on new evidence. The guidance was intended to be a "living" document to be updated as the science in this area evolves, and the present publication represents the first such update. In particular, this update takes into account the many new validated assay/test methods developed since the 2012 version of the GD (see Table A.1 and discussion in Section B). This document also provides additional guidance on evaluation of each validated assay/test method (Section C).

A.2. The OECD Conceptual Framework for Testing and Assessment of Endocrine Disrupting Chemicals

2. The OECD Conceptual Framework lists the OECD TGs and standardised test methods available, under development or proposed, that can be used to evaluate chemicals for endocrine disruption. It is not an exhaustive list and will be updated as new assays are developed. Assays other than those described in the list may also be valuable for assessing chemicals for endocrine disruption and could be assigned to a level based on the level descriptors. The CF is intended to provide a guide to the tests available which can provide information on assessment of endocrine disruption, but is not intended to be a testing strategy. Furthermore, the CF, as revised in 2017, does not include evaluation of exposure as it is intended for hazard identification/characterisation (see definitions in Section A.3).

The OECD Conceptual Framework for Testing and Assessment of Endocrine Disrupting Chemicals, revised 2017

Mammalian and non-mammalian toxicology		
Level 1 Existing data and existing or new non-test information	– Physical and chemical properties, e.g. molecular weight reactivity, volatility, biodegradability – All available (eco)toxicological data from standardised or non-standardised tests – Read-across, chemical categories, quantitative structure activity relationships and other *in silico* predictions, and absorption, distribution, metabolism and excretion model predictions	
Level 2 *In vitro* assays providing data about selected endocrine mechanism(s)/ pathway(s) (mammalian and non-mammalian methods)	– Estrogen (OECD TG 493) or androgen receptor binding affinity (US EPA TG OPPTS 890.1150) – Estrogen receptor transactivation (OECD TG 455, ISO 19040-3), yeast estrogen screen (ISO 19040-1 & 2) – Androgen receptor transactivation (OECD TG 458) – Steroidogenesis *in vitro* (OECD TG 456) – Aromatase assay (US EPA TG OPPTS 890.1200) – Thyroid disruption assays (e.g. thyroperoxidase inhibition, transthyretin binding) – Retinoid receptor transactivation assays – Other hormone receptors assays as appropriate – High-throughput screens	
	Mammalian toxicology[3]	**Non-mammalian toxicology[3]**
Level 3 *In vivo* assays providing data about selected endocrine mechanism(s)/ pathway(s)[1]	– Uterotrophic Assay (OECD TG 440) – Hershberger assay (OECD TG 441)	– Amphibian metamorphosis assay (AMA) (OECD TG 231) – Fish short-term reproduction assay (FSTRA) (OECD TG 229)[2] – 21-day fish assay (OECD TG 230) – Androgenised female stickleback screen (AFSS) (OECD GD 148) – EASZY Assay. Detection of Substances Acting through Estrogen Receptors using Transgenic cyp19a1b GFP Zebrafish Embryos (draft OECD TG) – *Xenopus* embryonic thyroid signalling assay (XETA) (draft OECD TG) – Juvenile medaka anti-androgen screening assay (JMASA) (draft OECD GD) – Short-term juvenile hormone activity screening assay using *Daphnia magna* (draft OECD TG) – Rapid androgen disruption adverse outcome reporter (RADAR) assay (draft OECD TG)
Level 4 *In vivo* assays providing data on adverse effects on endocrine-relevant endpoints[2]	– Repeated dose 28-day study (OECD TG 407) – Repeated dose 90-day study (OECD TG 408) – Pubertal development and thyroid function assay in peripubertal male rats (PP male assay) (US EPA TG OPPTS 890.1500) – Pubertal development and thyroid function assay in peripubertal female rats (PP female assay) (US EPA TG OPPTS 890.1450) – Prenatal developmental toxicity study (OECD TG 414) – Combined chronic toxicity and carcinogenicity studies (OECD TG 451-453)	– Fish sexual development test (FSDT) (OECD TG 234) – Larval amphibian growth and development assay (LAGDA) (OECD TG 241) – Avian reproduction assay (OECD TG 206) – Fish early life stage (FELS) toxicity test (OECD TG 210) – New guidance document on harpacticoid copepod development and reproduction test with *Amphiascus* (OECD GD 201)[2] – *Potamopyrgus antipodarum* reproduction test (OECD TG 242)[4] – *Lymnaea stagnalis* reproduction test (OECD TG 243)[4]

The OECD Conceptual Framework for Testing and Assessment of Endocrine Disrupting Chemicals, revised 2017 (*continued*)

	Mammalian toxicology[3]	**Non-mammalian toxicology**[3]
Level 4 (*continued*)	– Reproduction/developmental toxicity screening test (OECD TG 421) – Combined repeated dose toxicity study with the reproduction/developmental toxicity screening test (OECD TG 422) – Developmental neurotoxicity study (OECD TG 426) – Repeated dose dermal toxicity: 21/28-day study (OECD TG 410) – Subchronic dermal toxicity: 90-day study (OECD TG 411) – 28-day (subacute) inhalation toxicity study (OECD TG 412) – Subchronic inhalation toxicity: 90-day study (OECD TG 413) – Repeated dose 90-day oral toxicity study in non-rodents (OECD TG 409)	– Chironomid toxicity test (OECD TG 218-219) [4] – *Daphnia magna* reproduction test (with male induction) (OECD TG 211)[4] – Earthworm reproduction test (OECD TG 222)[4] – Enchytraeid reproduction test (OECD TG 220)[4] – Sediment water *Lumbriculus* toxicity test using spiked sediment (OECD TG 225)[4] – Predatory mite reproduction test in soil (OECD TG 226)[4] – Collembolan reproduction test in soil (TG OECD 232)[4]
Level 5 *In vivo* assays providing more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism[2]	– Extended one-generation reproductive toxicity study (EOGRTS) (OECD TG 443)[5] – Two-generation reproduction toxicity study (OECD TG 416, most recent update)	– Fish life cycle toxicity test (FLCTT) (US EPA TG OPPTS 850.1500) – Medaka extended one-generation reproduction test (MEOGRT) (OECD TG 240) – Avian two-generation toxicity test in the Japanese quail (ATGT) (US EPA TG OCSPP 890.2100/740-C-15-003) – Sediment water chironomid life cycle toxicity test (OECD TG 233)[4] – *Daphnia* multigeneration test for assessment of EDCs (draft OECD TG)[4] – Zebrafish extended one-generation reproduction test (ZEOGRT) (draft OECD TG)

Notes:

1. Some assays may also provide some evidence of adverse effects. 2. Some endpoints can be sensitive to more than one mechanism and may be due to non-endocrine mechanisms. 3. Depending on the guideline/protocol used, the fact that a substance may interact with a hormone system in these assays does not necessarily mean that when the substance is used it will cause adverse effects in humans or ecological systems. 4. At present, these invertebrate assays solely involve apical endpoints which are able to respond to some endocrine active substances and some non-endocrine active substances. Those in Level 4 are generally partial life cycle tests, while those in Level 5 are full or multiple life cycle tests. 5. The EOGRTS (OECD TG 443) is preferable for detecting endocrine disruption because it provides an evaluation of a number of endocrine endpoints in the juvenile and adult F1, which are not included in the two-generation study (OECD TG 416) adopted in 2001.

Notes to the OECD Revised Conceptual Framework:

Entering at all levels and exiting at all levels is possible and depends on the nature of existing information and needs for testing and assessment. The assessment of each chemical should be made on a case-by-case basis, taking into account all available information. The framework should not be considered as all inclusive at the present time. It includes assays that are either available, or for which validation is under way. With respect to the latter, these are provisionally included, and a few assays (e.g. the avian two-generation test) have only been validated at national level. At Level 2 some assays are not (yet) proposed for validation but are included because they may provide information on important molecular interactions.

A.3. Definitions and terms used

3. In the context of this document, the following terms have been defined according to published and generally well-accepted definitions. The definitions from the Berlin Worshop Consensus Statement (Solecki et al., 2017), where several scientists in the endocrine disruption field agreed statements, are also shown and accepted in this document.

Term	Definition	Comments	Reference
Endocrine disrupter (ED)	An ED is an exogenous substance or mixture that alters function(s) of the endocrine system and consequently causes adverse health effects in an intact organism, its progeny or (sub)populations.	It is acknowledged that many other definitions exist (e.g.Weybridge Conference, 1996), but the WHO/IPCS (2002) definitions have been used as working definitions for this document because they cover both human health and non-mammalian populations. Accepted by Solecki et al. (2017) and within the European Union (EC, 2016).	WHO/IPCS (2002)
Potential endocrine disrupter	A potential ED is an exogenous substance or mixture that possesses properties that might be expected to lead to endocrine disruption in an intact organism, its progeny or (sub)populations.		WHO/IPCS (2002)
Endocrine active substance (EAS)	A substance having the inherent ability to interact or interfere with one or more components of the endocrine system resulting in a biological effect, but need not necessarily cause adverse effects.		EFSA (2013)
Adverse effect	A change in morphology, physiology, growth, reproduction, development or lifespan of an organism which results in impairment of functional capacity or impairment of capacity to compensate for additional stress or increased susceptibility to the harmful effects of other environmental influences.	Widely accepted as the definition of "adverse effect" to accompany the WHO/IPCS (2002) definition of an ED, and also by Solecki et al. (2017).	WHO/IPCS (2009)
Adverse outcome pathway (AOP)	An AOP is a conceptual construct that portrays existing knowledge concerning the linkage between a direct molecular initiating event and an adverse outcome at a biological level of organisation relevant to risk assessment.	AOPs can be very helpful in establishing the links between an endocrine mechanism and its potential apical effects.	Ankley et al. (2010)
Intact organism	The term "intact organism" is understood to mean that the effect would occur *in vivo*, either observable in a test animal system, epidemiologically or clinically. However, it does not necessarily mean that the adverse effect has to be demonstrated in an intact test animal, but may be shown in adequately validated alternative test systems predictive of adverse effects in humans and/or wildlife. The importance of mechanistic data derived from experimental systems (*in vitro* or *in vivo* in which the animals have been surgically or genetically altered as part of a focused experiment) was also recognised.	"Intact organism" to accompany the WHO/IPCS (2002) definition of an ED.	Solecki et al. (2017)
Hazard identification	The identification of the type and nature of adverse effects that an agent has an inherent capacity to cause in an organism, system or (sub)population. Hazard identification is the first stage in hazard assessment and the first of four steps in risk assessment.	GD 150 only covers assessment of hazard, not risk. Exposure is not considered. The term hazard identification/characterisation is used in relevant places and may encompass elements of both of these definitions.	IPCS/WHO (2004)

Term	Definition	Comments	Reference
Hazard characterisation	The qualitative and, wherever possible, quantitative description of the inherent property of an agent or situation having the potential to cause adverse effects. This should, where possible, include a dose-response assessment and its attendant uncertainties. Hazard characterisation is the second stage in the process of hazard assessment and the second of four steps in risk assessment.		IPCS/WHO (2004)
Mode of action (MOA)	A set of key events and processes starting with the interaction of an agent with a cell, through physiological and tissue or organ changes, potentially resulting in an adverse outcome.	"Mode" of action is contrasted with "mechanism", which implies a more detailed biochemical and molecular description of causality. These definitions are implicit in the IPCS Human Relevance Framework and Adverse Outcome Pathways (AOP)	Dellarco and Fenner-Crisp (2012) Boobis et al. (2006; 2008) Ankley et al. (2010) OECD (2016)
Unknown or variable composition substances, complex reaction products or biological materials (UVCBs)	These are substances where the number of constituents is relatively large, the composition is largely unknown, or the variability of composition is high or unpredictable.		Substance identity – UVCB substances ECHA workshop 2 Feb. 2012
Weight of evidence (WOE)	A process in which all of the evidence considered relevant for a hazard identification/characterisation is evaluated and weighted.	This concept is central to the evaluation of endocrine active substances and endocrine disrupters.	WHO/IPCS (2009)

4. The above definitions implicitly refer not only to the chemical in question, but also to its endocrine-active impurities. For multi-constituent substances, UVCBs and mixtures, the definitions refer to relevant constituents. Furthermore, when reference is made to a chemical in this context, it implicitly also covers its relevant environmental transformation products and its metabolites that are formed in exposed organisms.

5. The following "concepts" related to endocrine disruption were also agreed at the Berlin Workshop (Solecki et al., 2017) and are reproduced below as they are central to an understanding of endocrine disruption:

- Alterations of the function of the endocrine system may arise from interaction with hormone receptors; changes in circulating levels of the hormone; and from the impact of chemical(s) on hormone synthesis, transport, metabolism and other factors.
- Certain hormones interact with their receptors according to an equilibrium reaction. Accordingly, the concentrations of both free hormone and free receptor are important variables controlling hormone action, explaining why different cells and tissues at different times during development are differentially sensitive to the hormone. These factors also vary between species.
- Experimental work has led to a better understanding of the role of hormones in development and during the maintenance of physiological functions. Disruption of the programming role of hormones during prenatal and postnatal development can cause adverse effects that do not become evident until later in life.
- Interference with the role of many hormones during the maintenance of physiological functions in adult life can also lead to adverse effects.

A.4. Objectives

6. The objectives of this guidance document are to:

- Provide guidance on assays that might indicate the potential for endocrine disruption, endpoints within these assays and interpretation of their results.
- Support regulatory authorities' decisions on the hazard of specific chemicals and toxicologically relevant metabolites when they receive test results from a TG, draft TG or other standardised assay for the screening/testing of chemicals for endocrine disrupting properties. The context for these decisions will vary, depending on local legislation and practice, so the advice is worded in such a way as to permit flexible interpretation.
- Provide guidance on how to interpret the outcome of individual tests and how to strengthen the weight of the evidence on whether or not a substance may be an endocrine disrupter (ED). Testing strategies or guidance on interpretation from a suite of tests are not given.

7. Hazard assessment methods in this document are arranged in a two-step process, with the intention of minimising animal testing globally through application of the 3Rs (replace, reduce and refine the use of laboratory animals in testing):

- Use of a harmonised framework for assessing test results together with existing information on likely or known hazards should avoid unnecessary animal testing.
- Recommendation of a test method that may be performed if regulatory authorities need more evidence. The test method is defined precisely to facilitate the mutual acceptance of data and to avoid unnecessary duplication of testing. The recommended test method will utilise non-animal tests where possible, although a few alternative scenarios are considered depending on existing information.
- Because hormone receptors and pathways are highly conserved across the vertebrates, cross-species extrapolations should be considered as a way to reduce vertebrate testing.

A.5. General approach

8. The general approach taken by this GD is primarily to consider the possible results that might be obtained from each endocrine disruption-responsive assay,[1] and to provide guidance about how these results might be interpreted in the light of data that may or may not already be available from other *in vitro* or *in vivo* assays. This should include all available data such as publications in the peer-reviewed literature as well as TGs. In order to inform this interpretation, background data on the assays addressed, non-testing approaches and other considerations relevant to the assays are discussed. These include cross-species extrapolations, read-across and multiple modes of action (MOA). The nature, quantity and quality of the existing and new data in each of the scenarios for the endocrine disruption-responsive assays should be evaluated systematically in a weight of evidence (WOE) approach (WOE and examples are also discussed), and there is generally no single "right" answer. Use of other technologies (for example gene expression analysis or "omics" data) may help in understanding the link between endocrine-related mechanisms and apical effects in a WOE approach. This GD should therefore be used flexibly in the light of local regulatory needs. The key questions addressed concern likely mechanisms of endocrine

action and any resulting apical effects that can be attributed to such action. Given the widely agreed definition of endocrine-disrupting chemicals (WHO/IPCS, 2002), the advice suggests that a chemical is an ED if an adverse *in vivo* effect can be plausibly linked to an endocrine MOA.

9. This document provides advice on the next step in testing (if any) which might be appropriate for a regulatory authority to take, given the various data scenarios. It should be noted that it has only been possible to cover the most likely scenarios. Advice on further testing which may be needed to assist in deciding if a chemical is an ED is generally limited to a single next step, and this GD therefore does not present an entire hazard testing strategy for endocrine disruption.

10. The key advice for each assay is given in tabular format listing a series of scenarios (see Section C). These scenarios describe combinations of different assay results, provide advice on interpreting assay results and on further testing. However, each table should be read in conjunction with the preceding text that explains issues related to the assay and for which there is insufficient space in the tabular format. Once again, it is important to note that these tables (so-called "building blocks") are purely advisory, so individual regulatory authorities are not in any way bound to follow the advice. This is all the more important given that the guidelines for testing for endocrine disruption are still relatively new and the field will probably develop further.

A.6. Scope and limitations

A.6.1. Assays and endocrine modalities covered

11. The scope of the main section of the GD is limited to providing guidance on how to interpret results from assays included in the OECD Conceptual Framework (CF) for testing and assessment of EDs (see Section A.2). As the field of endocrine disruption is still developing, the CF will be subject to periodic revisions. In fact, during the updating of the GD, the CF was revised for the second time. The assays discussed are most of those included in the original CF plus some additional assays added in 2017 that were considered relevant to assessment of endocrine disruption. Some other assays have been added to the CF that are not included in this GD but may be useful when new assays for EDs are considered and validated in the future. Guidance is provided on the endpoints for the assays discussed, with respect to the endocrine modalities listed below. This is followed by guidance on how to strengthen the WOE that a chemical is/is not an ED based on the result from the assay under consideration and other existing relevant information. Various scenarios are considered and the guidance suggests different considerations and the next test that may be performed in a single step.

12. Detailed guidance is given for the most relevant and fully validated assays in the CF from the perspective of ED identification, while more limited guidance is provided for newer assays which are still in the process of being validated. The GD is limited to endocrine mechanisms and hazard assessment. Information on chemical exposure (e.g. on use, volume, fate, levels, duration and route) is not considered.

13. The GD mainly covers the same endocrine modalities as the original CF, i.e.:

- estrogen mediated (E)
- androgen mediated (A)
- thyroid hormone mediated (T)

- steroidogenesis interference (S).

However, some assays covering apical responses to the juvenile hormone and ecdysteroid (Ec) modalities in arthropods are now included, although none have specific mechanistic endpoints for these modalities. Possible effects on the retinoic acid pathway are also included, following the publication of the *Detailed Review Paper on the State of the Science on Novel* In Vitro *and* In Vivo *Screening and Testing Methods and Endpoints for Evaluating Endocrine Disrupters* (OECD, 2012) and the draft "Detailed review paper on the retinoid system" (OECD, 2017).

14. Although the assays in this guidance are applicable to most types of EDs and endocrine active substances (EASs) which are currently known (i.e. those operating via estrogen/androgen/thyroid/steroidogenesis [E,A,T,S] modalities), it should be recognised that the assays may not be responsive to certain poorly understood chemical types or MOA. For example, it is unlikely that EDs that damage the corticosteroid system of non-mammalian species will be covered (Trenzado, Carrick and Pottinger, 2003) although the adrenals are examined in many mammalian assays, therefore providing an alert. Several modalities in vertebrates are also not covered by available assays. Equally, there are no validated assays available for assessing mechanisms of endocrine activity in invertebrates, so it is not at present possible to conclude that a chemical is an ED in invertebrates.

Epigenetic effects

15. There is a growing body of evidence that some EDs may operate through epigenetic mechanisms (although such effects are not confined to EDs). Such potential effects have been reviewed and discussed *inter alia* by Crews et al. (2014) and Nilsson and Skinner (2015), and a more extensive analysis of human data, with experimental systems is provided in recent reviews by Marczyo, Jacobs and Gant (2016) and Jacobs et al. (2017). In essence, an epigenetic effect is a change in phenotype or gene expression, inherited over rounds of cell division and sometimes transgenerationally, caused by mechanisms other than alterations in gene sequence (e.g. histone modifications, DNA methylation, RNAi mediated gene silencing). It has been suggested that epigenetic changes may result in transgenerational phenotypic effects that even occur in the absence of continued environmental exposures. It is currently unclear whether the long-term assays available for testing for endocrine disruption (e.g. insect, fish, avian and rodent life cycle tests) would reveal the full range of potential epigenetic responses. For vertebrate wildlife, the information is more scarce; for example, Brown, Schultz and Nagler (2009) failed to observe heritable reproductive defects in the offspring of male rainbow trout exposed to a strong estrogen. In contrast, two recent studies do correlate ED effects. A first report shows increased intragenic DNA methylation of the follicle stimulating hormone receptor (Fshr) gene within the gonad tissue of juvenile female European eels (*Anguilla anguilla*) from highly polluted compared with lightly polluted French waters: correlated with increased levels of gonadal persistent organic pollutants (POPs) and metals, decreased Fshr mRNA, and reduced gonad development in the highly polluted eels (Pierron et al., 2014). The second measured reproductive impairments in 75-85 differentially methylated DNA regions in the red blood cells sampled from adult male American alligators (*Alligator mississippiensis*) living in POP and metal contaminated lakes (Guillette et al., 2016). Genes associated with the differentially methylated DNA regions were within pathways of endocrine relevance. *In vitro* and *in vivo* testing methods of epigenomic endpoints for evaluating endocrine disruption were reviewed as part of the OECD *Detailed Review Paper on the State of the Science on Novel* In Vitro *and* In Vivo *Screening and Testing Methods and Endpoints for Evaluating Endocrine Disrupters* (OECD, 2012; Greally and Jacobs,

2013) and in the last five years more relevant *in vitro* test systems, and human cancer data are now available (Jacobs et al., 2017; Parfett and Desaulniers, 2017; Alavian-Ghanini and Rüegg, 2017) and are being integrated into OECD work on non-genotoxic carcinogenesis, including endocrine modes of carcinogenesis. In summary, there is scope for epigenetic factors to have a role in some types of endocrine disruption. Assays for these effects are currently being discussed but have not yet been standardised by the OECD.

A.6.2. Scope of assessment and restriction to single assays

16. This GD does not present a testing strategy as it is restricted to a single step when further testing is recommended or proposed for consideration. It only recommends the most appropriate assay that could be performed if authorities need more evidence to support a regulatory decision. The proposed guidance is not meant to encourage animal testing. It encourages the maximal use of all existing information consistent with the OECD's integrated approaches to testing and assessment (OECD, 2008).

17. The level of confidence about whether or not a compound impacts endocrine function will increase with combined lines of pertinent evidence from multiple studies and endpoints across taxa, and which encompass different life stage effects and a range of doses. The amount of evidence needed to decide whether a substance is an ED in a regulatory context will depend on different authorities' policies/frameworks and the regulatory decision context. For example, results from a particular test or building block may suffice when taking a decision for priority setting but may not be adequate for more predictive hazard identification/characterisation.

18. Detailed guidance is not given on the conduct of WOE evaluations, or the relevance for human health of results from the assays considered. Section B.5 provides a summary of current WOE guidance that may be helpful for endocrine assessment. It is acknowledged that some mechanisms of action in rodents may not be relevant for humans, but the human relevance of specific mechanisms is not discussed.

19. Furthermore, the guidance does not consider exposure; however, this should be included when deciding whether further testing is needed in order to avoid unnecessary animal tests. This may be particularly relevant to non-mammalian wildlife where environmental hazard assessment aims at deriving a safe exposure level in the form of a predicted no-effect concentration (PNEC) or an environmental quality criterion/ environmental quality standard for the chemical. Traditionally the approach aims at the protection of all species in the relevant environmental compartment. For this purpose it is relevant to compare the sensitivity of several species in the compartment in question. If data are available on potential ED effects in more than one species/taxon, further testing may first be performed with the most sensitive species/taxon provided that it is possible to identify this taxon based on available information. Lastly, as in any evaluation, it is essential that the degree of confidence and uncertainty be communicated in the characterisation of the conclusions.

A.6.3. Rationale for assay inclusion

20. Detailed guidance is provided on the validated[2] and/or mainly widely accepted assays in the 2017 revised CF; these are listed in Parts A and B of Table A.1. Those assays listed under (A) are established methods, either with endocrine active substance (EAS)-specific endpoints or with non-specific sensitivity to EASs, which have been validated and published as OECD test guidelines. Assays listed under (B) have not received full validation by the OECD, or are in the process of OECD validation, or are guidelines which

have been validated and published by other organisations. The terms "validation" and "validated assays" are used as defined in the OECD GD on the "Validation and international acceptance of new or updated test methods for hazard assessment" (OECD, 2005) (see also Glossary). Validation may have been conducted by the OECD or other organisations (e.g. the Interagency Coordinating Committee on the Validation of Alternative Methods [ICCVAM]). Note that the word "assay" is used here to be consistent with the terminology used in the CF and describes a "test method" as defined in OECD (2005): "a test method is an experimental system that can be used to obtain a range of information from chemical properties through the adverse effects of a substance. The term 'test method' may be used interchangeably with 'assay' for ecotoxicity as well as for human health studies". The word "screen" is used in this document to describe *in vitro* or *in vivo* assays which primarily provide information on an endocrine disruption mechanism, and also occasionally information on adverse effects for use in hazard identification/characterisation. However, some regulatory authorities may wish to use positive screening tests for preliminary hazard identification/characterisation. Screens are generally rapid, and often simple, test methods and may have a truncated response range. On the other hand, the word "test" covers *in vivo* assays which can provide evidence to support a conclusion that a chemical is an endocrine disrupter that may cause adverse effects in an intact organism. An example of a screen would be the estrogen binding assay which only measures receptor binding activity *in vitro*, whereas an example of a test would be the Medaka Extended One-Generation Reproduction Test (MEOGRT), which measures reproductive success in intact fish. "Screen" and "test" are also broadly defined in OECD (2005), but here the word "test" is used more precisely, see the Glossary for all terms.

21. All of the assays in Parts A and B of Table A.1 are now included in the 2017 revised CF. Assays with non-specific sensitivity to EAS (e.g. OECD TG 408 repeated dose 90-day oral toxicity study in rodents and OECD TG 451-3 combined chronic toxicity/carcinogenicity studies) contain relevant endocrine endpoints (e.g. weights and histopathology of sex organs), and are used as such for REACH (OECD TG 408) and pesticide dossier evaluation in the EU, for example. OECD TG 453 (combined chronic toxicity/carcinogenicity studies) provides information on carcinogenicity in endocrine tissues and is therefore very important for endocrine assessment of chemicals. OECD TG 421 (reproduction/developmental toxicity screening test) and OECD TG 422 (combined repeated dose toxicity study with the reproduction/developmental toxicity screening test) provide information on reproduction in addition to effects on endocrine organs and are also used for REACH.

22. The TGs in the CF are also included in regional testing frameworks, such as the United States Environmental Protetion Agency's (US EPA) Endocrine Disruptor Screening Program (EDSP), although it should be noted that the two-tier structure of the EDSP differs from the CF paradigm. All of the US EPA TGs within this tiered testing strategy utilise the CF assays. Further information on the EDSP can be found in Section B.6.

23. Assays in Parts A and B of Table A.1 and the modalities they can detect (of E, A, T, S) are shown in Table A.2. This table is intended to be a guide only and does not reflect which assays are most relevant or have the most endpoints for detecting these modalities. A more detailed listing of endpoints and their responses can be found in Table B.1.

A.6.4. Rationale for assay exclusion

24. Assays mentioned in the 2017 revised CF but not covered in this document are listed in Part C of Table A.1. Guidance for these assays has generally been omitted either

because they have not yet started validation (e.g *in vitro* assays for determining disruption of thyroid function), there is insufficient experience in their use (e.g. invertebrate life cycle assays), they are thought not to offer significant advantages over existing tests (e.g. fish hepatocyte vitellogenin assay) or because they failed validation (e.g. MCF-7 cell proliferation assay). The list has been expanded to include other non-TG *in vitro* assays in common use, regardless of their validation status.

25. Progress has been made in the development of *in vitro* screening assays for disruption of thyroid function (OECD, 2006). The thyroid scoping document (OECD, 2014) reviews several key biological mechanisms of thyroid system disruption for their "state of readiness" as candidates for validation: short term (A), intermediate (B) or long term (C). *In vitro* assays identified as A or B are listed in Part C of Table A.1. *In vitro* assays with long-term validation plans will not be further discussed here. The thyroperoxidase (TPO) assay is now included in the high throughput screens (HTS) conducted by the US EPA (https://actor.epa.gov/dashboard; https://pubchem.ncbi.nlm.nih.gov), and a QSAR model predicting TPO inhibition based on a comprehensive training set of this HTS has recently been published (Paul et al., 2014; Rosenberg et al., 2016). The US EPA has now developed several high throughput thyroid screening assays and the EU Joint Research Centre is currently initiating a validation study involving EURL ECVAM's European Union Network of Laboratories for the Validation of Alternative Methods (EU-NETVAL) to assess the performance of a number of assays for disturbance of thyroid hormone function.

26. *In vitro* and *in vivo* assays for disruption of the function of the thyroid hormone system were discussed in a recent workshop on thyroid disruption. The output of the workshop is available and has helped to inform further assay development (EU, 2017).

27. The yeast estrogen (YES) and yeast androgen screens have also not been included in Parts A or B of this OECD guidance, although they are commonly used as *in vitro* screens in ecotoxicology (Routledge and Sumpter, 1996; Sohoni and Sumpter, 1998). They may suffer from limitations such as problems with materials that have fungicidal activity or inhibit cell proliferation, solubility, permeability or transport issues across the cell wall (ICCVAM, 2003). It has also been reported that the YES assay is not sensitive for anti-estrogenic chemicals (Fang et al., 2000) The detailed review paper on "Environmental endocrine disrupter screening: The use of estrogen and androgen receptor binding and transactivation assays in fish" (OECD, 2010a) also describes these assays.

28. However, in spite of these limitations, they are widely used as they are easy to handle and require no sophisticated lab equipment. Furthermore, in the absence of complex gene-regulating networks in the yeast cells, no cross talk is possible between other hormonal pathways and only the respective hormonal signalling (i.e. receptor binding) is captured. Variants of the yeast-based assays (*Saccharomyces cerevisiae* and *Arxula adeninivorans*) carrying the human ERα-receptor have recently been standardised within the International Organization for Standardization (ISO): ISO 19040 series: Determination of the estrogenic potential of water and waste water,[3] together with human cell line based transactivation assays.

- ISO 19040-1. Water quality – Determination of the estrogenic potential of water and waste water – Part 1: Yeast estrogen screen (YES, *Saccharomyces cerevisiae*)
- ISO 19040-2. Water quality – Determination of the estrogenic potential of water and waste water – Part 2: Yeast estrogen screen (A-YES, *Arxula adeninivorans*)
- ISO 19040-3. Water quality – Determination of the estrogenic potential of water and waste water – Part 3: *In vitro* human cell-based reporter gene assay.

29. The three ISO standards have a similar core. This core covers issues such as scope, sample handling, glassware, etc. ISO 19040-1 and ISO 19040-2 use two different yeast species. Various properties differ between these species, e.g. ISO 19040-2 is more salt tolerant (used for seawater). They also differ in media and processing requirements, etc. ISO 19040-3 is a generic standard that covers human cell lines (not yeast). Any human cell line that meets the criteria of the standard (e.g. the EC50 of the reference), would be a valid cell line for this standard.

30. The YES assays will be finalised as ISO standards in 2018. The YES is therefore included in the CF. ISO may take on the validation of the androgen receptor transactivation assays (including the yeast androgen screen [YAS] assays) in the future. If this happens, this process will likely take several years to complete.The YES and YAS assays could be considered to be the functionally similar to the ER and the AR stably transfected transactivation assay (STTA) assays and many of the possible next steps to be taken would be the same. These "building blocks" could therefore be used cautiously to provide guidance for the YES and YAS assays, but noting the limitations described above. The guidance for the ER STTA (OECD TG 455) would cover the YES assay and is given in Section C.1.2. The guidance for the AR STTA would cover the YAS assay and is given in Section C.1.3.

31. Other nuclear hormone receptor-based transactivation assays have also become more commonly used in research, including the aryl hydrocarbon receptor, the peroxisome proliferator-activated receptor (alpha and gamma), the liver X receptor, the vitamin D receptor, retinoid receptors (retinoid X receptor, retinoic acid receptor), the constitutive androstane receptor, the pregnane X receptor and the growth hormone receptor. Although none of these assays have been formally validated, many of them are included in high throughput screens conducted by the US EPA and have now become publically available.[4]

32. Guidance about tests that are based on the induction of proliferation (e.g. the E-screen where proliferation in estrogen-responding cells, particularly in the MCF-7 human breast cancer cell line), is used to detect estrogenic activity (Soto and Sonnenschein, 2001) is also not included. Proliferation assays are not recommended by the ICCVAM (2003) because cell proliferation can be mediated through pathways other than those involving transcriptional activation of estrogen responsive genes. The National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods in the United States co-ordinated an international interlaboratory validation study of a MCF-7 cell proliferation test method. The validation study was completed in 2011. Although accuracy of the ER agonist protocol was high at the lead laboratory and sufficient in the partner labs, interlaboratory reproducibility was insufficient for the method to proceed further. The study indicated that the test method protocols, especially the antagonist protocol, required additional development before this method could be considered validated (https://ntp.niehs.nih.gov/pubhealth/evalatm/test-method-evaluations/endocrine-disruptors/index.html).

Table A.1. **Screens and tests for which guidance is provided in this document**

The definitions of "screen" and "test" are given in Section A.6.3 and the Glossary.

Those assays listed under (A) are established methods, either with endocrine active substance-specific endpoints or with non-specific sensitivity to endocrine active substances, which have been validated and published as OECD test guidelines. Assays listed under (B) have not received full validation by the OECD, or are in the process of OECD validation, or are guidelines which have been validated and published by other organisations. Guidance for both assay types can be found in Section C of this guidance document, but it is important to note that this guidance is not yet definitive for those assays still undergoing validation. Assays listed in this table under (C) are largely those appearing in the Conceptual Framework (as revised in 2017) but for which no guidance is provided because they are not yet the subject of a guideline, or assays where there is insufficient experience in their use with endocrine active substances. Guidance for assays under (C) will be written when the validation of the assays is more advanced. All assays have been sorted according to which level they should occupy in the CF, although some do not yet appear in it.

It is important to bear in mind that the CF (see Section A.2) is not a testing strategy to be followed linearly from Level 1 through to Level 5, although in cases where little or no information is available (i.e. for new chemicals), it could provide guidance about where to start testing. Level 1 gathers together existing information, Level 2 covers *in vitro* mechanistic assays, Level 3 covers *in vivo* assays providing some data about selected endocrine mechanism(s)/pathway(s) (in some cases they also provide information about generally recognised hazard endpoints which, however, in some cases with other data may be robust enough for regulatory decision making), Level 4 covers *in vivo* assays providing some data on adverse effects of endocrine-relevant endpoints, and Level 5 covers *in vivo* assays which provide more comprehensive data on adverse effects over more extensive parts of an organism's life cycle.

It should be noted that the invertebrate assays generally report on apical effects of various types, and do not allow conclusions about mechanism (with the possible exception of the short-term juvenile hormone activity screening assay using *Daphnia magna*). At present, it is therefore not possible in most cases to reach conclusions about whether chemicals are endocrine disruptors in invertebrates.

Conceptual Framework level	*In vitro* screens	Mammalian *in vivo* screens and tests	Non-mammalian *in vivo* screens and tests
A. OECD test guidelines with endocrine active substance-specific endpoints or with non-specific sensitivity to endocrine active substances			
2	– OECD TG 493: Performance-Based Test Guideline for Human Recombinant Estrogen Receptor (hrER) *In Vitro* Assays to Detect Chemicals with ER Binding Affinity – OECD TG 455: Performance-Based Test Guideline for Stably Transfected Transactivation *In Vitro* Assays to Detect Estrogen Receptor Agonists and Antagonists – OECD TG 458: Stably Transfected Human Androgen Receptor Transcriptional Activation Assay for Detection of Androgenic Agonist and Antagonist Activity of Chemicals – OECD TG 456: H295R Steroidogenesis Assay	Not applicable	Not applicable
3	Not applicable	– OECD TG 440: Uterotrophic Bioassay in Rodents (UT Assay) (including OECD GD 71 on the use of the assay to screen for anti-estrogenicity) (screen) – OECD TG 441: Hershberger Bioassay in Rats (H Assay) (including OECD GD 115 on the Weanling Hershberger Bioassay) (screen)	– OECD TG 229: Fish Short-Term Reproduction Assay (FSTRA) (screen/test) – OECD TG 230: 21-Day Fish Assay (screen) – OECD TG 231: Amphibian Metamorphosis Assay (AMA) (screen)
4	Not applicable	– OECD TG 407: Repeated Dose 28-Day Oral Toxicity Study in Rodents (test) – OECD TG 408: Repeated Dose 90-Day Oral Toxicity Study (test) – OECD TG 451-3: Combined Chronic Toxicity/Carcinogenicity Studies (test) – OECD TG 421: Reproduction/Developmental Toxicity Screening Test – OECD TG 422: Combined Repeated Dose Toxicity Study with the Reproduction/Developmental Toxicity Screening Test (test) – OECD TG 414: Prenatal Developmental Toxicity Study (test) – OECD TG 426: Developmental Neurotoxicity Study (test) – OECD TG 410: Repeated Dose Dermal Toxicity: 21/28-Day Study (test) – OECD TG 411: Subchronic Dermal Toxicity: 90-Day Study (test) – OECD TG 412 28-Day (Subacute) Inhalation Toxicity Study (test) – OECD TG 413: Subchronic Inhalation Toxicity: 90-Day Study (test) – TG 409: Repeated dose 90-Day Oral Toxicity Study in Non-Rodents (test)	– OECD TG 242: *Potamopyrgus antipodarum* Reproduction Test (test) – OECD TG 243: *Lymnaea stagnalis* Reproduction Test (test) – OECD TG 218-219: Chironomid Toxicity Test (test) – OECD TG 211: *Daphnia* Reproduction Test (with Male Induction) (test) – OECD TG 210: Fish Early Life Stage Toxicity Test (test) – OECD TG 234: Fish Sexual Development Test (FSDT) (test) – OECD TG 241: Larval Amphibian Growth and Development Assay (LAGDA) (test) – OECD TG 206: Avian Reproduction Test (test)

Conceptual Framework level	*In vitro* screens	Mammalian *in vivo* screens and tests	Non-mammalian *in vivo* screens and tests
5	Not applicable	– OECD TG 443: Extended One-Generation Reproductive Toxicity Study (EOGRTS) (test) – OECD TG 416: Two-Generation Reproduction Toxicity Study (test)	– OECD TG 233: Sediment Water Chironomid Life Cycle Toxicity Test (test) – OECD TG 240: Medaka Extended One-Generation Reproduction Test (MEOGRT) (test)
B. Guidelines that have not received full validation by the OECD, or are in the process of OECD validation or which have been validated and published by other organisations			
2	– AR Binding Assay (US EPA OPPTS 890.1150) – Aromatase Assay (US EPA OPPTS 890.1200)	Not applicable	Not applicable
3	Not applicable	No assays in this category	– Short-Term Juvenile Hormone Activity Screening Assay Using *Daphnia magna* (draft OECD TG) (screen) – OECD GD 148: Androgenised Female Stickleback Screen (AFSS) (screen) – EASZY Assay. Detection of Substances Acting Through Estrogen Receptors Using Transgenic cyp19a1b GFP Zebrafish Embryos (draft OECD TG) (screen) – Juvenile Medaka Anti-Androgen Screening Assay (JMASA) (draft OECD GD) (screen) – *Xenopus* Embryonic Thyroid Signalling Assay (XETA) (draft OECD TG) (screen) – Rapid Androgen Disruption Adverse Outcome Reporter (RADAR) Assay (draft OECD TG) (screen)
4	Not applicable	– Pubertal Development and Thyroid Function Assay in Peripubertal Male Rats (Male PP Assay) (US EPA OPPTS 890.1500) (screen) – Pubertal Development and Thyroid Function Assay in Peripubertal female Rats (Female PP Assay) (US EPA OPPTS 890.1450) (screen)	– New Guidance Document on Harpacticoid Copepod Development and Reproduction Test with *Amphiascus* (OECD GD 201)[2] (test)
5	Not applicable	No assays in this category	– *Daphnia* Multigeneration Test for Assessment of EDCs (draft OECD TG) (test) – Fish Life cycle Toxicity Test (FLCTT) (US EPA OPPTS 850.1500) (test) – Zebrafish Extended One-Generation Reproduction Test (ZEOGRT) (draft OECD TG) (test) – Avian Two-Generation Toxicity Test in the Japanese Quail (ATGT) (US EPA OCSPP 890.2100/ 740-C-15-003) (test)

Conceptual Framework level	*In vitro* screens	Mammalian *in vivo* screens and tests	Non-mammalian *in vivo* screens and tests
C. Assays corresponding to those in the Conceptual Framework (original or revised 2017) for which no guidance has been written at present			
2	See Section A.6.4 for more details. – Thyroperoxidase inhibition – Transthyretin binding – Thyroid binding globulin binding – Sodium-iodine symporter modulation – Thyroid hormone receptor and thyroid stimulating hormone receptor modulation – T3 deiodinase inhibition – Thyroid hormone receptor transactivation – Thyroid gland explant – Stably Transfected Human Thyroid Receptor Transactivation Assay (TR STTA) – Retinoid receptor transactivation assays (retinoic acid receptor, retinoid X receptor) – Liver receptor ransactivation assays (aryl hydrocarbon receptor, constitutive androstane receptor, pregnane X receptor, peroxisome proliferator-activated receptor, liver X receptor, vitamin D receptor) – Other hormone receptors (e.g. glucocorticoid receptor, growth hormone receptor) – Fish hepatocyte vitollegenin assay – Yeast transactivation assays (yeast estrogen screen and yeast androgen screen) – Proliferation-based screens e.g. MCF-7 cell proliferation assay (estrogen receptor ant/agonist) – High-throughput screens (see OECD Guidance Document No. 211 for Describing Non-Guideline *In Vitro* Test Methods)	Not applicable	Not applicable
3	Not applicable	No assays in this category	No assays in this category
4	Not applicable	No assays in this category	– Fish Reproduction Partial Life Cycle Test (when/if TG is available) (test) – Earthworm Reproduction Test (OECD TG 222) (test) – Enchytraeid Reproduction Test (OECD TG 220) (test) – Sediment Water *Lumbriculus* Toxicity Test Using Spiked Sediment (OECD TG 225) (test) – Predatory mite reproduction test in soil (OECD TG 226) (test) – Collembolan Reproduction Test in Soil (OECD TG 232) (test)
5	Not applicable	No assays in this category	No assays in this category

Table A.2. **List of assays for which guidance has been developed in GD 150, showing their responsiveness to various endocrine modalities**

Assay name	CF Level	Mode of action which may produce a response						
		E	A	S	T	JH	Ec	R
OECD TG 493: PBTG for Human Recombinant Estrogen Receptor (hrER) *In Vitro* Assays to Detect Chemicals with ER Binding Affinity	2	■						
OECD TG 455: PBTG for Stably Transfected Transactivation *In Vitro* Assays to Detect Estrogen Receptor Agonists and Antagonists	2	■						
OECD TG 458: Stably Transfected Human Androgen Receptor Transcriptional Activation Assay for Detection of Androgenic Agonist and Antagonist Activity of Chemicals	2		■					
OECD TG 456: H295R Steroidogenesis Assay	2			■				
AR Binding Assay (US EPA OPPTS 890.1150)	2		■					
Aromatase Assay (US EPA OPPTS 890.1200)	2			■				
OECD TG 440: Uterotrophic Bioassay in Rodents	3	■						
OECD TG 441: Hershberger Bioassay in Rats	3		■		■			
OECD TG 229: Fish Short-Term Reproduction Assay	3	■	■	■				
OECD TG 230: 21-Day Fish Assay	3	■	■	■				
OECD TG 231: Amphibian Metamorphosis Assay	3				■			
OECD GD 148: Androgenised Female Stickleback Screen	3		■					
EASZY Assay: Detection of Substances Acting through Estrogen Receptors using Transgenic cyp19a1b GFP Zebrafish Embryos	3	■						
JMASA: Juvenile Medaka Anti-Androgen Screening Assay	3		■					
XETA: *Xenopus* Embryonic Thyroid Signalling Assay	3				■			
RADAR: Rapid Androgen Disruption Adverse Outcome Reporter Assay	3		■					
SJHASA: Short-Term Juvenile Hormone Activity Screening Assay Using *Daphnia magna*	3					▒		
OECD TG 407: Repeated Dose 28-Day Oral Toxicity Study in Rodents	4	■	■	■	■			▒
OECD TG 408: Repeated Dose 90-Day Oral Toxicity Study	4	■	■	■	■			▒
OECD TG 451-3: Combined Chronic Toxicity/Carcinogenicity Studies	4	■	■	■	■			▒
OECD TG 421 and 422: Combined 28-Day Reproductive Screening Tests	4	■	■	■	■			▒
OECD TG 414: Prenatal Developmental Toxicity Study	4	■	■	■	■			▒
OECD TG 426: Developmental Neurotoxicity Study	4	■	■	■	■			▒
OECD TG 410: Repeated Dose Dermal Toxicity: 21/28-Day Study	4	■	■	■				▒
OECD TG 411: Subchronic Dermal Toxicity: 90-Day Study	4	■	■	■	■			▒
OECD TG 412: 28-Day (Subacute) Inhalation Toxicity Study	4	■	■	■	■			▒
OECD TG 413: Subchronic Inhalation Toxicity: 90-Day Study	4	■	■	■	■			▒
TG 409: Repeated Dose 90-Day Oral Toxicity Study in Non-Rodents	4	■	■	■	■			▒
Pubertal Development and Thyroid Function Assay in Peripubertal Male Rats (US EPA OPPTS 890.1500)	4	■	■	■	■			▒
Pubertal Development and Thyroid Function Assay in Peripubertal female Rats (US EPA OPPTS 890.1450)	4	■	■	■	■			▒
OECD TG 210: Fish Early Life Stage Toxicity Test	4				▒			
OECD TG 234: Fish Sexual Development Test	4	■	■	■	▒			
OECD TG 241: Larval Amphibian Growth and Development Assay	4	■	■	▒	■			
OECD TG 206: Avian Reproduction Test	4	▒	▒	▒	▒			
OECD TG 242: *Potamopyrgus antipodarum* Reproduction Test	4							▒
OECD TG 243: *Lymnaea stagnalis* Reproduction Test	4							▒
OECD TG 218-219: Chironomid Toxicity Test	4					▒	▒	
OECD TG 211: *Daphnia* Reproduction Test (with male induction)	4					▒	▒	

Table A.2. **List of assays for which guidance has been developed in GD 150, showing their responsiveness to various endocrine modalities** (*continued*)

Assay name	CF Level	Mode of action which may produce a response						
		E	A	S	T	JH	Ec	R
OECD GD 201: New Guidance Document on Harpacticoid Copepod Development and Reproduction Test with *Amphiascus*	4							
OECD TG 416: Two-Generation Reproduction Toxicity Study	5							
OECD TG 443: Extended One-Generation Reproductive Toxicity Study	5							
OECD TG 240: Medaka Extended One-Generation Reproductive Toxicity Study	5							
FLCTT: Fish Life Cycle Toxicity Test (US EPA OPPTS 850.1500)	5							
ZEOGRT: Zebrafish Extended One-Generation Reproduction Test	5							
ATGT: Avian Two-Generation Toxicity Test in the Japanese Quail (US EPA OCSPP 890.2100/ 740-C-15-003)	5							
OECD TG 233: Sediment Water Chironomid Life Cycle Toxicity Test	5							
DMGT: *Daphnia* Multigeneration Test for Assessment of EDCs	5							

Notes: Dark blue = some endpoints responsive to and diagnostic of modality; light blue = endpoints responsive to but not diagnostic of modality. This table is intended to be a guide only and does not reflect which assays are most relevant or have the most endpoints responsive to these modalities. Modality abbreviations: E,A,T,S: estrogen/androgen/thyroid/steroidogenesis; JH: juvenile hormone, Ec: ecdysone, R: retinoid, -related modalities.

Notes

1. Endocrine disruption-responsive assays are those *in vitro* or *in vivo* assays whose endpoints are known to respond positively to endocrine disrupters and/or endocrine active substances.
2. These are assays which have been validated at the national or international level, especially as OECD TGs.
3. See, for example, https://www.iso.org/standard/64451.html.
4. See, for example, https://actor.epa.gov/dashboard and https://pubchem.ncbi.nlm.nih.gov.

References

Alavian-Ghavanini, A. and J. Rüegg (2017), "Epigenetic effects of endocrine disrupting chemicals: From mechanisms to novel test methods", *Basic & Clinical Pharmacology & Toxicology*, Vol. 122/1, pp. 38-45, http://dx.doi.org/10.1111/bcpt.12878.

Ankley, G.T. et al. (2010), "Adverse outcome pathways: A conceptual framework to support ecotoxicology research and risk assessment", *Environmental Toxicology and Chemistry*, Vol. 29/3, pp. 730-741, http://dx.doi.org/10.1002/etc.34.

Boobis, A.R. et al. (2008), "IPCS framework for analyzing the relevance of a noncancer mode of action for humans", *Critical Reviews in Toxicology*, Vol. 38/2, pp. 87-96, http://dx.doi.org/10.1080/10408440701749421.

Boobis, A.R. et al. (2006), "IPCS framework for analyzing the relevance of a cancer mode of action for humans", *Critical Reviews in Toxicology*, Vol. 36/10, pp. 781-792, https://doi.org/10.1080/10408440600977677.

Brown, K.H., I.R. Schultz and J.J. Nagler (2009), "Lack of a heritable reproductive defect in the offspring of male rainbow trout exposed to the environmental estrogen 17-alpha-ethynylestradiol", *Aquatic Toxicology*, Vol. 91/1, pp. 71-74, https://doi.org/10.1016/j.aquatox.2008.10.009.

Crews, D. et al. (2014), "Nature, nurture and epigenetics", *Molecular and Cellular Endocrinology*, Vol. 398/1-2, pp. 42-52, https://doi.org/10.1016/j.mce.2014.07.013.

Dellarco, V. and P.A. Fenner-Crisp (2012), "Mode of action: Moving toward a more relevant and efficient assessment paradigm", *The Journal of Nutrition*, Vol. 142/12, pp. 2192S-2198S, https://doi.org/10.3945/jn.111.157396.

EC (2016), "Communication from the Commission to the European Parliament and the Council on endocrine disrupters and the draft Commission acts setting out scientific criteria for their determination in the context of the EU legislation on plant protection products and biocidal products", COM(2016)350, European Commission, Brussels.

EFSA (2013), "Scientific opinion on the hazard assessment of endocrine disrupters: Scientific criteria for identification of endocrine disrupters and appropriateness of existing test methods for assessing effects mediated by these substances on human health and the environment", *EFSA Journal*, Vol. 11/3, pp. 3132, http://dx.doi.org/10.2903/j.efsa.2013.3132.

EU (2017), *Supporting the Organization of a Workshop on Thyroid Disruption*, European Commission, Brussels, http://dx.doi.org/10.2779/921523.

Fang, H. et al. (2000), "Quantitative comparisons of *in vitro* assays for estrogenic activities", *Environmental Health Perspectives*, Vol. 108/8, pp. 723-729.

Greally, J.M. and M.N. Jacobs (2013), "*In vitro* and *in vivo* testing methods of epigenomic endpoints for evaluating endocrine disrupters", *ALTEX – Alternatives to Animal Experimentation*, Vol. 30/4, pp. 445-471, https://doi.org/10.14573/altex.2013.4.445.

Guillette, L.J. Jr. et al. (2016), "Epigenetic programming alterations in alligators from environmentally contaminated lakes", *General and Comparative Endocrinology*, Vol. 238, pp. 4-12, https://doi.org/10.1016/j.ygcen.2016.04.012.

ICCVAM (2003), "ICCVAM evaluation of *in vitro* test methods for detecting potential endocrine disrupters: Estrogen receptor and androgen receptor binding and transcriptional activation assays", NIH Publication No. 03-4503, National Institute of Environmental Health Sciences.

IPCS/WHO (2004), *ICPS Risk Assessment Terminology*, World Health Organization, Geneva, www.inchem.org/documents/harmproj/harmproj/harmproj1.pdf.

Jacobs, M.N. et al. (2017), "Marked for life: Epigenetic effects of endocrine disrupting chemicals", *Annual Review of Environment and Resources*, Vol. 42, pp. 105-160, https://doi.org/10.1146/annurev-environ-102016-061111.

Marczylo, E.L., M.N. Jacobs and T.W. Gant (2016), "Environmentally induced epigenetic toxicity: Potential public health concerns", *Critical Reviews in Toxicology*, Vol. 46/8, pp. 676-700, http://dx.doi.org/10.1080/10408444.2016.1175417.

Nilsson, E.E. and M.K. Skinner (2015), "Environmentally induced epigenetic transgenerational inheritance of reproductive disease", *Biology of Reproduction*, Vol. 93/6, 145, pp. 1-8, https://doi.org/10.1095/biolreprod.115.134817.

OECD (2017), "Detailed review paper on the retinoid system (draft)", OECD, Paris.

OECD (2016), "Users' handbook supplement to the guidance document for developing and assessing adverse outcome pathways", *OECD Series on Adverse Outcome Pathways*, No. 1, OECD Publishing, Paris, http://dx.doi.org/10.1787/5jlv1m9d1g32-en.

OECD (2014), *New Scoping Document on* In Vitro *and* Ex Vivo *Assays for the Identification of Modulators of Thyroid Hormone Signalling*, OECD Series on Testing and Assessment, No. 207, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264274716-en.

OECD (2012), *Detailed Review Paper on the State of the Science on Novel* In Vitro *and* In Vivo *Screening and Testing Methods and Endpoints for Evaluating Endocrine Disrupters*, OECD Series on Testing and Assessment, No. 178, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264221352-en.

OECD (2010a), "Detailed review paper on environmental endocrine disrupter screening: The use of estrogen and androgen receptor binding and transactivation assays in fish", OECD Series on Testing and Assessment, No. 135, OECD, Paris, www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2010)34&doclanguage=en.

OECD (2010b), "Workshop report on OECD countries' activities and management of endocrine disrupters", OECD Series on Testing and Assessment, No. 118, OECD, Paris, www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/JM/MONO(2010)2&doclanguage=en.

OECD (2008), "Workshop report on integrated approaches to testing and assessment", OECD Series on Testing and Assessment, No. 88, OECD, Paris, www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2008)10&doclanguage=en.

OECD (2006), "Detailed review paper on thyroid hormone disruption assays", OECD Series on Testing and Assessment, No. 57, OECD, Paris, www.oecd.org/env/ehs/testing/37235405.pdf.

OECD (2005), "Guidance document on the validation and international acceptance of new or updated test methods for hazard assessment", OECD Series on Testing and Assessment, No. 34, OECD, Paris, www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?doclanguage=en&cote=env/jm/mono(2005)14.

Parfett, C.L. and D. Desaulniers (2017), "A Tox21 approach to altered epigenetic landscapes: Assessing epigenetic toxicity pathways leading to altered gene expression and oncogenic transformation *in vitro*", *International Journal of Molecular Sciences*, Vol. 18/6, http://dx.doi.org/10.3390/ijms18061179.

Paul, K.B. et al. (2014), "Development of a thyroperoxidase inhibition assay for high-throughput screening", *Chemical Research in Toxicology*, Vol. 27/3, pp. 387-399, http://dx.doi.org/10.1021/tx400310w.

Pierron, F. et al. (2014), "Abnormal ovarian DNA methylation programming during gonad maturation in wild contaminated fish", *Environmental Science & Technology*, Vol. 48/19, pp. 11688-11695, http://dx.doi.org/10.1021/es503712c.

Rosenberg, S.A. et al. (2016), "Development of a QSAR model for thyroperoxidase inhibition and screening of 72,526 REACH substances abstracts", *Toxicology Letters*, Vol. 258S(2016), pp. S62-S324.

Routledge, E.J. and J.P. Sumpter (1996), "Estrogenic activity of surfactants and some of their degradation products assessed using a recombinant yeast screen", *Environmental Toxicology and Chemistry*, Vol. 15/3, pp. 241-248, https://doi.org/10.1002/etc.5620150303.

Sohoni, P. and J.P. Sumpter (1998), "Several environmental oestrogens are also anti-androgens", *Journal of Endocrinology*, Vol. 158/3, pp. 327-339.

Solecki, R. et al. (2017), "Scientific principles for the identification of endocrine-disrupting chemicals: A consensus statement", *Archives of Toxicology*, Vol. 91/2, pp. 1001-1006, https://doi.org/10.1007/s00204-016-1866-9.

Soto, A.M. and C. Sonnenschein (2001), "The two faces of Janus: Sex steroids as mediators of both cell proliferation and cell death", *Journal of the National Cancer Institute*, Vol. 93/22, pp. 1673-1675, https://doi.org/10.1093/jnci/93.22.1673.

Trenzado, C.E., T.R. Carrick and T.G. Pottinger (2003), "Divergence of endocrine and metabolic responses to stress in two rainbow trout lines selected for differing cortisol responsiveness to stress", *General and Comparative Endocrinology*, Vol. 133/3, pp. 332-340, https://doi.org/10.1016/S0016-6480(03)00191-6.

WHO/IPCS (2009), *Principles and Methods for the Risk Assessment of Chemicals in Food*, Environmental Health Criteria 240, World Health Organization, Geneva, www.who.int/foodsafety/publications/chemical-food/en.

WHO/IPCS (2002), "Global assessment of the state-of-the-science of endocrine disrupters", Damstra, T. et al. (eds.), WHO/PCS/EDC/02.2, World Health Organization, Geneva, www.who.int/ipcs/publications/new_issues/endocrine_disruptors/en.

B. General guidance on endocrine assessment, assays and endpoints

33. The purpose of this section is to provide background information on the relevance of various types of data for supporting decisions about the endocrine disrupting properties of chemicals and other test materials (e.g. effluents, natural waters, contaminated foods, etc.) in humans and non-mammalian vertebrates. Interpretation of results from some invertebrate test guidelines is also included, but due to the rather poor current understanding of endocrinology in most invertebrates, and the lack of diagnostic screening endpoints with these taxonomic groups (e.g. OECD [2010c]), guidance cannot yet be given for many of these assays. Nevertheless, non-OECD test assays, including those utilising invertebrate species, may provide information that can be used in a weight of evidence (WOE) approach. Furthermore, the document only deals with estrogen-, androgen- and thyroid-mediated endocrine disruption, and with interference with steroidogenesis (although some guidance is also provided for evaluation of juvenile hormone, ecdysteroid and retinoid activity). It does not cover other possible types of endocrine disruption, such as effects on the hypothalamus-pituitary-adrenal axis or other receptor pathways. Some advice on the endocrine control of neural development is provided, but this is only rudimentary. The section is organised according to the OECD Conceptual Framework (CF) (see Section A.2), as updated in 2017 with tests which were unavailable or not included when it was first proposed.

34. It is important to bear in mind that the CF **is not a testing strategy** to be followed linearly from Level 1 through to Level 5, although in cases where little or no information is available (i.e. for new chemicals), it could provide ideas about where to start testing. In principle, any test can be conducted at any time in the hazard assessment process, depending on the perceived need for information. However, the data generated at various levels have a range of differing applications and implications, and must be interpreted accordingly. The purpose of this guidance document (GD) is therefore to assist assessors of endocrine-relevant tests with data interpretation in the light of information that may already exist, and to provide **optional** suggestions for obtaining additional data, if required, to increase confidence in conclusions on the endocrine disrupting possibilities of a particular chemical. It is clear that decisions about whether to obtain further data will be largely driven by regulatory needs which vary between jurisdictions, so advice on "next steps which could be taken to strengthen weight of evidence" is in no sense mandatory. As stated earlier, this process of data interpretation and assessment involves the need for a **weight of evidence approach** that considers both mechanistic and apical information, and it is self-evident that the more data which support a particular conclusion, the more reliable that conclusion will be.

35. This guidance supplements other GDs available on identification and interpretation of changes indicative of endocrine disruption such as the "Guidance document on mammalian reproductive toxicity testing and assessment" (OECD, 2008b), the "Guidance document for histologic evaluation of endocrine and reproductive tests in rodents" (OECD, 2009), the "Guidance document on the diagnosis of endocrine-related histopathology in fish gonads"

(OECD, 2010a) and the guidance document in support of the "Test guideline on the extended one-generation reproductive toxicity study" (OECD, 2013).

36. Subsequent sections of this document will deal separately and in detail with *in vitro* mechanistic screens and *in vivo* screens and tests covering endpoints relevant for humans or vertebrate wildlife. In the context of non-mammalian wildlife screens and tests, the test species are fish, amphibians, birds, molluscs, crustaceans and insects. General issues concerning such screens/tests are briefly considered together in this section. The distinction between screening assays used only for possible hazard identification and tests that may be used for more definitive hazard identification/characterisation is also discussed. The ability of the different assays at the different levels of the CF to detect endocrine disrupters (EDs) and endocrine active substances (EASs) is discussed briefly here and in more detail in Section C.

37. It should be remembered that due to the molecular similarities of endocrine systems and receptor homologies across the vertebrates, there may be some potential for using information from non-mammalian vertebrate test assays for assessing endocrine activity in mammals (and vice versa), and especially for extrapolation between various *in vitro* screens (see Section B.3). This must be tempered with the knowledge that outcomes associated with a given endocrine modality can vary significantly across the vertebrates, in large part due to variations in toxicokinetics and in absorption, distribution, metabolism and excretion (ADME) processes. The *in vitro* screens in question (although at present based largely on mammalian receptors and/or enzymes) are generally capable of providing information applicable to both humans and vertebrate wildlife (OECD, 2010d). Such extrapolation of *in vitro* information is generally qualitative (e.g. "Does the chemical bind to the estrogen receptor?") rather than quantitative (e.g. "What is the potency of the chemical in a particular taxonomic group?").

38. On the other hand, the purposes of the two *in vivo* assay types (mammals and non-mammalian wildlife) are rather different. Whereas mammalian assays may contribute mainly to hazard identification/characterisation whose objective is to protect individual human beings, non-mammalian assays were originally intended to provide information to help predict possible impacts on non-mammalian wildlife populations. This in turn may affect the way in which assay data are interpreted. For example, in the latter assays, ecotoxicologically relevant adverse effect endpoints used for regulatory decision making generally relate to mortality/survival, growth, development or reproduction. This may also apply to mammalian assays used for hazard identification/characterisation for protection of mammalian wildlife. Such assays may anyway provide useful information for hazard identification/characterisation across vertebrate species, including humans, because the fundamental approaches to such assessments are similar.

B.1. Considerations on the assays addressed

39. The considerations set out below are based partly on ideas proposed in Table 2 of OECD (2010b). However, they have been augmented with information relevant for non-mammalian wildlife testing, and have also been amended in the light of recent scientific developments.

B.1.1. Conceptual Framework Level 1: Existing data and non-test information

40. It is important to emphasise that before conducting any assessment of data from an endocrine disruption screen or test, all existing scientifically relevant and reliable information on the test chemical should be collated. Such data should ideally include

physico-chemical properties, and fate and behaviour, as well as any toxicological and ecotoxicological information. However, it is recognised that all these types of information may not be available.

41. Data on structural analogues and from quantitative structure-activity relationship (QSAR) models should be considered, especially if data on the chemical under consideration are scarce. QSAR models predicting mechanism and endocrine activity can be used for prioritisation, ranking and hazard identification (see below for more details). More advanced models (e.g. mode of action [WHO, 2007] or adverse outcome pathway models [Schultz, 2010; Ankley et al., 2010]), have also been developed (see Section B.5). Information from non-test methods may not only be "existing information" which is already generated, but predictions, models, read-across cases, etc. may also be generated as part of the assessment.

42. All existing relevant data should be maximally used (e.g. structural; physico-chemical information; *in vivo* and *in vitro* guideline and non-guideline testing; QSAR models; computational and other non-testing assays; toxicokinetic, pharmacokinetic and toxicodynamic information; category and read-across assessment methodologies) in a WOE approach before entering any other level of the CF. Ball et al. (2016) and Zhu et al. (2016) provide guidance and examples to support read-across using biological data. Such existing data/knowledge may be of great value when interpreting the results of endocrine screens/tests, but before they are used, their quality must be evaluated. A quality scoring system such as that recommended by Klimisch, Andreae and Tillmann (1997) or Schneider et al. (2009) can be helpful in this regard (see also Section B.5). Other guidance on this subject is provided by SciRAP and SYRINA (Beronius et al., 2014; Molander et al., 2015; Vandenberg et al., 2016; Ågerstrand et al., 2018). It is also important to know whether an *in vivo* endocrine disruption test has been performed at doses or concentrations which would not be expected to cause systemic toxicity that could mask endocrine effects, or which could cause misleading endocrine changes secondary to general or specific (non-endocrine) organ toxicities.

43. Information on metabolism and toxicokinetics is also very valuable. Any available toxicokinetic data (e.g. if OECD TG 417 [toxicokinetics] has been carried out) may help with decisions about route of administration for *in vivo* studies, the relevance of metabolism for *in vitro* studies and the relevance of results from one species to another. For example, if a chemical is metabolised then the addition of metabolising systems to *in vitro* tests should be considered (see below). Toxicokinetic studies may also provide information on bioavailability, half-lives for absorption and elimination, and clearance rates, and any non-linear kinetics resulting from saturation of absorption, which may help with interpretation of toxicity and endocrine data. *In silico* systems are also being developed to predict metabolism, e.g. "Metapath" is a system for simulating xenobiotic metabolism of pesticides and structurally similar molecules developed by the joint US, EU, Canadian and Australian project of the OECD Working Group of Pesticides.

44. Another important issue concerning initial data collation is the value of extrapolating data from mammalian tests when interpreting data from non-mammalian vertebrates, and vice versa. The broad similarity of endocrine systems across the vertebrates means that such extrapolation can be of considerable value, so it is vital that mammalian toxicologists and non-mammalian ecotoxicologists who assess endocrine disruption-related data should not operate without reference to each other (see Section B.3).

B.1.1.1. QSAR models

45. QSAR models for some endocrine modes of action (MOA) are now available. Estrogen receptor (ER) interaction is historically the most developed model with androgen receptor (AR) interaction models now becoming available and a thyroperoxidase model has recently been published and predictions will be available in 2018 at the *DK QSAR Database* website (Rosenberg et al., 2017). The Danish Environmental Protection Agency (DK EPA), the United States Environmental Protection Agency (US EPA) and the OECD provide websites that link to the models (some of which are free to users). The websites also host databases that provide outputs for thousands of chemicals and allow combined interrogation for many effects such as bioaccumulation, reproductive toxicity, etc.

46. The output of these models can be applicable (with caution) for interpretation of the mechanisms underlying *in vivo* results with vertebrates. Furthermore, these QSAR methodologies can be used to identify groups of chemicals and structural alerts that are linked to *in vivo* effects, thereby elucidating possible key MOA or mechanisms. The websites also provide links to further models. They may predict metabolic transformation (ADME) and possible cytochrome P450 metabolism that could be used in interpretation of, for example, disagreement between *in vitro* and *in vivo* results. Some sources of reference to QSARs and QMRF (QSAR model reporting formats) can be found on the websites and in Lo Piparo and Worth (2010), Jensen et al. (2011), Mombelli (2012), Dybdahl et al. (2012), Jónsdóttir et al. (2012), JRC (2013), Vuorinen et al. (2015), and Mansouri et al. (2016). See also the following paragraph.

B.1.1.2. Integrated approaches and models

47. Integrated approaches and models are now becoming commonly used. These combine HTS methods, human or non-human cell-based systems, model organism data and computational models, and may be used to replace testing or for data collection (Bell et al., 2017; Casey, 2016). The US EPA has developed a model for ER bioactivity that makes similar predictions to the ER binding, ER transactivation (ERTA) and Uterotrophic Assays (Browne et al., 2015) and a model that has been proposed as an alternative to the AR binding/AR transactivation assays (Kleinstreuer et al., 2016). The National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods have developed the Integrated Chemical Environment (ICE), which is a publically available web-based resource. ICE currently includes curated *in vivo* test data, reference chemical information, *in vitro* assay data (including ToxCast™ HTS data) and *in silico* model predictions. The ICE data integrator allows users to retrieve and combine datasets and to develop hypotheses through data exploration. Similarly, the US EPA EDSP21 Dashboard has been provided to help the Endocrine Disrupter Screening Program evaluate chemicals for endocrine-related activity, but is publically available.[l] The data for the dashboard comes from various sources: rapid, automated (or *in vitro* high-throughput) chemical screening data generated by the EPA's Toxicity Forecaster (ToxCast) project and the federal Toxicity Testing in the 21st century (Tox21) collaboration; chemical exposure data and prediction models (ExpoCastDB); high-quality chemical structures and annotations (DSSTox); physchem properties database (PhysChemDB) (all can be accessed through the EPA Chemistry Dashboard). It is also important to evaluate the relevance of metabolic activation (i.e. is the substance structure input into the model actually the one that cells will be exposed to?), and it would be helpful to ensure that a substance falls within the applicability domain of the model when determining the validity of any prediction. Additional guidance has also been developed and tested by the European Food and Safety Authority (EFSA) on the use and applications of such tools (EFSA, 2014).

B.1.1.3. Integrated approaches to testing and assessment

48. Integrated approaches to testing and assessment (IATA) are pragmatic, science-based approaches for chemical hazard or risk characterisation that rely on an integrated analysis of existing information in a WOE assessment coupled with the generation of new information using testing strategies. IATA follow an iterative approach to answer a defined question in a specific regulatory context, taking into account the acceptable level of uncertainty associated with the decision context. The OECD has an ongoing initiative to develop IATA; further information can be found in the "Guidance document for the use of adverse outcome pathways in developing integrated approaches to testing and assessment (IATA)" (OECD, 2016).

B.1.2. Conceptual Framework Level 2: **In vitro** *assays providing data about selected endocrine mechanism(s)/pathway(s)*

49. Assays at this level are screening assays used for hazard detection. Several assays are OECD performance-based test guidelines (PBTG) which describe the methodology for mechanistically and functionally similar test methods and facilitate the development of new, similar or modified test methods. At present, the hER binding assay (OECD TG 493) and the estrogen receptor transactivation assay (OECD TG 455) are PBTGs. PBTGs have the validated test methods annexed to the OECD TG. A separate document describing performance standards enables the development and validation of similar test methods for the same hazard endpoint to allow for timely amendment of the PBTG with new similar test methods. The similar test methods benefit from the mutual acceptance of data once they have been validated and accepted by the OECD.

50. These assays can provide identification of possible mechanisms and MOA, prediction of adverse outcome pathways (AOPs), priority-setting and WOE-based judgements leading to a conclusion. It is envisaged that a battery of *in vitro* tests would be carried out wherever possible because a single test will usually only provide information on one specific aspect of a modality. The results from a combination of tests will increase WOE.

51. Certain types of test data might be used to derive preliminary or more advanced judgements about a test chemical. Most *in vitro* assays can also provide "potency" data, e.g. binding data will provide a relative ranking of binding affinity based on a proposed scheme. The *in vitro* potency is not, however, predictive of *in vivo* potency in all cases. These assays are in most cases deliberately over-responsive (compared with many *in vivo* systems) towards chemicals that bind to a receptor as they are designed to provide alerts for endocrine activity. In other words, they will provide positives for some chemicals which give no *in vivo* responses, but are intended to minimise the risk that EASs will go undetected. It is noted that lack of metabolic systems in *in vitro* assays may lead to false negatives for chemicals which are bio-transformed to endocrine active metabolites but may potentially also lead to false positives for endocrine active chemicals which are very quickly transformed to endocrine inactive metabolites. Some cell-based assays for EASs do have metabolic capability (Combes, 2000; Jacobs et al., 2013) and it is important to establish whether or not this is the case when starting to use an assay.

52. Positive *in vitro* test results provide information about an endocrine mechanism/mode of action (MOA), and indicate the possibility of endocrine disruption effects *in vivo*. Current *in vitro* tests covered by the CF are largely based on mammalian systems, but their results can be used with caution to draw conclusions about possible EASs in other vertebrates, although potency and adverse consequences may differ. The activity of a chemical in a specific assay does not necessarily mean that it will cause toxicity or an adverse health outcome. There are many factors that determine whether a chemical will cause a specific adverse health outcome. Careful review is required to determine the use of the data in a particular decision context.

53. Negative *in vitro* results alone cannot be used to exclude possible endocrine activity because of their inherent limitations, such as inability or unknown capacity to metabolically activate toxicants. In addition, chemicals can interfere with the endocrine system in other ways than through the receptor, such as effects on the hypothalamic-pituitary-gonadal axis (HPG) that can only be detected in whole animal studies. For example, chemicals can interfere with the hormonal feedback loops in the HPG axis which could only be revealed in intact animals (e.g. by changes in hormone levels). Each *in vitro* assay measures a certain mechanism and thus conclusions can be drawn only in the context of what the *in vitro* assay evaluates. However, negative *in vitro* effects should only be interpreted as a tentative indication of a lack of endocrine activity for a specific aspect of the modality in question, if it can be substantiated that the compound does not undergo metabolic activation (e.g. by the use of ADME information).

54. *In vitro* screens can provide mechanistic data that are useful for the design of further *in vivo* studies. Again, cautious extrapolation to non-mammalian vertebrate *in vivo* tests is feasible.

55. *In vitro* screens are relevant for effects in humans and vertebrate wildlife because many are based on highly conserved hormone receptors or interaction with key enzymes or other key molecules involved in the regulation of hormone levels in all vertebrates. Chemicals that bind to these receptors or otherwise interfere with key processes of hormone regulation have the potential to cause effects in *in vivo* studies of vertebrate wildlife, assuming concentrations that reach the target are sufficiently high (e.g. dependent on ADME). Some *in vitro* screens are also available to detect juvenile hormone and ecdysteroid activity in arthropods (e.g. Dinan et al. [2001]; Smagghe et al. [2003]; Swevers et al. [2003]), but none of these have yet been standardised and validated internationally. However, an AOP now exists for ecdysone receptor agonism in arthropods (Song et al., 2017).

B.1.2.1 Possible sources of uncertainty and interference in in vitro *assays*

56. When using *in vitro* assays for regulatory purposes, possible sources of interference and factors causing variability need to be eliminated where possible. The use of proficiency chemicals and the requirements of validation processes have shaped the *in vitro* TGs and help them to be robust and reliable in practice. Nevertheless, there are many factors to be considered when conducting or evaluating these assays. A "Guidance document on good *in vitro* method practices (GIVIMP) for the development and implementation of *in vitro* methods for regulatory use in human safety assessment" has recently been drafted with this purpose (OECD, 2017b). The aim of this document is to reduce the uncertainties in cell- and tissue-based *in vitro* method derived predictions by applying all the necessary good scientific, technical and quality practices from *in vitro* method development to *in vitro* method implementation for regulatory use. Solubility of test substances, factors affecting

solubility, methods of determining solubility and recommendations for conducting assays with rather insoluble substances are also provided in the GIVIMP document (OECD, 2017b).

57. Possible sources of uncertainty may be interference from other receptors, e.g. the glucocorticoid receptor (GR) that may affect some AR transactivation assays (although interference is negligible in OECD TG 458). Alternatively, different cell types may express different isoforms (e.g. in TG 455 the ERα-HeLa-9903 cell line only expresses ERα whilst the VM7Luc4E2 cell line expresses ERα and ERβ). This may be advantageous or disadvantageous according to the assay's objective, but should be considered at the outset. Other sources of interference may be due to reporter gene product stabilisation or compound aggregation. Interfering factors in luciferase-based assays and in HTS have been reviewed by Thorne, Inglese and Auld (2010) and Thorne, Auld and Inglese (2010), who suggest practices that may reduce them. Hornung et al. (2017) have also reviewed artifacts in ER binding and agonist/antagonist assays and suggest ways to avoid false positives and optimise identification of true negatives. They suggest the use of endpoints such as toxicity, pH, precipitate formation, determination of inhibitor dissociation constants and the use of two different concentrations of estradiol tested in combination with graded concentrations of test chemical to distinguish true competitive antagonism from apparent antagonism. It is particularly important to exclude results obtained in the presence of cytotoxicity and precipitation. A detailed discussion on cytotoxicity, its measurement and the role it may play in disturbing the system can be found in OECD (2017b).

B.1.2.2 Metabolising systems in in vitro *assays*

58. Consideration should be given to the inclusion of metabolising systems in *in vitro* screens: see OECD (2008a) and Jacobs et al. (2008). It should be noted, however, that these systems are not applied on a regular basis with many *in vitro* assays (e.g. due to cytotoxicity). Some cell-based Level 2 assays may have limited metabolic capability and this may need to be assessed when setting up the assay. Another possible way of including metabolism is to carry out *in vitro* metabolism studies prior to the Level 2 assays. Identified metabolites or reaction mixture extracts containing metabolites could then be tested. It should be noted that *in vitro* metabolising systems may differ in some respects from *in vivo* systems, so their use still implies some uncertainty. The relative activities of different xenobiotic metabolising enzymes may differ *in vivo* and *in vitro* depending on availability of cofactors, stability of the enzymes or loss of subcellular compartments. However, many groups are now using metabolising systems and this area has recently been reviewed (Jacobs et al., 2013). Validation of AR and ER transactivation assays with metabolising systems added has recently been started via the OECD Validation Management Group for non-animal testing. The US EPA has also started projects addressing metabolic competence, using an "extracellular approach" where metabolism occurs in the media of cell-based and cell-free assays; or an "intracellular approach", where metabolism occurs inside the cell of cell-based assays.

B.1.3. Conceptual Framework Level 3: In vivo *assays providing data about selected endocrine mechanism(s)/pathway(s)*

59. Assays at Level 3 provide *in vivo* screening for **possible** endocrine disruption activity. In some cases they may also provide data on apical effects that could be caused by an endocrine MOA, but drawing sufficiently robust conclusions for regulatory decision making about possible adverse effects may not be possible, depending on the case and regulatory needs/requirements. They are designed to provide a yes/no (qualitative) answer

about the ability to interact with estrogen, androgen and thyroid hormone receptor mediated modalities, or interfere with steroidogenesis. Other non-receptor processes such as inhibition of iodination of thyroid hormones are also detected. It should be noted that although Level 3 (and 4) vertebrate assays do not generally expose organisms for a large proportion of their life cycle, and therefore are incapable of revealing the full spectrum of endocrine effects, experience to date suggests that they are sufficiently responsive to identify some EASs.

60. Assays at this level are screening assays designed primarily for hazard identification and for revealing mechanistic information. Some authorities may also seek to use them for taking regulatory decisions in some circumstances, but extrapolating from apical effects in screening tests to adverse effects may in some cases when evaluated with other data be sufficient for hazard assessment or for the identification of an ED, depending on the case and regulatory needs/requirements. These assays are designed to provide alerts to chemicals with possible endocrine disrupting properties, and detect alterations in endocrine-sensitive tissues. Therefore they are of deliberately high responsiveness (e.g. use in some cases of castrated/immature animal models without an intact or fully functional HPG axis, which are therefore unable to compensate fully for endocrine perturbations). In the case of immature animals, their responsiveness is comparable with the high sensitivity of some sensitive periods in the lifetime of higher mammals. The route of exposure may also not be representative of the natural situation, making direct extrapolation to the real world difficult (e.g. subcutaneous exposure in an assay when human exposure is dermal or oral).

61. They generally include the possibility for metabolic activation (albeit metabolism specific to rodents, fish or amphibians) of a chemical, a feature recommended for, but often absent from, current *in vitro* screens.

62. Assays are short in duration (e.g. the Uterotrophic [UT] and Hershberger [H] assays generally have three-day and ten-day dosing periods respectively whilst the Amphibian Metamorphosis Assay [AMA] and fish screens employ three weeks' aqueous exposure) and they generally only use very few (or a single) concentrations or dose levels. These assays also provide some information about the potency of a chemical *in vivo*, with respect to the magnitude of a change and the dose/concentration at which the change occurs.

63. It should be noted that both the 21-Day Fish Assay (OECD TG 230), the Fish Short-Term Reproduction Assay (FSTRA; OECD TG 229) and the AMA (OECD TG 231) are *in vivo* screens that primarily give information about endocrine disruption mechanisms in adult fish and/or larval amphibians. Additionally, OECD TG 229 includes apical endpoints (i.e. fecundity and by direct association also fertility) which can be affected both by some endocrine disrupting chemicals and some other chemicals toxic to reproduction. OECD TG 231 also contains an endpoint (amphibian metamorphosis) which could be considered as apical and potentially adverse, but the degree to which a given delay in metamorphosis is likely to be harmful to amphibian populations is case dependent. A delay of a few days may or may not have significant ecological consequences, depending on the biology and ecology of a given species, and should be carefully interpreted in the absence of longer term data, whereas a complete cessation of metamorphosis would have a major ecological impact. However, delayed metamorphosis may also be induced by other MOA than endocrine disruption. Hence, thyroid histopathology data may (depending on the availability of other relevant information) be needed for using an effect on metamorphosis in TG 231 to conclude that a chemical is an ED. Therefore, observations of delayed development (metamorphosis) in TG 231 may require long-term data obtainable from the

Larval Amphibian Growth and Development Assay (LAGDA; OECD TG 241) before a more definitive conclusion can be drawn about endocrine disruption.

64. A **positive** outcome (i.e. a statistically and biologically significant change(s) in an EAS-specific endpoint) of Level 3 assays indicates a possibility for adverse effects in the reproductive and developmental studies at Levels 4 and 5 and may in certain cases inform about effects in immature animals (which may be considered of concern). The specific criteria for a positive result in these assays are given in the "building blocks" in Section C but are generally significant changes in sex organ weight (UT and H assays), development (AMA), secondary sexual characteristics, and biomarkers such as vitellogenin or spiggin (fish screens). It should be noted that, depending on the guideline/protocol used, the fact that a substance may interact with a hormone system in these assays does not necessarily mean that when the substance is used it will cause adverse effects in humans or ecological systems. Conversely, apical endpoints in some of these assays (e.g. fecundity in the FSTRA and metamorphosis in the AMA) can provide evidence of adverse effects which may, in combination with mechanistic evidence, contribute to a conclusion that the test substance is an ED.

65. However, a compound found **negative** in Level 3 assays can be regarded as inactive against the specific modalities and life stages evaluated by those assays, but could still have endocrine disrupting properties mediated through other mechanisms or operative at more sensitive life stages (e.g. development or reproduction). These may be detected by a more comprehensive Level 4 or 5 assay than those *in vivo* screening assays covered by Level 3, although it is assumed selection of Level 3 assays is generally targeted on a previously suspected MOA.

66. The results from these *in vivo* screens can be used to decide if higher tier *in vivo* tests should be performed to reduce uncertainty about certain effects of EASs *in vivo* and to gain more information about potency. They may or may not provide data which can be used with confidence in human or vertebrate wildlife hazard identification/characterisation because the information does not always indicate whether, or to what extent, adverse effects on apical endpoints have occurred. Also, Level 3 screens do not encompass all possible modes by which E,A,T,S systems can be affected.

B.1.4. Conceptual Framework Level 4: **In vivo** *assays providing data on adverse effects on endocrine-relevant endpoints*

67. Assays at Level 4 can provide a more thorough assessment (in comparison with Level 3 assays) of the possible or actual endocrine disrupting effects and endocrine mechanism(s)/pathways of a chemical in developing or adult organisms because they are sensitive to more than one mode of endocrine disrupting action. A compound found to be positive indicates a possibility for adverse effects and which may require further investigation. However, if sufficient other data for decision making are available, further animal testing is not necessary. At this level, assays have numerous endpoints and therefore the criteria for a positive result are more complex than at lower levels, but generally a chemically induced, biologically significant change in an endocrine endpoint would be considered a positive result. A compound found to be negative is inactive under the specific conditions evaluated by the assay. A compound found negative in a Level 4 assay may still have endocrine disrupting properties either mediated through mechanisms not covered by the assay or because the assay was not sufficiently sensitive. Overall, a negative conclusion regarding endocrine disruption requires combined lines of evidence, if possible at various levels of the CF, e.g. Level 3 + Level 4 (or 5).

68. When conducting Level 4 and 5 tests it is important that the dose/concentration levels are high enough to detect relevant adverse effects. In the dose selection the investigator should also consider and ensure that data generated are adequate to fulfil the regulatory requirement across OECD countries as appropriate (e.g. hazard and risk assessment and labelling, ED assessment, etc.). In order to provide information relevant for assessing whether or not a chemical may fulfil the WHO/IPCS (2002) definition of an ED, the study design has to be sufficiently robust to demonstrate the presence or absence of effects. In the dose selection the investigator should also consider and ensure that data generated are adequate to fulfil the regulatory requirement across OECD countries as appropriate (e.g. hazard and risk assessment and labelling, ED assessment, etc.). The top dose or concentration should be sufficiently high to give clear systemic (i.e. non-endocrine specific) toxicity in order to ensure that a wide range of exposures (high to low) is tested. However, endocrine effects observed solely in the presence of clear systemic toxicity should be interpreted with caution and may be disregarded when sufficiently justified to be caused by secondary effects which are unlikely to be due to endocrine activity. The reason for this advice is a concern that some EAS-sensitive assays are being run at doses/concentrations of EASs that are too low to trigger direct impacts on the endocrine system. This GD is not the place to address this issue directly, but it should be considered when EAS-sensitive TGs are revised in the future. In addition, the number and spacing of dose/concentration levels should also be adequate to fulfil the objectives of the study (e.g. to demonstrate dose response relationships if this is required).

69. This level includes assays that are not specifically designed to detect EDs but have endpoints that are highly relevant for their detection. These assays include many standard repeated dose mammalian toxicology tests (e.g. OECD TG 407 [28-day Repeated Dose Toxicity Test] and OECD TG 408 [90-Day Repeated Dose Toxicity Test]). Most of these standard toxicology tests have not been validated for detection of EASs/EDs, with the exception of the 28-Day Repeated Dose Toxicity Test (OECD TG 407). This updated assay has been validated for some endocrine endpoints, but the sensitivity of the assay is not sufficient to identify all E,A,T,S-mediated EDs. The validation of the assay (OECD, 2006) showed that it identified strong and moderate EDs acting through the ER and AR; and EASs/EDs weakly and strongly affecting thyroid function. It was relatively insensitive to weak EASs/EDs acting through the ER and AR. It may also detect steroidogenesis inhibition although only one (potent) chemical was used in the validation study (CGS 18320B) (OECD, 2006). The 2017 version of the CF also includes standard repeated dose mammalian toxicology tests where administration is via dermal and inhalation routes and also where non-rodent mammalian test species are used. It was recognised that these assays also include some endocrine-sensitive endpoints. OECD TG 408 has recently been updated with endocrine endpoints (e.g. thyroid hormones and thyroid weight) and it is likely that the other repeat dose toxicity tests will also follow.

70. The reproduction/developmental screening tests OECD TG 421 and TG 422 are included in Level 4 as supplemental tests because they give limited but useful information on interaction with endocrine systems. EDs may be detected by effects on reproduction (gestation, gestation length, dystocia, implantation losses), genital malformations in offspring, changes in anogenital distance (AGD) in both sexes, and/or increased nipple retention in males, changes in histopathology of sex organs or effects on the thyroid hormonal system. These assays were updated in 2015 and 2016 to include more endocrine-sensitive endpoints, following a feasibility study (OECD, 2015). Other assays (e.g. OECD TG 414) are also being similarly updated.

71. Anogenital distance and nipple retention are sensitive endpoints of endocrine effects; however, their utility as apical endpoints or as biological indicators of endocrine action may require further experience in their use. Increased nipple retention and reduced AGD in male offspring are hallmarks of anti-androgenicity. Nevertheless, "retained nipples/areolae" as a qualitative endpoint may have high biological variability (e.g. Melching-Kollmuss et al., 2017) and alteration of AGD can occur via other MOA (e.g. Miyagawa et al., 2011; Seifert et al., 2009). However, current OECD guidance on these endpoints can be found in OECD GDs 43 and 151 and it is clear that these should be considered as apical endpoints. With regard to AGD, OECD GD 43 (OECD, 2008b) states: "A statistically significant change in AGD that cannot be explained by the size of the animal indicates effects of the exposure and should be used for setting the [no observed adverse effect level (NOAEL)]". With regard to nipple retention, OECD GD 151 (OECD, 2013) states: "a statistically significant change in nipple retention should be evaluated similarly to an effect on AGD as both endpoints indicate an adverse effect of exposure and should be considered in setting a NOAEL".

72. The feasibility report on OECD TG 421 and TG 422 (OECD, 2015) indicated that the sensitivity for detecting effects based on qualitative nipple retention (i.e. the number of males with or without nipples) was quite low irrespective of the number of litters included. However, nipple retention is a sensitive endpoint if measured quantitatively, i.e. if the number of nipples from 0 to 12 is recorded. This endpoint of quantitative nipple retention in the male pups was therefore included in these Level 4 study updates.

73. The one-generation assay (OECD TG 415) was also included at this level in earlier versions of the CF but this OECD TG has now been deleted as it has been made redundant following the introduction of the Extended One-Generation Reproductive Toxicity Study (EOGRTS) (OECD TG 443). This Level 5 assay provides a more thorough assessment of effects on reproduction and development than OECD TG 421/422.

74. The Prenatal Developmental Toxicity (OECD TG 414) and the Developmental Neurotoxicity (OECD TG 426) studies are also included in Level 4 as they involve repeated dosing of pregnant females and therefore potential exposure of the developing fetus. Both assays include some endpoints that may detect endocrine disruption (e.g. abnormalities of male and female genitalia in OECD TG 414). OECD TG 414 has also recently been updated with endocrine-sensitive endpoints (AGD and hormone levels in the dams), similar to OECD TG 421/422.

75. All assays at this level include apical endpoints and are designed for hazard identification/characterisation. The use of intact animal models provides an evaluation under normal physiological conditions but the responsiveness of these assays may be lower than Level 3 assays as hormone feedback mechanisms may provide some compensation in the case of EASs. Depending on the guideline/protocol used, the fact that a substance may interact with a hormone system in these assays does not necessarily mean that it will cause adverse effects in humans, e.g. the results for a chemical tested in the male or female pubertal assays with only two dose levels may not provide sufficient information on adverse effects. Interpretation may, however, be specific to regulatory authorities. For ecotoxicological tests, effects on apical endpoints at this level, such as fecundity, altered sex ratio and growth, are generally considered adverse because they are population relevant. Further investigation (e.g. conducting a relevant Level 5 assay that addresses effects on the next generation) may be required in order to determine if and how adverse effects observed at Level 4 may lead to adverse effects that are population relevant.

76. Experience with serum hormone determinations in Levels 4 and 5 rodent assays has revealed that their detection/measurement in rodent studies can be challenging. A recent workshop on "Practicability of Hormonal Measurements" was organised by the BfR (Germany) and the findings from this workshop will be published (Kucheryavenko et al., 2018). The OECD Expert Group on Reproductive and Developmental Toxicity recommends that to demonstrate proficiency for thyroid hormones measurement, a laboratory should be able to show results from a separate study using a positive control substance. Laboratories may also submit their calibration curves, standard curves, as well as data on the levels of quantification and detection. This group is also establishing a historical control database with thyroid toxicant positive controls. Level 4 assays may provide information about the potency of a compound which may be investigated further at Level 5, although some of these assays (e.g. the Fish Sexual Development Test and the Peripubertal Assays) may test relatively few concentrations or dose levels, thus limiting the precision of the results, and hence their usefulness for identifying a no-observed-effect-concentration/lowest-observed-effect-concentration/x% effect concentration (NOEC/LOEC/ECx) for all relevant types of adverse effects in environmental species. Effects on some endpoints included in the assays can, however, be considered as relevant adverse apical impacts on the (typically rather small) populations tested in the laboratory (e.g. major histopathologic changes in reproductive organs in rats; biased phenotypic sex ratios in developing fish), while others represent an effect on an indicator of hormonal activity for either humans or vertebrate wildlife (e.g. changes in thyroid hormone levels or vitellogenin titres).

77. Level 4 tests (e.g. the Fish Sexual Development Test or the 28-Day Repeated Dose Toxicity Test [OECD TG 407]) may also support an evaluation about whether specific endocrine-mediated effects may be influenced by general toxicity. This, of course, only applies if the tests have sufficient statistical power, test an appropriate range of concentrations and are conducted under conditions comparable to standard tests.

78. Some Level 4 assays (e.g. the Fish Sexual Development Test or the 28-Day Repeated Dose Toxicity Test [OECD TG 407]), but not all, can therefore provide data on adverse effects which may be sufficient for use in hazard identification/characterisation. Although most do not provide more comprehensive information about possible endocrine disrupting effects such as those obtainable from life cycle experiments (Level 5), they may often produce sufficiently robust data on adverse effects to obviate any need for Level 5 testing. In order to avoid unnecessary vertebrate testing, Level 5 tests should not be systematically conducted; rather, there should be a clear rationale based on available data collected at lower CF levels for requesting/performing Level 5 tests. This rationale should clarify why such a test is needed and how the information is intended to be used for the purpose of ED identification/characterisation. However, due to the low sensitivity of some Level 4 assays, a lack of endocrine-related adverse effects in one or more of them may not, depending on the case, remove a concern for ED activity raised by other available information.

B.1.5. Conceptual Framework Level 5: In vivo *assays providing more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism*

79. The developmental and reproductive toxicity studies at Level 5 provide data on adverse effects and endocrine mechanisms/pathways and are especially useful for hazard identification/characterisation as they add to the WOE concerning the potential for impacts in humans and vertebrate wildlife, and provide data on dose/concentration-response. The effects observed in reproductive tests with rodents, and in partial or full life cycle toxicity

studies with fish, amphibians and birds, may be due to endocrine disruption or other mechanisms, but the effect or pattern of effects (e.g. decreased AGD, increased nipple retention and malformations of reproductive organs in male rats) may indicate that effects mediated via impact on the endocrine system are involved. Some of these tests may also include measurement of endpoints which are indicative both of endocrine disruption activity and of adversity (e.g. altered sex ratio in the Medaka Extended One-Generation Reproduction Test [MEOGRT – OECD TG 240], alteration of puberty onset, or decrease in AGD or increase in nipple retention in male offspring in mammalian multigeneration tests).

80. Among the current OECD test guidelines for mammalian reproductive toxicity, exposure during all vulnerable periods of development is performed in the Extended One-Generation Reproductive Toxicity Study (EOGRTS) (OECD TG 443) and the Two-Generation Reproductive Toxicity Study design (OECD TG 416). The EOGRTS is the most sensitive assay for detection of endocrine disruption and this assay is preferred over the Two-Generation Reproductive Toxicity Study (OECD TG 416). However, if an adequate two-generation reproductive toxicity study is available, then an additional EOGRTS may depend on the case and/or regulatory needs and may not be required.

81. The EOGRTS (OECD TG 443) includes more endpoints sensitive to endocrine disruption than OECD TG 416 and it is expected that it will replace OECD TG 416 for mammalian reproductive toxicity testing. This test is also expected to have greater sensitivity than OECD TG 416 as it requires an increased number of pups to be examined. Endpoints sensitive to endocrine disruption include areola/nipple retention (PND 13), mandatory assessment of AGD at birth, measurement of thyroid hormones and TSH levels. Effects on the developing nervous and immune systems can also be assessed if the relevant cohorts are included in the study. These systems may also be sensitive to endocrine influences. Decisions on whether to produce and assess the F2 generation, omit the developmental neurotoxicity or developmental immunotoxicity have to be taken on a case-by-case basis depending on existing knowledge and regulatory purpose.

82. The Two-Generation Reproductive Toxicity Study (OECD TG 416) was updated in 2001 with endocrine disruption sensitive endpoints such as vaginal opening (VO), preputial separation (PPS), estrous cyclicity, evaluation of primordial follicle counts, AGD at postnatal day (PND) 0 (triggered by alterations in F1 sex ratio or timing of sexual maturation). This study provides information about endocrine disruption-relevant endpoints, particularly if combined with data from long-term repeat dosing studies, e.g. the 90-Day Repeated Dose Test (OECD TG 408) where the histopathology of the thyroid and mammary gland and possibly hormone data could be available. However, older reproductive toxicity studies that lack sensitive endpoints (e.g. onset of puberty) cannot fully exclude the possibility that chemicals tested negative may still be EDs. The updated OECD TG 416 does not include some endocrine disruption-related sensitive endpoints such as nipple retention, and anogenital distance is only investigated in the F2 generation if changes in sex ratio are observed in the F1 generation, which is not a particularly sensitive endpoint in respect of endocrine disruption. Thus, Two-Generation Reproductive Toxicity Studies (OECD TG 416) conducted before the inclusion of sensitive endocrine endpoints (e.g. sexual maturation) by themselves may not be considered adequate for demonstrating the probable absence of endocrine disrupting activity, although they still provide much valuable data (mainly restricted to fertility and effects on reproductive organs). In summary, the EOGRT study (OECD TG 443) is preferable for detecting endocrine disruption because it provides an evaluation of a number of endocrine endpoints in the

juvenile and adult F1 which are not included in the Two-Generation Reproductive Toxicity Study (OECD TG 416) adopted in 2001.

83. Late effects becoming manifest after weaning of the animals are partly covered in young adults, in OECD TG 416 and OECD TG 443, especially in relation to reproductive function, and to a more limited extent in relation to developmental neurotoxicity. However, effects on sexual dimorphism of the brain are not thoroughly investigated unless specific investigations are requested, for example in the developmental neurotoxicity (DNT) cohorts of the EOGRTS. The DNT cohort investigations may, according to OECD TG 443, if warranted, be supplemented with tests on memory and learning. Other potentially important late effects such as premature reproductive senescence (Cooper et al., 2007) are also not assessed. Effects becoming manifest during ageing are not included in any current guidelines for reproductive toxicity, but are being reviewed by the OECD. It is recognised that at the present time Level 5 assays do not cover all endocrine outcomes and this review should address these gaps.

84. In contrast, fish single- or multigeneration life cycle tests and the Avian Two-Generation Test (ATGT) include evaluation of exposure of many endocrine disruption-sensitive processes, and thus there is a higher level of confidence in the results. For multi-generation tests, the degree of confidence will be constrained by the statistical power of the test and the ability to control study conditions across multiple generations. This applies to the MEOGRT (OECD TG 240). While a recent publication (Flynn et al., 2017) evaluated nine studies that informed the development of the MEOGRT, there have been few completed tests that used the final MEOGRT guidelines, as published by the US EPA (890.2200) or the OECD (TG 240). The test guideline may be modified, if necessary, when more experience has been gained in its operation. The assay covers *inter alia* the possibility of detecting effects partly caused by the maternal transfer to offspring of certain EDs.

B.2. Endpoints in the various assays of the Conceptual Framework

85. In order to facilitate the interpretation of hazard data derived from screens and tests in the revised Conceptual Framework, Table B.1 presents a list of possible endpoints and their applicability for identifying endocrine disrupting modes of action and/or effects resulting from the four major modalities under consideration (i.e. estrogen-mediated activity, androgen-mediated activity, thyroid-related activity and steroidogenesis disruption related-activity). It should be borne in mind that agonism/antagonism and thereby the terms "estrogenic"/"anti-estrogenic", "androgenic"/"anti-androgenic" used in Table B.1, and throughout the document, are context-dependent (i.e. dependent on dose, life stage, tissue, etc.) and may have various meanings. When using these terms, it is recommended to consider whether they describe a molecular initiating event, one or more key events of an AOP, or one or more of the adverse outcomes (AO) of an AOP – or the whole AOP. In addition, effects resulting from interference with juvenile hormone in non-mammalian assays are included. Effects on the retinoid system have been included in Table B.1 as a recent draft "Detailed review paper on the retinoid system" (OECD, 2017a) indicates that many endpoints sensitive to E,A,T,S modalities may also be affected by substances acting on the retinoid system in developing animals. Other endocrine MOA (e.g. in DRP No. 178; OECD, 2012a) may also affect these endpoints. A recent publication has evaluated endpoints in existing regulatory tests with respect to their ability to provide diagnostic information on E,A,T,S modalities and several other endocrine axes such as the

hypothalamus/pituitary/adrenal axis, somatotropic axis and vitamin D-signaling (Manibusan and Touart, 2017).

86. Where possible, the direction of change is indicated for the endpoints in Table B.1. Care should be taken, however, when information in Table B.1 is used to interpret observations of effects induced by specific substances *in vivo*. The data from validation studies on the assays has been used to guide the changes as much as possible, but in some cases it has not been possible to generalise and in other cases extrapolations have been made across similar endpoints in different studies (e.g. OECD TG 416 has not been validated for thyroid-related activities but it is reasonable to suppose that thyroid changes in OECD TG 416 would be similar to those seen in the OECD TG 407 and the pubertal assays). In all cases, the direction of change is illustrative and not all possibilities are given (e.g. for steroidogenesis disruption, only inhibition of steroidogenic enzymes is illustrated, reflecting the chemicals used in validation studies whereas in theory induction may be possible). Specific chemicals may induce a range of effects *in vivo* which cannot be clearly assigned to only one endocrine MOA. There may be good biological reasons for this, including that many chemicals act through multiple MOA. Even the reference chemicals used in validation studies are recognised in many cases to have more than one MOA, and therefore the effects on endpoints should be taken as indicative rather than definitive. Table B.1 also lists those endpoints which may not be directly linked to E,A,T,S-related mechanisms.

87. Endpoints for hormonal-mediated activity and endpoints potentially sensitive to, but not diagnostic of, hormonal-mediated activity listed in Table B.1 can be affected by a variety of non-endocrine factors, such as marked systemic toxicity, handling stress or infections (e.g. Dang, 2014). In the context of infections, it should be noted that pathogens and parasites can lead to systemic toxicity, but also very specific interactions with the endocrine system have been reported in invertebrates (Morley, 2006; Rodgers-Gray et al., 2004) as well as in vertebrates (Larralde et al., 1995; Sitjà-Bobadilla, 2009; Trubiroha et al., 2010). Furthermore, care should be taken to avoid diets or caging materials which can be sources of endocrine activity (Beresford et al., 2011; Thigpen et al., 2013). It is important to consider possible confounding factors and use a WOE approach when interpreting changes in a single study or a battery of studies. Changes in endpoints should not be evaluated in isolation without any other corroborating evidence of an endocrine MOA of the test item.

88. Changes in endpoints may depend on factors such as dose, tissue, life stage and the endogenous hormone levels. For example, in a life stage where endogenous serum estrogen levels are low, a weakly acting "estrogenic" xenobiotic may cause agonistic effects because binding to the unoccupied ERs causes their activation. In another life stage where the serum estrogen is relatively high but some ERs are not occupied, it may at low dose also be agonistic because it binds to the unoccupied ERs and causes activation. At higher doses, however, where all ERs are occupied by either endogenous estrogen or the xenobiotic estrogen, it may act antagonistically. In this case, the xenobiotic "estrogen" may compete with and replace the receptor-bound endogenous estrogen, so that the normal endogenous estrogen activation is weakened by ER binding a molecule with lower potency than endogenous estrogen. This issue has implications for both the interpretation of test results and for how those results are generalised in respect of possible *in vivo* situations that the test results should inform about. Therefore, care should be taken when conclusions are drawn about agonistic and antagonistic MOA because often such conclusions are oversimplifications of what may happen *in vivo*.

89. The endpoints listed in Table B.1 are those specified in the guideline, or those most commonly used in an assay, for methods for which no guidelines are available. Specific data transformations, e.g. anogenital distance expressed as cubic root of body weight (as calculated in Gallavan et al. [1999]) are not shown in Table B.1 but are specified in the relevant TGs. Other endpoints may be added, particularly changes in titres of hormones such as estradiol, testosterone, luteinising hormone (LH), follicle stimulating hormone (FSH), etc., and are frequently added to OECD TG 407 and OECD TG 412 for example. Several of the OECD TGs for developmental and reproductive toxicity have recently been updated, with others in progress. Beekhuijzen et al. (2016) suggest practical considerations for these updates, based on experiences within one laboratory.

90. However, it should be noted that several assays with non-mammalian wildlife species (especially the Larval Amphibian Growth and Development Assay [LAGDA], the Avian Reproduction Test, and the MEOGRT) and the CF Levels 4 and 5 mammalian assays are not solely designed to detect the effects of endocrine disrupters, but they are expected to be sensitive to many such chemicals, as well as to other reproductively toxic materials. Furthermore, many of these assays with non-mammalian wildlife species are still in development, so a full description of their reactions to the types of EASs under consideration here cannot yet be given. Finally, it is important to note that although a number of invertebrate assays with apical endpoints have been included in this document, these assays rarely provide information on MOA and they may also respond to non-EASs. As yet, the OECD has not standardised any mechanistic *in vitro* assays for MOA which occur in invertebrates. This implies that it may be currently impossible to conclude whether a substance is an ED in these phyla, although non-standardised *in vitro* assays are available for some MOA in invertebrates (e.g. ecdysteroid and juvenile hormone activity in arthropods). Similar, though less severe, issues arise with avian multigeneration data because of a relative lack of understanding of endocrinology in birds.

Table B.1. **Endpoints relevant for endocrine disruption modalities in test guidelines and other EAS-sensitive assays (in the Conceptual Framework) for which guidance with interpretation of data has been developed**

Probable direction of change is indicated where possible**.** However, in all cases, the direction of change is illustrative and not all possibilities are given, e.g. for steroidogenesis disruption, only inhibition of steroidogenic enzymes is illustrated, reflecting the chemicals used in validation studies, whereas induction may also be possible.

Note that for many assays, individual endpoints may not in themselves be diagnostic of an endocrine disruption modality. Such diagnosis often relies on a combination of endpoints or assays in a weight of evidence assessment. The term diagnostic does not refer to clinical diagnosis, but rather to conclusive evidence.

The symbol "?" in this table indicates a lack of knowledge about whether the modality causes a response in the respective assay.

Test guideline or other test method (reference to interpretation table within this document)	Endpoints for estrogen-mediated activity		Endpoints for androgen-mediated activity		Endpoints for thyroid-related activity	Endpoints for steroidogenesis-related activity	Endpoints for juvenile hormone-related activity	Endpoints potentially sensitive to, but not diagnostic of, E,A,T,S, juvenile hormone, ecdysone or retinoid modalities
	Agonistic	Antagonistic	Agonistic	Antagonistic				
***In vitro* screens (CF Level 2)**								
A. OECD test guidelines with endocrine active substance-specific endpoints or with non-specific sensitivity to endocrine active substances								
OECD TG 493: Performance-Based Test Guideline for Human Recombinant Estrogen Receptor (hrER) *In Vitro* Assays to Detect Chemicals with ER Binding Affinity (Table C.1.1)	Displacement of ligand from receptor. Binding cannot distinguish between agonism or antagonism.		Nil	Nil	Nil	Nil	Nil	Nil
OECD TG 455: Performance-Based Test Guideline for Stably Transfected Transactivation *In Vitro* Assays to Detect Estrogen Receptor Agonists and Antagonists (Table C.1.2)	Activation of reporter gene linked to estrogen receptor (ER). ER agonists may also inhibit if they can compete with the activating ligand.	Inhibition of activation of reporter gene linked to ER. ER agonists may also inhibit if they can compete with the activating ligand.	Nil	Nil	Nil	Nil	Nil	Activators of the aryl hydrocarbon receptor may inhibit activation of reporter gene linked to ER through crosstalk at the DNA level.
OECD TG 458: Stably Transfected Human Androgen Receptor Transcriptional Activation Assay for Detection of Androgenic Agonist and Antagonist Activity of Chemicals (Table C.1.3)	Nil	Nil	Activation of reporter gene linked to AR.	Inhibition of activation of reporter gene linked to AR.	Nil	Nil	Nil	Nil

Test guideline or other test method (reference to interpretation table within this document)	Endpoints for estrogen-mediated activity		Endpoints for androgen-mediated activity		Endpoints for thyroid-related activity	Endpoints for steroidogenesis-related activity	Endpoints for juvenile hormone-related activity	Endpoints potentially sensitive to, but not diagnostic of, E,A,T,S, juvenile hormone, ecdysone or retinoid modalities
	Agonistic	Antagonistic	Agonistic	Antagonistic				
OECD TG 456: H295R Steroidogenesis Assay (Table C.1.4)	Nil	Nil	Nil	Nil	Nil	Inhibition and/or induction of estradiol and testosterone synthesis.	Nil	Nil
B. Guidelines that have not received full validation by the OECD, or are in the process of OECD validation, or which have been validated and published by other organisations								
AR Binding Assay (US EPA OPPTS 890.1150) (Table C.1.5)	Nil	Nil	Displacement of ligand from receptor. Binding cannot distinguish between agonism or antagonism.		Nil	Nil	Nil	Nil
Aromatase Assay (US EPA OPPTS 890.1200) (Table C.1.6)	Nil	Nil	Nil	Nil	Nil	Inhibition of aromatase (CYP 19) activity.	Nil	Nil
Non-mammalian *in vivo* screens and tests (CF Levels 3-5)								
A. OECD test guidelines with endocrine active substance-specific endpoints or with non-specific sensitivity to endocrine active substances								
OECD TG 229: Fish Short-Term Reproduction Assay (FSTRA) (Table C.2.1)	Vitellogenin (VTG) induction in males or females. Depression of male 2º sex characteristics in fathead minnow or medaka. Specific gonad histopathologic findings as listed in OECD (2010a).[3]	VTG depression in females (assuming no systemic toxicity). Specific gonad histopathologic findings as listed in OECD (2010a).[3]	Induction of male 2º sex characteristics in female fathead minnow or medaka. Specific gonad histopathologic findings as listed in OECD (2010a).[3] Possible VTG depression in females.	Depression of male 2º sex characteristics in fathead minnow or medaka. Specific gonad histopathologic findings as listed in OECD (2010a).[3]	Nil	Possible effects on: – VTG depression in females (assuming no systemic toxicity) – gonad histo-pathology (e.g. Leydig cell hyperplasia; see OECD 2010a).[3]	Nil	E,A,T,S and/or other activity can affect the following: – fecundity depression – certain histopathologic findings not related to endocrine activity – behaviour.
OECD TG 230: 21-Day Fish Assay (Table C.2.2)	VTG induction in males or females. Depression of male 2º sex characteristics in fathead minnow or medaka.	VTG depression in females (assuming no systemic toxicity).	Induction of male 2º sex characteristics in female fathead minnow or medaka. Possible VTG depression in females.	Depression of male 2º sex characteristics in fathead minnow or medaka.	Nil	Possible effects on: VTG depression in females (assuming no systemic toxicity).	Nil	E,A,T,S and/or other activity can affect the following: – behaviour – certain histopathologic findings (see OECD [2010a]).

Test guideline or other test method (reference to interpretation table within this document)	Endpoints for estrogen-mediated activity		Endpoints for androgen-mediated activity		Endpoints for thyroid-related activity	Endpoints for steroidogenesis-related activity	Endpoints for juvenile hormone-related activity	Endpoints potentially sensitive to, but not diagnostic of, E,A,T,S, juvenile hormone, ecdysone or retinoid modalities
	Agonistic	Antagonistic	Agonistic	Antagonistic				
OECD TG 231:Amphibian Metamorphosis Assay (AMA) (Table C.2.3)	Nil	Nil	Nil	Nil	Developmental stage.[2] Hind limb length.[2] Snout-vent length.[2] Thyroid gland histopathology. Time to metamorphosis. (see OECD TG 231 for interpretation of combined effects – individual changes may not be diagnostic).	Nil	Nil	E,A,T,S modalities can affect: – body weight.
OECD TG 242: *Potamopyrgus antipodarum* Reproduction Test (Table C.2.4)	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Retinoid X receptor (RXR) agonists and various other activities may affect embryo production.
OECD TG 243: *Lymnaea stagnalis* Reproduction Test (Table C.2.5)	Nil	Nil	Nil	Nil	Nil	Nil	Nil	RXR agonists and various other activities may affect fecundity.
OECD TG 218-219: Chironomid Toxicity Test (Table C.2.6)	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Juvenile hormone or ecdysteroid agonists and antagonists can interfere with metamorphosis, moulting, time to emergence and growth.
OECD TG 211: *Daphnia* Reproduction Test (with Male Induction) (Table C.2.7)	Nil	Nil	Nil	Nil	Nil	Nil	Production of male neonates, but note that various natural stressors (e.g. starvation) can also lead to male neonate production.	Juvenile hormone or ecdysteroid agonists and antagonists can interfere with moulting and growth.
OECD TG 210: Fish Early Life Stage Toxicity Test (Table C.2.8)	Nil	Nil	Nil	Nil	Some thyroid-active chemicals may interfere with embryonic development and metamorphosis, but this is not diagnostic for thyroid activity.	Nil	Nil	Nil

Test guideline or other test method (reference to interpretation table within this document)	Endpoints for estrogen-mediated activity		Endpoints for androgen-mediated activity		Endpoints for thyroid-related activity	Endpoints for steroidogenesis-related activity	Endpoints for juvenile hormone-related activity	Endpoints potentially sensitive to, but not diagnostic of, E,A,T,S, juvenile hormone, ecdysone or retinoid modalities
	Agonistic	Antagonistic	Agonistic	Antagonistic				
OECD TG 234: Fish Sexual Development Test (FSDT) (Table C.2.9)	Female-biased phenotypic sex ratio.[1] VTG induction in males and females. Specific gonad histopathologic findings (optional) as listed in OECD (2010a).[3]	Male-biased phenotypic sex ratio.[1] Increase in sexually undifferentiated fish. VTG depression in females, assuming no systemic toxicity. Specific gonad histopathologic findings (optional) as listed in OECD (2010a).[3]	Male-biased phenotypic sex ratio.[1] Possible VTG depression in females, assuming no systemic toxicity. Specific gonad histopathologic findings (optional) as listed in OECD (2010a).[3]	Induction of intersex fish. Female-biased phenotypic sex ratio.[1] Specific gonad histopathologic findings (optional) as listed in OECD (2010a).[3]	Some thyroid-active chemicals may interfere with embryonic development and metamorphosis, but this is not diagnostic for thyroid activity.	Possible effects on: – male-biased phenotypic sex ratio[1] – VTG depression in females, assuming no systemic toxicity.	Nil	E,A,T,S and/or other activity can affect the following: – body length – body weight – morphological abnormalities – anormal behaviour – certain histopathologic findings not related to endocrine activity.
OECD TG 241: Larval Amphibian Growth and Development Assay (LAGDA) (Table C.2.10)	Feminisation of testes. Induction of vitellogenin in males. Female bias in sex ratio.	?	Masculinisation of ovaries. Reduction of vitellogenin in females (assuming no systemic toxicity). Male bias in sex ratio.	?	Depending on the type of interference with the HPT axis, changes in the following: – thyroid histopathology – time to metamorphosis.	?	Nil	E,A,T,S modalities can affect: – mortality – behaviour – growth.
OECD TG 206: Avian Reproduction Test *Note:* No endpoints specific to a particular endocrine disruption modality are included at present but diagnostic endpoints could be added (e.g. vitellogenin). (Table C.2.11)	Nil	Nil	Nil	Nil	Nil	Nil	Nil	E,A,T,S modalities can affect: – egg production – cracked eggs – eggshell thickness – egg viability – hatchability – body weight – gross pathology.
OECD TG 233: Sediment Water Chironomid Life Cycle Toxicity Test (Table C.2.12)	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Juvenile hormone or ecdysteroid agonists and antagonists can interfere with metamorphosis, moulting, growth and/or reproduction.

Test guideline or other test method (reference to interpretation table within this document)	Endpoints for estrogen-mediated activity		Endpoints for androgen-mediated activity		Endpoints for thyroid-related activity	Endpoints for steroidogenesis-related activity	Endpoints for juvenile hormone-related activity	Endpoints potentially sensitive to, but not diagnostic of, E,A,T,S, juvenile hormone, ecdysone or retinoid modalities
	Agonistic	Antagonistic	Agonistic	Antagonistic				
OECD TG 240: Medaka Extended One-Generation Reproduction Test (MEOGRT) (Table C.2.13)	Female-biased phenotypic sex ratio (phenotypic sex compared with genetic sex). VTG induction in males and females. Specific gonad histopathology as listed in OECD (2010a).[3]	Male-biased phenotypic sex ratio. VTG depression in females (assuming no systemic toxicity). Increase in sexually undifferentiated fish. Specific gonad histopathology as listed in OECD (2010a).[3]	Male-biased phenotypic sex ratio. Possible VTG depression in females (assuming no systemic toxicity). Induction of male secondary sexual characteristics (anal fin papillae) in females.	Female-biased phenotypic sex ratio. Induction of intersex fish. Reduction in number of anal fin papillary processes in males. Specific gonad histopathology as listed in OECD (2010a).[3]	Some thyroid-active chemicals may interfere with embryonic development and metamorphosis.	Male-biased phenotypic sex ratio. VTG depression in females (assuming no systemic toxicity). Reduction in number of anal fin papillary processes in males (for substances interfering with androgen biosynthesis).	Nil	E,A,T,S modalities can affect: – hatching success – weight – length – behaviour – gross morphology – gonado-somatic index – multiple organ histopathology – time to maturity (time to first spawn) – fecundity – fertilisation success.
B. Guidelines that have not received full validation by the OECD, or are in the process of OECD validation, or which have been validated and published by other organisations								
Draft OECD TG SJHASA: Short-Term Juvenile Hormone Activity Screening Assay Using *Daphnia magna* (Table C.2.14)	Nil	Nil	Nil	Nil	Nil	Nil	Production of male neonates, but note that various natural stressors (e.g. starvation) can also lead to male neonate production.	Juvenile hormone or ecdysteroid agonists and antagonists can interfere with moulting and growth.
OECD GD 148: Androgenised Female Stickleback Screen (AFSS) (Table C.2.15)	Nil	Nil	Spiggin induction.	Spiggin depression.	Nil	Nil	Nil	Nil
Draft OECD TG EASZY Assay: Detection of Substances Acting through Estrogen Receptors using Transgenic cyp19a1b GFP Zebrafish Embryos (Table C.2.16)	Induction of green fluorescent protein (GFP).	Nil	Nil	Nil	Nil	Nil	Nil	Nil
Draft OECD TG JMASA: Juvenile Medaka Anti-Androgen Screening Assay (Table C.2.17)	Nil	Nil	Nil	Reduction in number of anal fin papillary processes in males.	Nil	Reduction in number of anal fin papillary processes in males (for substances interfering with androgen biosynthesis).	Nil	Nil

Test guideline or other test method (reference to interpretation table within this document)	Endpoints for estrogen-mediated activity		Endpoints for androgen-mediated activity		Endpoints for thyroid-related activity	Endpoints for steroidogenesis-related activity	Endpoints for juvenile hormone-related activity	Endpoints potentially sensitive to, but not diagnostic of, E,A,T,S, juvenile hormone, ecdysone or retinoid modalities
	Agonistic	Antagonistic	Agonistic	Antagonistic				
Draft OECD TG RADAR: Rapid Androgen Disruption Adverse Outcome Reporter Assay (Table C.2.24)	Nil	Nil	Increase in GFP.	Decrease in GFP during simultaneous exposure to an androgen.	Nil	Aromatase inhibition can lead to accumulation of testosterone which could cause an increase in GFP.	Nil	Nil
Draft OECD TG XETA: *Xenopus* Embryonic Thyroid Signalling Assay (Table C.2.18)	Nil	Nil	Nil	Nil	Increase or decrease in GFP, depending on precise mode of thyroid activity.	Nil	Nil	Nil
OECD GD 201: New Guidance Document on Harpacticoid Copepod Development and Reproduction Test with *Amphiascus* (Table C.2.19)	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Juvenile hormone or ecdysteroid agonists and antagonists can interfere with metamorphosis, moulting, growth and/or reproduction.
Draft OECD TG DMGT: *Daphnia* Multigeneration Test for Assessment of EDCs (Table C.2.20)	Nil	Nil	Nil	Nil	Nil	Nil	Induction of male neonates, but note that various natural stressors (e.g. starvation) can also lead to male neonate production.	Juvenile hormone or ecdysteroid agonists and antagonists can interfere with moulting and growth.
Fish Life Cycle Toxicity Test (FLCTT) (US EPA OPPTS 850.1500, possibly with endocrine-sensitive additions) *Note:* No endpoints specific to a particular E,A,T,S modality are included at present but endpoints indicative of endocrine activity could be added if validated. (Table C.2.21)	Female-biased phenotypic sex ratio.[1] VTG induction in males.	?	Male-biased phenotypic sex ratio.*	?		Possible effects on: – VTG depression in females, if no systemic toxicity.	Nil	E,A,T,S modalities can affect: – hatching success – weight – length – behaviour – gross morphology – gonado-somatic index – multiple organ histopathology – time to maturity (time to first spawn) – fecundity – fertilisation success.

Test guideline or other test method (reference to interpretation table within this document)	Endpoints for estrogen-mediated activity		Endpoints for androgen-mediated activity		Endpoints for thyroid-related activity	Endpoints for steroidogenesis-related activity	Endpoints for juvenile hormone-related activity	Endpoints potentially sensitive to, but not diagnostic of, E,A,T,S, juvenile hormone, ecdysone or retinoid modalities
	Agonistic	Antagonistic	Agonistic	Antagonistic				
Draft OECD TG ZEOGRT: Zebrafish Extended One-Generation Reproduction Test (Table C.2.22)	Female-biased phenotypic sex ratio. VTG induction in males and females. Specific gonad histopathology as listed in OECD (2010a).[3]	Male-biased phenotypic sex ratio. VTG reduction in females (assuming no systemic toxicity). Increase in sexually undifferentiated fish. Specific gonad histopathology as listed in OECD (2010a).[3]	Male-biased phenotypic sex ratio. VTG reduction in females (assuming no systemic toxicity). Specific gonad histopathology as listed in OECD (2010a).[3]	Female-biased phenotypic sex ratio. VTG induction in females. Induction of intersex fish. Specific gonad histopathology as listed in OECD (2010a).[3]	Some thyroid-active chemicals may interfere with embryonic development and metamorphosis.	Male-biased phenotypic sex ratio. VTG reduction in females (assuming no systemic toxicity).	Nil	E,A,T,S modalities can affect: – time to hatching – hatching success – weight – length – behaviour – gross morphology – gonado-somatic index – multiple organ histopathology – time to maturity (time to first spawn) – fecundity – fertilisation success.
US EPA OCSPP 890.2100/740-C-15-003 ATGT: Avian Two-Generation Toxicity Test in the Japanese Quail (Table C.2.23)	Phenotypic and genotypic sex ratio. Gonad histopathology. Estradiol and testosterone titres.	Phenotypic and genotypic sex ratio. Gonad histopathology. Estradiol and testosterone titres.	Phenotypic and genotypic sex ratio. Gonad histopathology. Estradiol and testosterone titres.	Phenotypic and genotypic sex ratio. Gonad histopathology. Estradiol and testosterone titres.	T3/T4	?	Nil	E,A,T,S modalities can affect: – mortality – growth – fecundity – fertility – time to sexual maturity – shell thickness – shell breakage – hatching success – gross morphology.
Mammalian *in vivo* screens and tests (CF Levels 3-5)								
A. OECD test guidelines with endocrine active substance-specific endpoints or with non-specific sensitivity to endocrine active substances								
OECD TG 440: Uterotrophic Bioassay in Rodents (UT assay) (including OECD GD 71 for Antiestrogenicity Screen) (immature female or adult after ovariectomy) (Table C.3.1)	Uterine weight (wet and blotted) increase. Optional: keratinisation and cornification of vagina, proliferation of endometrial epithelium, changes in uterine histopathology.	Reduction of estrogen-stimulated uterine weight increase. *Note:* TG does not include antagonist determination, which is described in OECD GD 71. Optional: reduction of other estrogen-stimulated histopathologic changes.	Uterine weight (wet and blotted) increase. (Aromatisable) androgens can increase uterine weight in both immature and ovariectomised female rats.	Nil	Nil	Nil	Nil	The immature rodent assay where the hypothalamic/pituitary/gonadal axis is intact, may detect other modes of action (e.g. related to GnRH inhibition).

Test guideline or other test method (reference to interpretation table within this document)	Endpoints for estrogen-mediated activity		Endpoints for androgen-mediated activity		Endpoints for thyroid-related activity	Endpoints for steroidogenesis-related activity	Endpoints for juvenile hormone-related activity	Endpoints potentially sensitive to, but not diagnostic of, E,A,T,S, juvenile hormone, ecdysone or retinoid modalities
	Agonistic	Antagonistic	Agonistic	Antagonistic				
OECD TG 441: Hershberger Bioassay (H assay) (Adult Male after Castration) (including OECD GD 115 for Weanling Hershberger Bioassay) (Table C.3.2)	Nil	Nil	Increase in weight of ventral prostate, seminal vesicles, levator ani plus bulbocavernosus muscle complex (LABC), cowpers glands, glans penis (+ve outcome if 2 or more tissues are increased). *Note in the weanling H assay:* glans penis is not included, testis weight is decreased. Optional: changes in serum hormones.	Reduction of androgen-stimulated weights of ventral prostate, seminal vesicles, LABC, cowpers glands, glans penis (+ve outcome if 2 or more tissues are decreased). *Note in the weanling H assay:* glans penis is not included, testis weight is increased. Optional: changes in serum hormones.	Optional: Possible liver weight increase (in combination with other thyroid-related endpoints). Reduction in serum T4 and T3 (anti-thyroid). Agonistic changes are opposite.	Nil	Nil	Optional: Adrenal weight.
OECD TG 407: Repeated Dose 28-Day Oral Toxicity Study in Rodents (Table C.3.3)	Histopathologic changes in ovary, uterus/cervix, vagina. Decrease in weight of epididymides, prostate + seminal vesicles with coagulating glands. Histopathologic changes in testes, epididymides, prostate + seminal vesicles with coagulating glands. Other endpoints: – increase in weight of uterus (slight), decrease in weight of ovaries – changes in estrous cyclicity – histopathologic (proliferative) changes in mammary glands (males).	Changes may occur in the following: Histopathologic changes in ovary, uterus/cervix, vagina. Increase in weight of epididymides, prostate + seminal vesicles with coagulating glands. Histopathologic changes in testes, epididymides, prostate + seminal vesicles with coagulating glands Other endpoints: – uterine/ovary weight decrease – changes in estrous cyclicity – histopathologic changes in mammary glands.	Histopathologic changes in ovary, uterus/cervix, vagina. Increase in weight of prostate + seminal vesicles with coagulating glands. Decrease in weight of testes. Histopathologic changes in testes, epididymides, Other endpoints: – ovary weight (decrease) – changes in estrous cyclicity – histopathologic changes in mammary glands.	Decrease in weight of epididymides, prostate + seminal vesicles with coagulating glands. Histopathologic changes in testes, epididymides, prostate + seminal vesicles with coagulating glands. Other endpoints: – ovary weight (decrease).	Possible liver weight increase (in combination with other thyroid-related endpoints). Histopathologic changes in thyroid (follicular cell height increase and colloid area decrease). Other endpoints: – serum T3 and T4 decreased, TSH increased – increased thyroid weight (anti-thyroid) – agonistic changes are opposite.	Possible effects on: – histopathologic changes in ovary, uterus/cervix, vagina – weight of prostate + seminal vesicles with coagulating glands. Other endpoints: – uterine and ovary weight – changes in vaginal smears – histopathologic changes in mammary gland. Note that changes depend on nature of interference (e.g. inhibition of estrogen synthesis results in antiestrogen-like effects on endpoints).	Nil	Changes in adrenal weight. Histopathologic changes in adrenal. Other endpoints: Histopathologic changes in pituitary and mammary glands.

Test guideline or other test method (reference to interpretation table within this document)	Endpoints for estrogen-mediated activity		Endpoints for androgen-mediated activity		Endpoints for thyroid-related activity	Endpoints for steroidogenesis-related activity	Endpoints for juvenile hormone-related activity	Endpoints potentially sensitive to, but not diagnostic of, E,A,T,S, juvenile hormone, ecdysone or retinoid modalities
	Agonistic	Antagonistic	Agonistic	Antagonistic				
OECD TG 408: Repeated Dose 90-Day Oral Toxicity Study (Table C.3.4)	Increased uterus weight, decreased ovary weight. Histopathologic changes in ovary, uterus/cervix, vagina and female mammary gland. Decrease in weight of epididymides, prostate + seminal vesicles with coagulating glands. Histopathologic changes in testes, epididymides, prostate + seminal vesicles with coagulating glands. Histopathologic changes in male mammary gland. Optional endpoints: Changes in estrous cyclicity. Changes in serum hormones. Reductions in sperm parameters: sperm numbers, sperm motility, sperm morphology.	Changes may occur in the following: – uterus and ovary weight (decrease) – histopathologic changes in ovary, uterus/cervix, vagina and female mammary gland – testes and epididymides weights (increase) – histopathologic changes in testes, epididymides, male accessory sex organs and male mammary gland. Optional endpoints: Changes in estrous cyclicity. Changes in serum hormones. Changes in sperm parameters: sperm numbers, sperm motility, sperm morphology.	Decreased ovary weight. Histopathologic changes in ovary, uterus/cervix, vagina. Increased weight of epididymides, decreased testes weight. Histopathologic changes in testes, epididymides, male accessory sex organs. Optional endpoints: Changes in estrous cyclicity. Changes in serum hormones. Changes in sperm parameters: sperm numbers, sperm motility, sperm morphology.	Histopathologic changes in ovary, uterus/cervix, vagina. Decreased weight of epididymides, increased testes weight. Histopathologic changes in testes, epididymides, male accessory sex organs. Optional endpoints: Changes in estrous cyclicity. Changes in serum hormones. Changes in sperm parameters: sperm numbers, sperm motility, sperm morphology.	Possible liver weight increase (in combination with other thyroid-related endpoints). Serum T4, T3 decreased, TSH increased. Histopathologic changes in thyroid gland. (Anti-thyroid changes, agonistic changes are opposite). Changes to HDL/LDL ratio (in combination with other thyroid-related endpoints).	Possible effects on: – male accessory sex organs and male mammary gland. Optional endpoints show possible changes in: estrous cyclicity, serum hormones. Changes in sperm parameters: sperm numbers, sperm motility, sperm morphology. Note that changes depend on nature of interference (e.g. inhibition of estrogen synthesis results in antiestrogen-like effects on endpoints).	Nil	Changes in adrenal weight. Histopathologic changes in adrenal, and pituitary glands.
OECD TG 451-3: Combined Chronic Toxicity/ Carcinogenicity Studies (Table C.3.5)	Increased uterus weight, decreased ovary weight. Histopathologic changes in ovary, uterus/cervix, vagina and female mammary gland. Decrease in weight of epididymides. Histopathologic changes in testes, epididymides, male accessory sex organs and male mammary gland.	Changes may occur in the following: – uterus and ovary weight decrease) – histopathologic changes in ovary, uterus/cervix, vagina and female mammary gland – testes and epididymides weights (increase) – histopathologic changes in testes, epididymides, male accessory sex organs and male mammary gland.	Decreased ovary weight. Histopathologic changes in ovary, uterus/cervix, vagina. Increased weight of epididymides, decreased testes weight. Histopathologic changes in testes, epididymides, male accessory sex organs.	Histopathologic changes in ovary, uterus/cervix, vagina. Decreased weight of epididymides, increased testes weight. Histopathologic changes in testes, epididymides, male accessory sex organs.	Increased thyroid weight. Possible liver weight increase (in combination with other thyroid-related endpoints). Histopathologic changes in thyroid gland.	Possible effects on: – uterus and ovary weight – histopathologic changes in ovary, uterus/cervix, vagina and female mammary gland – weight of testes and epididymides – histopathologic changes in testes, epididymides, male accessory sex organs and male mammary gland. Note that changes depend on nature of interference (e.g. inhibition of estrogen synthesis results in estrogen antagonism-like effects on endpoints).	Nil	Changes in adrenal weight. Histopathologic changes in adrenal, and pituitary glands. Tumour types.

Test guideline or other test method (reference to interpretation table within this document)	Endpoints for estrogen-mediated activity		Endpoints for androgen-mediated activity		Endpoints for thyroid-related activity	Endpoints for steroidogenesis-related activity	Endpoints for juvenile hormone-related activity	Endpoints potentially sensitive to, but not diagnostic of, E,A,T,S, juvenile hormone, ecdysone or retinoid modalities
	Agonistic	Antagonistic	Agonistic	Antagonistic				
OECD TG 421 Reproduction/ Developmental Toxicity Screening Test and OECD TG 422 Combined Repeated Dose Toxicity Study with the Reproduction/ Developmental Toxicity Screening Test (Table C.3.6)	Change in anogenital distance (AGD) in male (decrease) and female pups. Changes in estrus cyclicity. Genital abnormalities in male pups. Increased uterus weight, decreased ovary weight. Decrease in weight of epididymides, prostate, seminal vesicles (+ coagulating glands). Other sex accessory organs optional. Histopathologic changes in ovary and uterus. Histopathologic changes in testes, epididymides and male accessory sex organs and mammary gland.	Changes may occur in the following: – change in AGD in male and female pups – estrus cyclicity – genital abnormalities in male pups – uterine/ovary weight decrease – increase in weights of: epididymides, prostate, seminal vesicles (+ coagulating glands). Other sex accessory organs optional. Histopathologic changes in ovary and uterus. Histopathologic changes in testes, epididymides, male accessory sex organs and mammary gland.	Change in AGD in male (increase) and female pups. Genital abnormalities in male pups. Changes in weights of: uterus, ovaries (decrease). Increase in weights of: epididymides, prostate, seminal vesicles (+ coagulating glands). Other sex accessory organs optional. Decreased testes weight. Histopathologic changes in ovary and uterus. Histopathologic changes in testes, epididymides, male accessory sex organs.	Change in AGD in male (decrease) and female pups. Genital abnormalities in male pups. Nipple retention. Changes in weights of: uterus, ovaries. Decrease in weights of: epididymides, prostate, seminal vesicles (+ coagulating glands). Other sex accessory organs optional. Increased testes weight. Histopathologic changes in ovary and uterus. Histopathologic changes in testes, epididymides, male accessory sex organs.	Increased thyroid weight. Histopathologic changes in thyroid gland. Serum T4, decreased, TSH increased. Agonistic changes are opposite.	Possible effects on: – AGD in male and female pups – estrus cyclicity – weights of: uterus, ovaries, testes, epididymides, prostate, seminal vesicles (+ coagulating glands). Other sex accessory organs optional. Histopathologic changes in the above organs and in mammary glands. Note that changes depend on nature of interference (e.g. inhibition of estrogen synthesis results in estrogen antagonism-like effects on endpoints).	Nil	Changes in adrenal and pituitary weight. Histopathologic changes in adrenals and pituitary. Changes in fertility, reproduction or fetal development. Reproductive organ development may be affected by retinoid modulation. Gestation length. Dystocia. Placental weight. Number of implantations, corpora lutea. Number of live births and pre- and post-implantation loss.
OECD TG 414: Prenatal Developmental Toxicity Study (Table C.3.7)	Genital abnormalities in male pups. Change in AGD in male (decrease) and female fetuses.	Possible genital abnormalities. Change in AGD in male and female fetuses.	Possible genital abnormalities. Change in AGD in male (increase) and female fetuses.	Genital abnormalities in male pups. Change in AGD in male (decrease) and female fetuses.	Increased thyroid weight. Histopathologic changes in thyroid gland. Serum T4, decreased, TSH increased in dams. Agonistic changes are opposite.	Possible genital abnormalities. Possible change in AGD in male and female fetuses. Note that changes depend on nature of interference (e.g. inhibition of estrogen synthesis results in antiestrogen - like effects on endpoints).	Nil	Changes in : – number of implantations, corpora lutea – number of live births and post-implantation loss – litter size – sex ratio – litter/fetal weight – external, soft tissue and skeletal changes.

Test guideline or other test method (reference to interpretation table within this document)	Endpoints for estrogen-mediated activity		Endpoints for androgen-mediated activity		Endpoints for thyroid-related activity	Endpoints for steroidogenesis-related activity	Endpoints for juvenile hormone-related activity	Endpoints potentially sensitive to, but not diagnostic of, E,A,T,S, juvenile hormone, ecdysone or retinoid modalities
	Agonistic	Antagonistic	Agonistic	Antagonistic				
OECD TG 426: Developmental Neurotoxicity Study (Table C.3.8)	Decreased age at vaginal opening (VO) in offspring. Increased age at preputial separation (PPS) in offspring.	Possible effects on:. – age at VO in offspring (advance) – age at PPS in offspring.	Possible effects on:. – age at VOin offspring – age at PPS in offspring (reduction).	Decreased age at VOin offspring. Increased age at PPS in offspring.	Nil	Possible effects on: – age at VOin offspring – age at PPS in offspring. Note that changes depend on nature of interference (e.g. inhibition of estrogen synthesis results in antiestrogen-like effects on endpoints).	Nil	Changes in : – gestation length – litter size – pup survival index – litter/fetal weight – sex ratio – motor activity (including habituation), motor and sensory function, learning and memory in offspring – brain weight and histopathologic-al examination – morphometric (quantitative) evaluation of the brain.
OECD TG 410: Repeated Dose Dermal Toxicity: 21/28-Day Study (Table C.3.9)	Changes in weights of: testes. Other (target) organs may also be examined.	Possible: – changes in weights of testes. Other (target) organs may also be examined.	Possible: Changes in weights of: testes. Other (target) organs may also be examined.	Changes in weights of: testes. Other (target) organs may also be examined.	Nil	Possible: Changes in weights of: testes. Other (target) organs may also be examined. Note that changes depend on nature of interference (e.g. inhibition of estrogen synthesis results in antiestrogen-like effects on endpoints).	Nil	Changes in adrenal weight.
OECD TG 411: Subchronic Dermal Toxicity: 90-Day Study (Table C.3.10)	Changes in weights of: testes. Histopathologic changes in uterus, ovaries, testes, epididymides, prostate, seminal vesicles and female mammary gland	Possible: – changes in weights of testes – histopathologic changes in uterus, ovaries, testes, epididymides, prostate, seminal vesicles and female mammary gland.	Possible: – changes in weights of: testes – histopathologic changes in uterus, ovaries, testes, epididymides, prostate, seminal vesicles and female mammary gland.	Changes in weights of: testes. Histopathologic changes in uterus, ovaries, testes, epididymides, prostate, seminal vesicles and female mammary gland.	Histopathologic changes in thyroid gland.	Possible: – changes in weights of: testes – histopathologic changes in uterus, ovaries, testes, epididymides, prostate, seminal vesicles and female mammary gland. Note that changes depend on nature of interference (e.g. inhibition of estrogen synthesis results in antiestrogen-like effects on endpoints).	Nil	Changes in adrenal weight. Histopathologic changes in adrenals and pituitary.

Test guideline or other test method (reference to interpretation table within this document)	Endpoints for estrogen-mediated activity		Endpoints for androgen-mediated activity		Endpoints for thyroid-related activity	Endpoints for steroidogenesis-related activity	Endpoints for juvenile hormone-related activity	Endpoints potentially sensitive to, but not diagnostic of, E,A,T,S, juvenile hormone, ecdysone or retinoid modalities
	Agonistic	Antagonistic	Agonistic	Antagonistic				
OECD TG 412: 28-Day (Subacute) Inhalation Toxicity Study (Table C.3.11)	Changes in weights of: uterus (increase), ovaries (decrease), testes, epididymides (decrease). Histopathologic changes in uterus, ovaries, testes, epididymides, prostate, seminal vesicles and female mammary gland.	Possible: – changes in weights of: uterus/ovaries (decrease), testes/epididymides (increase) – histopathologic changes in uterus, ovaries, testes, epididymides, prostate, seminal vesicles and female mammary gland.	Possible: – changes in weights of: uterus, ovaries, testes, epididymides – histopathologic changes in uterus, ovaries, testes, epididymides, prostate, seminal vesicles and female mammary gland.	Changes in weights of: uterus, ovaries, testes, epididymides (decreases). Histopathologic changes in uterus, ovaries, testes, epididymides, prostate, seminal vesicles and female mammary gland.	Increased thyroid weight. Histopathologic changes in thyroid gland.	Possible: – changes in weights of: testes – histopathologic changes in uterus, ovaries, testes, epididymides, prostate, seminal vesicles and female mammary gland. Note that changes depend on nature of interference (e.g. inhibition of estrogen synthesis results in estrogen antagonism-like effects on endpoints).	Nil	Changes in adrenal weight. Histopathologic changes in adrenals and pituitary.
OECD TG 413: Subchronic Inhalation Toxicity: 90-Day Study (Table C.3.12)	Changes in weights of: uterus (increase), ovaries (decrease), testes, epididymides (decrease). Histopathologic changes in uterus, ovaries, testes, epididymides, prostate, seminal vesicles and female mammary gland.	Possible: – changes in weights of: uterus/ovaries (decrease), testes/epididymides (increase) – histopathologic changes in uterus, ovaries, testes, epididymides, prostate, seminal vesicles and female mammary gland.	Possible: – changes in weights of: uterus, ovaries, testes, epididymides – histopathologic changes in uterus, ovaries, testes, epididymides, prostate, seminal vesicles and female mammary gland.	Changes in weights of: uterus, ovaries, testes, epididymides (decreases). Histopathologic changes in uterus, ovaries, testes, epididymides, prostate, seminal vesicles and female mammary gland.	Increased thyroid weight. Histopathologic changes in thyroid gland.	Possible: – changes in weights of: testes – histopathologic changes in uterus, ovaries, testes, epididymides, prostate, seminal vesicles and female mammary gland. Note that changes depend on nature of interference (e.g. inhibition of estrogen synthesis results in estrogen antagonism-like effects on endpoints).	Nil	Changes in adrenal weight. Histopathologic changes in adrenals and pituitary.
OECD TG 409: Repeated Dose 90-Day Oral Toxicity Study in Non-rodents (Table C.3.13)	Changes in weights of: uterus, ovaries, testes, epididymides. Histopathologic changes in uterus, ovaries, testes, epididymides, prostate, seminal vesicles and female mammary gland.	Possible: – changes in weights of: uterus, ovaries, testes, epididymides – histopathologic changes in uterus, ovaries, testes, epididymides, prostate, seminal vesicles and female mammary gland.	Possible: – changes in weights of: uterus, ovaries, testes, epididymides – histopathologic changes in uterus, ovaries, testes, epididymides, prostate, seminal vesicles and female mammary gland.	Changes in weights of: uterus, ovaries, testes, epididymides. Histopathologic changes in uterus, ovaries, testes, epididymides, prostate, seminal vesicles and female mammary gland.	Increased thyroid weight. Histopathologic changes in thyroid gland.	Possible: – changes in weights of: testes – histopathologic changes in uterus, ovaries, testes, epididymides, prostate, seminal vesicles and female mammary gland. Note that changes depend on nature of interference (e.g. inhibition of estrogen synthesis results in estrogen antagonism-like effects on endpoints).	Nil	Changes in adrenal weight. Histopathologic changes in adrenals and pituitary.

Test guideline or other test method (reference to interpretation table within this document)	Endpoints for estrogen-mediated activity		Endpoints for androgen-mediated activity		Endpoints for thyroid-related activity	Endpoints for steroidogenesis-related activity	Endpoints for juvenile hormone-related activity	Endpoints potentially sensitive to, but not diagnostic of, E,A,T,S, juvenile hormone, ecdysone or retinoid modalities
	Agonistic	Antagonistic	Agonistic	Antagonistic				
OECD TG 416: Two-Generation Reproduction Toxicity Study (Table C.3.14)	Change in AGD in male (decrease) and female pups. Changes in estrus cyclicity (P, F1). Decreased age at VO (F1). Increased age at PPS (F1). Changes in weights of: (P, F1) uterus (increase), ovaries, testes, epididymides (decrease), prostate, seminal vesicles (+ coagulating glands). Histopathologic changes in vagina, uterus (+ cervix), ovaries, testis, epididymis, prostate, seminal vesicles and coagulating glands. Reductions in sperm parameters: sperm numbers (testicular homogenization-resistant spermatids and cauda epididymal sperm reserves), sperm motility, sperm morphology (P, F1).	Changes may occur in the following: – AGD in male and female pups – estrus cyclicity (P, F1) – age at VO (F1) – age at PPS (F1) – weights of: (P, F1) uterus, ovaries, testes, epididymides, prostate, seminal vesicles (+ coagulating glands) – histopathologic changes in the above organs. Sperm parameters: sperm numbers (testicular homogenization-resistant spermatids and cauda epididymal sperm reserves), sperm motility, sperm morphology (P, F1).	Studies using androgens are lacking. However, changes may occur in the following: – increased AGD in male pups, change in AGD in female pups – estrus cyclicity (P, F1) – age at VO (F1) – age at PPS (F1) – weights of: (P, F1) uterus, ovaries, testes, epididymides, prostate, seminal vesicles (+ coagulating glands) – histopathologic changes in the above organs. Sperm parameters: sperm numbers (testicular homogenization-resistant spermatids and cauda epididymal sperm reserves), sperm motility, sperm morphology (P, F1).	Decreased AGD in male pups, change in AGD in female pups. Changes in estrus cyclicity (P, F1). Changes in age at VO (F1). Increased age at PPS (F1). Changes in weights of: (P, F1) uterus, ovaries, testes, epididymides (decrease), prostate, seminal vesicles (+ coagulating glands). Histopathologic changes in the above organs. Reductions in sperm parameters: sperm numbers (testicular homogenization-resistant spermatids and cauda epididymal sperm reserves), sperm motility, sperm morphology (P, F1).	Increased thyroid weight. Possible liver weight increase (in combination with other thyroid-related endpoints). Histopathologic changes in thyroid (follicular cell height increase and colloid area decrease).	Possible effects on: –AGD in male and female pups – estrus cyclicity (P, F1) – age at VO (F1) – age at PPS (F1) – changes in weights of: (P, F1) uterus, ovaries, testes, epididymides, prostate, seminal vesicles (+ coagulating glands) – histopathologic changes in the above organs. Reductions in sperm parameters: sperm numbers (testicular homogenization-resistant spermatids and cauda epididymal sperm reserves), sperm motility, sperm morphology (P, F1). Note that changes depend on nature of interference (e.g. inhibition of estrogen synthesis results in estrogen antagonism-like effects on endpoints).	Nil	Changes in: – weights of adrenals – time to mating – male fertility – female fertility – gestation length – dystocia – placental weight – number of implantations, corpora lutea – number of live births and pre- and post-implantation loss – litter size – sex ratio (F1, F2) – litter/pup weight – pup survival index – abnormalities in pup development (F1, F2). Reproductive organ development may be affected by retinoid modulation.

Test guideline or other test method (reference to interpretation table within this document)	Endpoints for estrogen-mediated activity		Endpoints for androgen-mediated activity		Endpoints for thyroid-related activity	Endpoints for steroidogenesis-related activity	Endpoints for juvenile hormone-related activity	Endpoints potentially sensitive to, but not diagnostic of, E,A,T,S, juvenile hormone, ecdysone or retinoid modalities
	Agonistic	Antagonistic	Agonistic	Antagonistic				
OECD TG 443: Extended One-Generation Reproductive Toxicity Study (EOGRTS) (Table C.3.15)	Change in AGD in male and female pups. Changes in estrus cyclicity (P, F1). Decreased age at VO (F1). Increased age at PPS (F1). Genital abnormalities. Changes in weights of: (P, F1) uterus, ovaries, testes, epididymides, prostate, seminal vesicles (+ coagulating glands). Histopathologic changes in the above organs. Histopathologic changes (proliferative) in E,A,T,S mammary glands. Changes in sperm parameters: sperm numbers sperm motility, sperm morphology (P, F1).	Changes may occur in the following: – change in AGD in male and female pups – estrus cyclicity (P, F1) – age at VO (F1) – age at PPS (F1) – genital abnormalities – weights of: (P, F1) uterus, ovaries, testes, epididymides, prostate, seminal vesicles (+ coagulating glands) – histopathologic changes in the above organs – histopathologic changes in mammary glands – changes in sperm parameters: sperm numbers sperm motility, sperm morphology (P, F1).	Studies using androgens are lacking. However, changes may occur in the following: – increased AGD in male pups, change in AGD in female pups – decreased age at PPS (F1) – genital abnormalities – weights of: (P, F1) uterus, ovaries, testes, epididymides, prostate, seminal vesicles (+ coagulating glands) – histopathologic changes in the above organs and in mammary glands – changes in sperm parameters: sperm numbers sperm motility, sperm morphology (P, F1).	Decreased AGD in male pups, change in AGD in female pups. Increased age at PPS (F1). Genital abnormalities. Nipple retention. Changes in weights of: (P, F1) testes, epididymides, prostate, seminal vesicles (+ coagulating glands). Histopathologic changes in the above organs and in mammary glands. Changes in sperm parameters: sperm numbers sperm motility, sperm morphology (P, F1).	Increased thyroid weight. Possible liver weight increase (in combination with other thyroid-related endpoints). Histopathologic changes in thyroid. Serum T4, decreased, TSH increased.	Possible effects on: – AGD in male and female pups – estrus cyclicity (P, F1) – age at VO (F1) – age at PPS (F1) – genital abnormalities – changes in weights of: (P, F1) uterus, ovaries, testes, epididymides, prostate, seminal vesicles (+ coagulating glands) – histopathologic changes in the above organs – changes in sperm parameters: sperm numbers sperm motility, sperm morphology (P, F1) – histopathologic changes in mammary glands.	Nil	Changes in weights of adrenals and pituitary. Histopathologic changes in adrenals. Changes in : – time to mating – male fertility – female fertility – dystocia – gestation length – number of implantations, corpora lutea – number of ovarian follicles – number of live births and post-implantation loss – litter size – viability index – placental weight – sex ratio (F1) – litter/pup weight – pup survival index – abnormalities in pup development (F1). Reproductive organ development may be affected by retinoid modulation. Apical endpoints from the developmental neuro- and immunotoxicity cohorts may be sensitive to endocrine modulation. Specifically: – Effects on brain weight and histopathological examination. Morphometric (quantitative) evaluation of the brain. – Effects in: auditory startle test, functional observation battery, motor activity tests.

Test guideline or other test method (reference to interpretation table within this document)	Endpoints for estrogen-mediated activity		Endpoints for androgen-mediated activity		Endpoints for thyroid-related activity	Endpoints for steroidogenesis-related activity	Endpoints for juvenile hormone-related activity	Endpoints potentially sensitive to, but not diagnostic of, E,A,T,S, juvenile hormone, ecdysone or retinoid modalities
	Agonistic	Antagonistic	Agonistic	Antagonistic				
B. Guidelines that have not received full validation by the OECD, or are in the process of OECD validation, or which have been validated and published by other organisations								
Pubertal Development and Thyroid Function Assay in Peripubertal Male Rats (Male PP Assay) (US EPA OPPTS 890.1500) (Table C.3.16)	Assay is not designed to detect this modality but the following changes may occur: – increased age at PPS – decreased weight of seminal vesicles (+ coagulating glands), ventral prostate, dorsolateral prostate, LABC, epididymides – decreased testis weight – histopathologic changes in testes, epididymides – iIncreased serum testosterone.	Assay is not designed to detect this modality. However, the following changes may occur in the following endpoints: – age at PPS – weight of seminal vesicles (+ coagulating glands), ventral prostate, dorsolateral prostate, LABC, epididymides – testis weight – histopathologic changes in testes, epididymides – serum testosterone	Decreased age at PPS. Increased weight of seminal vesicles (+ coagulating glands), ventral prostate, dorsolateral prostate, LABC, epididymides. Decreased testis weight. Histopathologic changes in testes, epididymides. Decreased serum testosterone.	Increased age at PPS. Decreased weight of seminal vesicles (+ coagulating glands), ventral prostate, dorsolateral prostate, LABC, epididymides. Increased testis weight. Histopathologic changes in testes, epididymides. Increased serum testosterone.	Increased thyroid weight. Possible liver weight increase (in combination with other thyroid-related endpoints). Histopathologic changes in thyroid (follicular cell height increase and colloid area decrease). Serum T4 decreased, TSH increased.	Possible effects on: – PPS – weight of seminal vesicles (+ coagulating glands), ventral prostate, dorsolateral prostate, LABC, epididymides – histopathologic changes in testes, epididymides – serum testosterone.	Nil	Changes in weight of pituitary and/or adrenals.
Pubertal Development and Thyroid Function Assay in Peripubertal Female Rats (Female PP Assay) (US EPA OPPTS 890.1450) (Table C.3.17)	Decreased age at VO. Increased weight of uterus and decreased weight of ovaries. Histopathologic changes in uterus and ovaries. Decreased age at first estrus. Changes in estrus cyclicity.	Tthe following changes may occur: – increased age at VO – decreased weight of uterus – histopathologic changes in uterus and ovaries – increased age at first estrus – changes in estrus cyclicity.	Assay is not designed to detect this modality but the following changes may occur: – increased age at VO – decreased weight of uterus and ovaries – histopathologic changes in uterus and ovaries – increased age at first estrus – changes in estrus cyclicity.	Assay is not designed to detect this modality but the following changes may occur: – decreased age at VO – decreased weight of ovaries – histopathologic changes in uterus and ovaries.	Increased thyroid weight. Possible liver weight increase (in combination with other thyroid-related endpoints). Histopathologic changes in thyroid (follicular cell height increase and colloid area decrease). Serum T4 decrease, TSH increased.	Possible effects on: – age at VO – weight of uterus and ovaries – histopathologic changes in uterus and ovaries – estrus cyclicity.	Nil	Changes in weight of pituitary and/or adrenals.

Notes: 1. Simultaneous measurement of genotypic sex ratio (in Japanese medaka or stickleback at present) allows a more powerful detection of any effects on phenotypic sex ratio. However, sufficient power can be achieved by using an appropriate number of animals with phenotypic sexing alone, as specified in the guideline. 2. Accelerated or asynchronous development is considered by many authorities to be diagnostic of thyroid active chemicals, in addition to abnormal thyroid histopathology. Retarded development may be due either to thyroid-active chemicals or to systemic toxicants. 3. Primary histopathological criteria in gonads include the following: males – increased spermatogonia; testis-ova; testicular degeneration; Leydig cell hyperplasia/hypertrophy. Females – increased oocyte atresia; perifollicular cell hyperplasia/hypertrophy; decreased yolk formation; changes in ovarian staging. Although these endpoints are indicative of endocrine activity, care should be taken in their interpretation because some (e.g. oocyte atresia) can also be caused by certain types of systemic toxicity.

B.3. Cross-species extrapolations

91. Cross-species extrapolations should be considered during data assessment. Endocrine systems with respect to hormone structure, receptors, synthesis pathways, hormonal axes and degradation pathways are well conserved across vertebrate taxa especially in the case of estrogen, androgen and thyroid hormones and steroidogenesis. In invertebrates, many systems are distinct from those in vertebrates and are not fully understood; however, the retinoic acid system is also relevant in many species (OECD, 2017b). When interpreting data for endocrine assessment, this conservation should be borne in mind as results from tests using human *in vitro* or non-human mammalian (*in vitro and in vivo*) systems may be highly relevant for vertebrate wildlife species and vice versa. In addition, results from non-human mammalian studies are also highly relevant for mammalian wildlife species. Caution should be exercised, however, when extrapolating in this way, as species differences in exposure pathways, ADME, organ physiology, effects of hormones at different life stages across taxa/classes and other differences should be considered. The consequences of the action of a hormone may be different in different species, even if the molecular initiating event is the same.

92. Cross-species conservation was clearly demonstrated by Ankley and Gray (2013), who conducted an analysis using model chemicals acting (primarily) as ER agonists (17α-ethynylestradiol, methoxychlor, bisphenol A), AR agonists (methyltestosterone, 17β-trenbolone), AR antagonists (flutamide, vinclozolin, p,p^1-DDE) or inhibitors of steroidogenic enzymes (ketoconazole, fadrozole, fenarimol, prochloraz). All chemicals had been tested in the US EPA Endocrine Disruptor Screening Program (EDSP) Fish Short-Term (21-day) Reproduction Assay (FSTRA, OECD TG 229) and in one or more of the four *in vivo* US EPA EDSP Tier 1 screens with rats (Uterotrophic, Hershberger, male and female pubertal assays). There was a high concordance between the fish and rat assays with respect to identifying chemicals that impacted specific endocrine pathways of concern. Although most chemicals were detected as positive in both rat and fish assays, the degree of effect did vary. For example, the effects of competitive inhibitors of steroid hormone synthesis were far more obvious in the fish assay, whereas the activity of androgen receptor antagonists was clearer in mammalian assays.

93. Another example of useful cross-species extrapolation concerns thyroid activity in amphibians and mammals. Pickford (2010) studied 41 thyroid-active chemicals which in many cases had been tested both in thyroid-sensitive amphibian screens (more or less similar to the Amphibian Metamorphosis Assay [AMA]) and in thyroid-sensitive mammalian screens such as the male and female rat pubertal assays. Consistent with the work of Ankley and Gray (2013), there was strong concordance between the results of mammalian assays and those with a non-mammalian vertebrate. In only one case (methoxychlor) was thyroid activity seen in amphibians but not in mammals, and none of the chemicals active in mammals were negative in amphibians. As with the rat/fish comparisons, the types and degrees of effect varied considerably between rats and frogs, but there is no doubt that useful predictions of *in vivo* thyroid activity are possible right across the vertebrate spectrum, either from amphibians to mammals or vice versa. Hence, there seems to be a good foundation for extrapolation of qualitative screening level information between these two animal groups, although it should be noted that only the AMA, the LAGDA and the *Xenopus* Embryonic Thyroid Signalling Assay (XETA, not yet fully validated) are able to identify thyroid agonists and disturbance to peripheral tissue deiodination.

94. Mammalian toxicity studies are aimed at identifying potential hazards relevant for protecting human health, the primary goal being to protect the individual. For ecotoxicology, the primary goal is the protection of populations, and therefore the relevance of findings may differ (for example, see Marty et al. [2017]). In particular, it is important to note that an adverse apical outcome as determined in a Level 4 or 5 study does not necessarily imply that adverse effects would follow in an exposed wildlife population, effects which are part of the definition of an ED. Marty et al. (2017) describe the various considerations which they believe should be made when extrapolating from effects on individuals to impacts on populations. In some jurisdictions, however, effects on growth, reproduction and development are considered as population-relevant hazard endpoints and used as such in regulatory decision making. However, studies designed to determine endocrine effects have many commonalities, for example they need to use adequately sensitive species and life stages; have mechanistic endpoints that are diagnostic for endocrine pathways of concern; and in some cases they also show linkage between mechanistic responses and apical, adverse outcomes (Coady et al., 2017).

95. To help predict susceptibility across species, the US EPA has developed an online screening tool (SeqAPASS) that allows extrapolation of toxicity information across species (LaLone et al., 2016). SeqAPASS extrapolates from data-rich model organisms to thousands of other non-target species to evaluate their specific potential chemical susceptibility. The sensitivity of a species to a chemical is determined by a number of factors, one of which is the presence or absence of proteins that interact with chemicals ("protein targets"). Linking to various databases, SeqAPASS evaluates the similarities of amino acid sequences and protein structure to identify whether a protein target is present for a chemical interaction in other non-target species. A chemical interaction with the protein target could disrupt biological processes, leading to unintended adverse effects on survival, growth, development and reproduction. This method, for example, can be used to predict whether a pesticide, developed to control a pest species, would affect other, non-target species such as pollinators or protected species.

B.4. Considering potential for multiple modes of endocrine action

96. When assessing results from an assay or a combination of assays, although it might be assumed that EASs will have a single, highly specific mode of endocrine action, this is often not the case. To take a few examples, it has been shown in various *in vitro* assays that: zearalenone is both an estrogen agonist and an androgen antagonist (Molina-Molina et al., 2014); some metabolites of brominated flame retardants are both anti-estrogenic and anti-androgenic (Fic et al., 2014); some triazole fungicides such as epoxyconazole are both aromatase inhibitors and anti-androgens (Kjaerstad et al., 2010); and bisphenol-A and some other phenol derivatives are both estrogenic and anti-androgenic (Paris et al., 2002). It should also be noted in passing that some chemicals show promiscuous activity in nuclear hormone receptor assays which are not necessarily predictive of adverse outcomes but may be attributable to such factors as assay interference and cytotoxicity, etc.

97. Such effects can also be found *in vivo*. In fish, Ankley et al. (2001; 2005; 2007) have demonstrated that methyltestosterone is both androgenic and less potently estrogenic (probably via aromatisation); that the azole fungicides ketoconazole and prochloraz can both inhibit aromatase (leading to masculinisation) but also inhibit testosterone production (probably via inhibition of CYP17). Other azoles such as prochloraz are also AR antagonists (i.e. they are true anti-androgens), and can weakly block both the fish and mammalian ARs. In rat studies, administration of prochloraz during pregnancy causes increased nipple retention in males and increased anogenital distance in female pups

(Vinggaard et al., 2005; Melching-Kollmuss et al., 2017). It is also positive in the Hershberger assay (Vinggaard et al., 2002; Blystone et al., 2007). All are hallmarks of AR antagonism. See also case studies for OECD GD 150 in OECD (2012b). There are many other examples of such multiple effects. The breast cancer drug tamoxifen is a classical example of a substance with multiple MOA as it is a weak ER agonist in the mammary gland at low doses but becomes a potent antagonist at high doses (Kuiper, van den Bemd and van Leeuwen, 1999; Jordan, 1992; Vandenberg et al., 2012). In addition, the estrogenic and antiestrogenic effects of tamoxifen may also result from interaction of ERα and ERβ within a given cell, because ERβ may function as a dominant negative regulator (Pettersson, Delaunay and Gustafsson, 2000; Sotoca et al., 2008; Huang, Warner and Gustafsson et al., 2015; Madeira et al., 2013).

98. It is also possible that different MOA are manifested differently in different species or within different organs. Continuing the example of tamoxifen, it acts as an ER agonist and antagonist in the uterus, and as an agonist in bone in rats and humans (Kim et al., 2002; Kleinstreuer et al., 2016; Lufkin, Wong and Deal, 2001; Fontana and Delmas, 2003). In general terms, for many substances, it appears that one MOA usually predominates (i.e. one MOA has a higher potency than the others). Tamoxifen, for example, is generally considered an ER antagonist. However, phenomena such as those described above can obviously lead to difficulties in the interpretation of assay data since a very clear pattern of effects *in vivo* reflecting only one mechanism/mode of action can only seldom be expected. It may be possible for a substance's agonistic effects, for example when conclusions are drawn based on specific test data with certain dose selections, to be obscured by its antagonistic effects, thus leading to a false-negative conclusion.

99. Although these examples are from the E,A and S pathways, multiple MOA are not limited to these. For example, genistein and daidzein can activate both ERs and PPARs causing dose-dependent effects (Dang et al., 2003; Dang and Lowik, 2004). In addition, different MOA may operate at different doses (Dang, 2009). The estrogenic and antiestrogenic activity of genistein or daidzein can be explained by an activation of ERs at low doses (estrogenic) and an interaction between ERs and PPARs at high doses (antiestrogenic) (Dang and Lowik, 2005).

100. It is in cases such as these that the value of a WOE approach becomes clear. Multiple MOA may well be revealed by *in silico* modelling, or by a battery of *in vitro* assays. Such results should then alert those interpreting *in vivo* data to look out for apparently anomalous or equivocal results. For example, although the observation that methyl testosterone simultaneously causes masculinised secondary sexual characteristics and elevated vitellogenin titres in fish (Ankley et al., 2001) could be dismissed as experimental error, careful scrutiny of *in vitro* and other available data may reveal a genuine underlying cause.

101. The development of such understanding is important when establishing links between an endocrine MOA and an adverse apical effect, an essential component of the hazard evaluation of EDs. However, it is also critical to appreciate that the most important issue is whether or not the combined apical effect is considered adverse.

B.5. Use of weight of evidence and adverse outcome approaches

102. Although assessment of the potential of a substance to interact with the endocrine system and possibly whether it is an ED requires a WOE evaluation, detailed guidance is not provided here because there are many guidance documents already written, both generic for chemical assessment and specific for assessment of endocrine disruption. WOE has been defined by the World Health Organization as "a process in which all of the evidence

considered relevant for a hazard identification/characterisation is evaluated and weighted" (WHO/IPCS, 2009). Selection of appropriate guidance may depend on the objective of the evaluation and regional approaches or frameworks (e.g. a regulatory requirement for assessment of a substance within the EU). Several of these have been published, for example Solecki et al. (2017); US EPA (2011); Vandenberg et al. (2016); National Academies of Sciences, Engineering and Medicine (2017). A WOE assessment can be considered to consist of three basic steps: 1) assembling the evidence; 2) weighing the evidence; and 3) integrating the evidence (EFSA, 2017a). The information in the current GD may help to define endocrine endpoints and interpret data with respect to endocrine activity/disruption. Endpoints and their relevance to (eco)toxicity are also discussed in Manibusan and Touart (2017) and Marty et al. (2017).

103. Relevance and reliability of the assembled evidence should be addressed. Globally, different chemical legislations already require assessment and use of relevant and reliable published literature. Relevance is usually assessed first, often at the point of acquiring abstracts from a literature search. Reliability is then assessed for only those papers/reports that are considered relevant. In this context, reliability refers to data quality. There are many methods available for addressing reliability, and it is important to use a transparent process to identify high-quality data (using specific criteria). The EFSA (2011) suggests several methods; the ToxR tool (Schneider et al., 2009); and the methods of Klimisch, Andreae and Tillmann (1997) are frequently used. The ToxR tool is a useful tool that is very easy to use for assessing the reliability of publications, although some authorities claim that it is biased in favour of Good Laboratory Practice (GLP) studies. Fenner-Crisp and Dellarco (2016); Kaltenhäuser et al. (2017); and Moermond et al. (2017) review and discuss issues specific to the use of all types of data for regulatory decision making. There has been some debate about the use of studies conducted according to GLP and standardised test guidelines (which is generally the case for the present guidance), compared with non-standard or non-GLP literature data (Zoeller et al., 2015), but all information used should be scientifically robust. Essentially any information that is deemed scientifically relevant and reliable should be included in the evaluation.

104. Once the information has been assessed for relevance and reliability, then it is helpful to assemble the data in a framework in order to collate data on effects relevant for assessing the endocrine axes. The OECD Conceptual Framework may be used as a guide for collating assays at the different levels, distinguishing screening data from test data, and determining whether effects seen in higher tier tests are corroborated by lower tier data and whether they are biologically plausibly linked to endocrine activity. Such an approach was carried out in case studies described in Matthiessen et al. (2017).

105. Analysis of MOA may be required for substances acting via interactions with endocrine pathways. Human (and population) relevance should also be considered. By default the relevance to human or population should be assumed, unless the opposite has been demonstrated. Guidance on these using the Bradford Hill criteria and several case studies has been published (WHO, 2007). Applying endocrine-specific MOA may, however, be challenging, to distinguish between responses that are adaptive versus adverse, especially in non-mammalian species (Dang, 2016; Wheeler and Coady, 2016; Mihaich et al., 2017; EFSA, 2017b).

106. In order to weigh and integrate the evidence, a framework may be used, as described above, or expert judgement without a framework (although this is less transparent). Table B.2 provides a summary of some of the published approaches to WOE assessment for EAS and their attributes and uncertainties.

Table B.2. **A selection of evidence approaches for assessment of endocrine effects (in order of publication date)**

Reference	Comments
Boobis et al. (2006, 2008); WHO (2007)	– Analyses the relevance of cancer and non-cancer modes of action (MOA) for humans using the International Programme on Chemical Safety (IPCS) framework.
OECD (2008)	– Workshop report on integrated testing approaches.
OECD (2010b)	– Workshop report on endocrine disrupters.
CEFIC-EMSG (2010)	– Guidance for human health and vertebrate wildlife. Addresses the issues of data relevance, quality and significance – using a weight of evidence (WOE). Indicates whether, and what action needs to be taken, in order to assess the hazards and risks of a substance. – Slightly outdated. Some more recent assays are missing.
DK EPA (2011)	– A scientific WOE approach to the establishment of Criteria for Endocrine Disrupters and Options for Regulation in the EU (REACH, PPPR, BPR).
Bars et al. (2011)	– Output from European Centre for Ecotoxicology and Toxicology of Chemicals workshop. – Suggests scientific criteria for the determination of endocrine disrupting properties that integrate information from both regulatory (eco)toxicity studies and mechanistic/screening studies. – The criteria suggested are designed for EU regulatory requirements but the paper also discusses the US approach.structurally related chemicals.
Borgert et al. (2011)	– WOE approach for the US EPA Endocrine Disruptor Screening Program (EDSP), but relevant generally. – Suggests hypothesis testing with quantitative weightings for endpoints to give a WOE score and a narrative developed to clearly describe the final determinations.
US EPA (2011)*	– Suggested WOE approach for US EPA assessment of EDSP Tier 1 studies and need for Tier 2. – Conclusions regarding the potential of a substance to interact with the estrogenic, androgenic or thyroidal hormonal pathways. Uses alignment table of endpoints from all studies across taxa.
Juberg et al. (2013)	– Case study example of use of the OECD Conceptual Framework, assays, endpoints, etc. – Applicable across regulatory areas.
EFSA (2013)	– Provides opinion on criteria, test methods and critical aspects. Uses WHO definition. – An endocrine disrupter is defined by three criteria: 1) an adverse effect in an intact organism or a (sub)population; 2) an endocrine activity; and 3) a plausible causal relationship between the two.
Weltje et al. (2013)	– Update to Bars et al. (2011) with a focus on ecotoxicology.
Borgert et al. (2014)	– Follow-up to Borgert et al. (2011) with detailed rationale for weighting the EDSP endpoints. – Output from expert panel (Endocrine Policy Forum). – Case study example in de Peyster and Mihaich (2014).
van Der Kraak et al. (2014)	– Quantitative WOE approach used for evaluation of atrazine in fish, amphibians and reptiles. – All studies scored for relevance of response to adverse outcomes and strength of methods.
Simon et al. (2014); Meek et al. (2014)	– Updates to IPCS human relevance framework.
Lutter et al. (2015)	– Review of WOE approaches in literature. Some discussion of US EPA and the European Chemials Agency approaches. – Not specific for endocrine disrupters, no decision-making tools.
Becker et al. (2015)	– Use the Bradford-Hill considerations of biological plausibility, empirical support (dose-response, temporality and incidence) and essentiality in building adverse outcome pathways. OECD approach. – WOE evaluations and case studies.
Christiansen et al. (2015)	– Information/testing strategy for identification of substances with endocrine disrupting properties in the EU. Suggests information/testing strategies for adequate identification of endocrine disrupters. Based on OECD GD 150 and OECD Fish Toxicity Framework (OECD STA 171).
Becker et al. (2017)	– Proof of concept extension of the IPCS framework for scoring confidence in the supporting data to improve scientific justification for MOA. Not specific for endocrine disrupters.
Vandenberg et al. (2016)	– Proposes a framework for systematic literature review and integrated assessment (SYRINA) of endocrine studies. Tailored to the IPCS/WHO definition of an endocrie disrupter. – Recommended by The Endocrine Society

Table B.2. **A selection of evidence approaches for assessment of endocrine effects (in order of publication date)** (*continued*)

Reference	Comments
National Academies of Sciences, Engineering and Medicine (2017)	– Describes the application of systematic literature review methodology and development of a generic strategy for evaluating evidence of low-dose effects of EAS. – Recommended by the Endocrine Society.
Beronius and Vandenberg (2016)	– Discusses the advantages and challenges of applying systematic literature review methodology in the identification and assessment of endocrine disrupters.
Gross et al. (2017)	– Reviews WOE approaches to distil key recommendations for the evaluation of potential endocrine disrupter properties of chemicals. Makes recommendations for use within EU regulatory contexts.
ECHA (2016)	– Guidance in use of WOE for REACH.
EFSA (2017a; 2017b)	– Addresses the use of the WOE generally (in areas under EFSA's remit) using both qualitative and quantitative approaches. Several case studies illustrate the applicability of the proposed approach (2017a). Biological relevance addressed in 2017b. Not specific for endocrine activity.
EFSA-ECHA (2017)	– Guidance document for the identification of endocrine disruptors in the context of Regulation (EU) No. 528/2012 and (EC) No. 1107/2009. Currently in draft form.

*. https://www.regulations.gov/document?D=EPA-HQ-OPPT-2010-0877-0021.

107. One approach that is incorporated into many WOE processes and the OECD CF is the concept of adverse outcome pathways (AOPs). AOPs are analytical constructs that describe a sequential chain of causally linked events at different levels of biological organisation that lead to an adverse health or ecotoxicological effect (see Ankley et al., 2010; OECD, 2016). AOPs are not chemical-specific but use chemicals as examples that cause the effects. In the context of a WOE analysis, an AOP could provide a basis for identifying regulatory data needs and supporting test interpretation. AOPs are available, or an AOP can be constructed, for the linkage between a substance acting via a known molecular initiating event, such as activation of the estrogen receptor, and adverse "downstream" consequences (e.g. altered sexual differentiation). Since the linkages between the molecular initiating event and subsequent key events leading to an adverse outcome are causal in nature, the basic construct directly informs WOE analyses. An example of this type of AOP-based WOE analysis for the effects of inhibition of sex steroid synthesis (aromatase activity) on reproduction in fish is described in Becker et al. (2015). AOPs help to organise the information available from studies dedicated to the identification on ED- and non-ED related key events. In itself though, an AOP cannot be used as a decision scheme in a regulatory context.

108. The OECD has an ongoing AOP Development Programme, overseen by the extended Advisory Group on Molecular Screening and Toxicogenomics (EAGMST). The OECD AOP-Knowledge Base (AOP-KB) can be found via the e.AOP portal. This enables searching and browsing of AOPs, links to published AOPs, informs on the status of AOPs, and allows browsing of AOP external review reports. The future of AOP development and regulatory decision making is discussed in LaLone et al. (2017).

109. Integrated approaches to testing and assessment mayalso be integrated in WOE and AOPs. Information on this can be found in Section B.1.1.3.

B.6. Regulatory experience of endocrine assessment

110. Use of ED/EAS-sensitive screens and tests in a regulatory, as opposed to research, context is relatively new. In the European Union, a few chemicals with endocrine activity have been evaluated under the REACH legislation, while in the United States, the Endocrine Disruptor Screening Program (EDSP) has so far screened a few dozen chemicals

(mainly pesticides) to which humans and/or vertebrate wildlife are exposed and subjected those which screened positive to higher tier testing. In addition, Japan has conducted endocrine screening and testing of some chemicals which are widespread in the Japanese environment.[2] Although experience is still sparse, it is helpful to consider it briefly in more detail because it provides some realistic perspectives on the somewhat theoretical advice in this document.

B.6.1. Regulatory experience in the United States

111. One example of the application of these assays in a regulatory context exists within the US EPA's EDSP. The EDSP uses validated assays and/or models to determine, based on the WOE, if there is a disruption in the endocrine system for the estrogen, androgen and/or thyroid (E, A, or T) pathways. This is accomplished through a tiered-testing approach, including: screening (Tier 1) and identification of any adverse endocrine-related effect and quantification of dose-response relationships for hazard identification/characterisation (Tier 2).

112. Tier 1 screening consists of a battery of complementary *in vitro* and *in vivo* assays meant to maximise the sensitivity and reliability for determining the potential of a chemical to interact with the E, A or T pathways. In addition to the available Tier 1 assay data, other scientifically relevant information, including general toxicity data and open literature studies of sufficient quality, are considered in the WOE assessment. The diversity of endocrine endpoints and test species in the battery allow for the evaluation of the consistency of responses.

113. In the US EPA Tier 1 WOE analysis, the EPA assembles and integrates information from individual lines of evidence within the conceptual framework of an AOP on the basis of complementarity and redundancy. Complementarity refers to the concordance of endpoints within an assay that measures multiple endpoints and redundancy refers to the concordance of endpoints/responses across assays. These concepts are described further in US EPA's WOE guidance document (US EPA, 2011). This guidance outlines four main steps that serve as the foundation for WOE evaluations. The first step is to evaluate the individual studies for their scientific quality and relevance in assessing potential endocrine interaction(s). The second step is to integrate the data across different levels of biological organisation while examining the extent of complementarity and redundancy in the observed responses across these different levels of biological organisation. As part of this evaluation, the magnitude, direction (i.e. increase or decrease) and diagnostic specificity of responses are important to consider. As recommended by the US Federal Insecticide, Fungicide and Rodenticide Act (FIFRA) Scientific Advisory Panel in 2013 (US EPA, 2013), little weight is placed on endocrine effects that are noted exclusively at substance levels inducing overt toxicity (e.g. decreased survival or body weight), The third step is to characterise the main lines of evidence as well as any conclusions. Finally, the last step is to evaluate whether additional testing is needed based on the evidence and conclusions described above.

114. The US EPA has released its reviews of the Tier 1 screening assay results for the first 52 pesticide chemicals (active and inert ingredients) in the EDSP. For each chemical, the EPA decided whether additional (Tier 2) testing is necessary. The WOE assessments and associated data evaluation records are publically available at: https://www.epa.gov/endocrine-disruption/endocrine-disruptor-screening-program-tier-1-screening-determinations-and. In broad terms, this programme has shown the value of subjecting test chemicals to a battery of *in vitro* and *in vivo* screens, the results of which

are then used to identify a much smaller subset of chemicals for definitive testing. The cost of this approach is high, and it seems likely that cheaper high-throughput screening will ultimately become more widely used.

B.6.2. Regulatory experience in the European Union

115. In 1999 the European Commission adopted a "Community strategy for endocrine disrupters" (EC, 1999) with short-, medium- and long-term actions intended to contribute to a better environment and improved health of people within the European Union. Regulatory action under this strategy addressed endocrine disrupters in environmental and substance-specific legislation, e.g. for industrial chemicals (REACH), biocides, plant protection products and cosmetics. Under REACH (EC, 2006), substances having endocrine disrupting properties may be identified as substances of very high concern for human health and/or the environment. More information on these substances can be found on the European Chemicals Authority (ECHA) website.

116. Substances are listed in the Community Rolling Action Plan for substance evaluation under REACH during the period 2012-17 due to concerns about suspected endocrine disrupting properties. They were selected by screening the information in registration dossiers submitted to the European Chemicals Agency (ECHA) and on external data, and based on national priorities of member state competent authorities.

117. During evaluation of the available databases, the guidance provided in OECD GD 150 has been widely used and found to be of value when deciding and justifying the next steps.

118. For the chemicals evaluated under the Community Rolling Action Plan until 2017, a conclusion on endocrine disrupting properties was possible for a few substances. Mostly, however, the available information was considered not sufficient and further information on adverse effects and/or MOA was requested. The information requests address all levels of the OECD Conceptual Framework, but predominantly Level 4 and 5 studies. For human health, one of the most frequently requested single tests was the Extended One-Generation Reproductive Toxicity Study (EOGRTS, OECD TG 443), often with substance-tailored modifications of the test design. Often, there has been a need to address several concerns. For instance, the decision on whether to address a concern for developmental neurotoxicity (DNT) in OECD TG 426 or by conducting an EOGRTS with a DNT cohort including the option to modify the test in accordance with Paragraph 50 of the test guideline to include additional investigations (e.g. of learning and memory may depend on the level of concern and/or data already available for reproductive toxicity).

119. Concerns for endocrine disruption in environmental organisms have led to requests for a wider variety of tests. This may be because the standard dataset under REACH contains less ecotoxicity studies that already include ED-relevant endpoints compared to the dataset for mammalian toxicity. The information requests to address ED concerns have included the AMA (OECD TG 240) or LAGDA (OECD TG 241), Androgenised Female Stickleback Screens (AFSS, OECD TG 230 modified), Fish Sexual Development Tests (FSDT, OECD TG 234), Medaka or Zebrafish Extended One-Generation Reproduction Tests (MEOGRT or ZEOGRT), and Fish Short-Term Reproduction Assays (OECD TG 229). In several cases, modifications were made to the standard test guideline/method, e.g. collection of gonads for histopathology or measurements of vitellogenin induction in a fish bioaccumulation study (OECD TG 305). Such substance-specific tailoring of the test design is facilitated where OECD test guidelines already contain optional endpoints or guidance on how to combine studies.

120. Substance and dossier evaluation decisions taken under REACH (EU, 2006) are published on the ECHA website. The decisions contain the information requested and the rationale for the information requests. Once a substance evaluation is finalized, the conclusion documents are also published.

B.6.3. Experience in the chemical industry – views of the OECD Business and Industry Advisory Committee

121. In concert with the OECD Conceptual Framework for Testing and Assessment of Endocrine Disrupters, OECD GD 150 is considered a useful tool with which to organise and evaluate existing data. It assists in making conclusions on whether a substance is or is not an endocrine disrupter, as well as helping to guide what additional testing, if any, may be needed. The GD helps facilitate evaluations by substance producers as well as by regulatory agencies and serves as a common frame of reference for facilitating discussions. Elements of the GD that have been particularly important in this respect are:

1. The promotion of the concept of weight of evidence and that a conclusion can only be made by evaluating all of the relevant data collectively.
2. A stepwise approach to data generation. In some cases this has helped to avoid animal testing, as the GD indicated that the next most appropriate step was the generation of *in vitro* data.
3. The GD is clear that there is a need for flexibility in approach and that there may be a need to consider/generate data not specifically discussed in the guidance itself.
4. The GD provides a clear grounding to the test guidelines, many of which are relatively new to regulatory application.

122. There have also been some challenges associated with use of the GD:

1. Although the stepwise approach to data generation has obvious merits, the GD details "next step(s) which could be taken to strengthen weight of evidence if necessary" without providing guidance on how to decide if additional evidence is necessary. This is particularly problematic for substances that have been shown to have no endocrine effects throughout lower tiers of the Conceptual Framework, with more data from higher tiers (up to Level 5) being requested in order to increase the evidence for no effect and minimise uncertainty. This has the potential to increase the number of animal-intensive studies requested and performed.
2. At times the guidance suggests approaches that have not been formally validated, which creates uncertainty since most regulatory programmes require this. This includes the suggestion to perform *in vitro* assays incorporating metabolic activation, the Avian Two-Generation Reproduction Test and the Androgenised Female Stickleback Screen.
3. The GD only briefly touches on human relevance considerations and does not address the population relevance of effects. It would be helpful if the GD could suggest approaches to evaluate human/population relevance of an endocrine effect before suggesting a next step to strengthen WOE of the effect.

123. Overall, OECD GD 150 is a useful document to support sound, science-based regulatory decisions. It outlines a reasonable and pragmatic approach to the evaluation of potential endocrine disrupting properties of substances.

B.6.4. Other regulatory experience

124. Other than in the United States and the EU, regulation of chemicals based on their endocrine disrupting properties has not yet been formally implemented, although the apical effects of endocrine disupting chemicals (EDCs) (e.g. interference with reproduction) are widely used to evaluate chemicals in traditional hazard and hazard identification/ characterisation programmes. However, government-sponsored research programmes on EDCs are widespread, perhaps most prominently in Japan, where chemicals causing significant exposure to humans and vertebrate wildlife have been extensively tested for ED properties using approaches and assays which are broadly in line with those recommended in OECD GD 150.[3] The Japanese Fourth Program on Endocrine Disrupting Effects of Chemical Substances: EXTEND 2016, is currently in operation. EXTEND 2016 and its predecessors have resulted in 67 chemicals in the Japanese environment being listed as suspected EDCs.[4]

Notes

1. See: https://actor.epa.gov/edsp21.
2. See: www.env.go.jp/en/chemi/ed.html.
3. See: www.env.go.jp/en/chemi/ed.html.
4. See: www.chemsafetypro.com/Topics/Japan/Endocrine_Disrupters_Regulations_and_Lists_in_Japan.html.

References

Ågerstrand, M. et al. (2018), "A call for action: Improve reporting of research studies to increase the scientific basis for regulatory decision making", *Journal of Applied Toxicology*, Vol. 38/5, pp. 783-785, http://dx.doi.org/10.1002/jat.3578.

Ankley, G.T. and L.E. Gray (2013), "Cross-species conservation of endocrine pathways: A critical analysis of tier 1 fish and rat screening assays with 12 model chemicals", *Environmental Toxicology and Chemistry*, Vol. 32/5, pp. 1084-1087, https://doi.org/10.1002/etc.2151.

Ankley, G.T. et al. (2010), "Adverse outcome pathways: A conceptual framework to support ecotoxicology research and risk assessment", *Environmental Toxicology and Chemistry*, Vol. 29/3, pp. 730-741, http://dx.doi.org/10.1002/etc.34.

Ankley, G.T. et al. (2007), "Ketoconazole in the fathead minnow (*Pimephales promelas*): Reproductive toxicity and biological compensation" *Environmental Toxicology and Chemistry*, Vol. 26/6, pp. 1214-1223, https://doi.org/10.1897/06-428R.1.

Ankley, G.T. et al. (2005), "Effects of two fungicides with multiple modes of action on reproductive endocrine function in the fathead minnow (*Pimephales promelas*)", *Toxicological Sciences*, Vol. 86/2, pp. 300-308, https://doi.org/10.1093/toxsci/kfi202.

Ankley, G.T. et al. (2001), "Description and evaluation of a short-term reproduction test with the fathead minnow (*Pimephales promelas*)", *Environmental Toxicology and Chemistry*, Vol. 20/6, pp. 1276-1290, https://doi.org/10.1002/etc.5620200616.

Ball, N. et al. (2016), "Toward good read-across practice (GRAP) guidance", *ALTEX – Alternatives to Animal Experimentation*, Vol. 33/2, pp. 149-166, https://doi.org/10.14573/altex.1601251.

Bars, R. et al. (2011), "Science based guidance for the assessment of endocrine disrupting properties of chemicals", *Regulatory Toxicology and Pharmacology*, Vol. 59/1, pp. 37-46, https://doi.org/10.1016/j.yrtph.2010.09.003.

Becker, R.A. et al. (2017), "Quantitative weight of evidence to assess confidence in potential modes of action", *Regulatory Toxicology and Pharmacology*, Vol. 86, pp. 205-220, https://doi.org/10.1016/j.yrtph.2017.02.017.

Becker, R.A. et al. (2015), "Increasing scientific confidence in adverse outcome pathways: Application of tailored Bradford-Hill considerations for evaluating weight of evidence", *Regulatory Toxicology and Pharmacology*, Vol. 72/3, pp. 514-537, https://doi.org/10.1016/j.yrtph.2015.04.004.

Beekhuijzen, M. et al. (2016), "Update of OECD DART guidelines with endocrine disrupter relevant endpoints: Practical considerations", *Reproductive Toxicology*, Vol. 64, pp. 64-71, https://doi.org/10.1016/j.reprotox.2016.04.002.

Bell, S.M. et al. (2017), "An integrated chemical environment to support 21st century toxicology", *Environmental Health Perspectives*, Vol. 125/5.

Beresford, N. et al. (2011), "Estrogenic activity of tropical fish food can alter baseline vitellogenin concentrations in male fathead minnow (*Pimephales promelas*)", *Environmental Toxicology and Chemistry*, Vol. 30/5, pp. 1139-1145, https://doi.org/10.1002/etc.479.

Beronius, A. and L.B. Vandenberg (2016), "Using systematic review for hazard and risk assessment of endocrine disrupting chemicals", *Reviews in Endocrine and Metabolic Disorders*, Vol. 16/4, pp. 273-287, https://doi.org/10.1007/s11154-016-9334-7.

Beronius, A. et al. (2014), "Facilitating the use of non-standard *in vivo* studies in health risk assessment of chemicals: A proposal to improve evaluation criteria and reporting", *Journal of Applied Toxicology*, Vol. 34/6, pp. 607-617, https://doi.org/10.1002/jat.2991.

Blystone, C.R. et al. (2007), "Prochloraz inhibits testosterone production at dosages below those that affect androgen-dependent organ weights or the onset of puberty in the male Sprague Dawley rat", *Toxicological Sciences*, Vol. 97/1, pp. 65-74, https://doi.org/10.1093/toxsci/kfm004.

Boobis, A.R. et al. (2008), "IPCS framework for analyzing the relevance of a noncancer mode of action for humans", *Critical Reviews in Toxicology*, Vol. 38/2, pp. 87-96, http://dx.doi.org/10.1080/10408440701749421.

Boobis, A.R. et al. (2006), "IPCS framework for analyzing the relevance of a cancer mode of action for humans", *Critical Reviews in Toxicology*, Vol. 36/10, pp. 781-792, https://doi.org/10.1080/10408440600977677.

Borgert, C.J. et al. (2014), "Relevance weighting of tier 1 endocrine screening endpoints by rank order", *Birth Defects Research Part B: Developmental and Reproductive Toxicology*, Vol. 101/1, pp. 90-113, https://doi.org/10.1002/bdrb.21096.

Borgert, C.J. et al. (2011), "Hypothesis-driven weight of evidence framework for evaluating data within the US EPA's Endocrine Disrupter Screening Program", *Regulatory Toxicology and Pharmacology*, Vol. 61/2, pp. 185-191, https://doi.org/10.1016/j.yrtph.2011.07.007.

Browne, P. et al. (2015), "Screening chemicals for estrogen receptor bioactivity using a computational model", *Environmental Science & Technology*, Vol. 49/14, pp. 8804-8814, http://dx.doi.org/10.1021/acs.est.5b02641.

Casey, W.M. (2016), "Advances in the development and validation of test methods in the United States", *Toxicological Research*, Vol. 32/1, pp. 9-14, http://dx.doi.org/10.5487/TR.2016.32.1.009.

CEFIC-EMSG (2010), "Towards the establishment of a weight of evidence approach to prioritizing action in relation to endocrine disruption".

Christiansen, S. et al. (2015), "Information/testing strategies for identification of substances with endocrine disrupting properties", *Toxicology Letters*, Vol. 238/2, Supplement, pp. S41-105, http://dx.doi.org/10.1016/j.toxlet.2015.08.112.

Coady, K.K. et al. (2017), "Current limitations and recommendations to improve testing for the environmental assessment of endocrine active substances", *Integrated Environmental Assessment and Management*, Vol. 13/2, pp. 302-316, https://doi.org/10.1002/ieam.1862.

Combes, R.D. (2000), "Endocrine disrupters: A critical review of *in vitro* and *in vivo* testing strategies for assessing their toxic hazard to humans", *ATLA*, Vol. 28/1, pp. 81-118.

Cooper, R.L. et al. (2007), "Atrazine and reproductive function: Mode and mechanism of action studies", *Birth Defects Research Part B: Developmental and Reproductive Toxicology*, Vol. 80, pp. 98-112, https://doi.org/10.1002/bdrb.20110.

Dang, Z. (2016), "Interpretation of fish biomarker data for identification, classification, risk assessment and testing of endocrine disrupting chemicals", *Environment International*, Vol. 92-93, pp. 422-441, https://doi.org/10.1016/j.envint.2016.04.003.

Dang, Z.C. (2014), "Fish biomarkers for regulatory identification of endocrine disrupting chemicals", *Environmental Pollution*, Vol. 185, pp. 266-270, https://doi.org/10.1016/j.envpol.2013.11.006.

Dang, Z.C. (2009), "Dose-dependent effects of soy phyto-oestrogen genistein on adipocytes: Mechanisms of action", *Obesity Reviews*, Vol. 10/3, pp. 342-349, https://doi.org/10.1111/j.1467-789X.2008.00554.x.

Dang, Z.C. and C. Lowik (2005), "Dose-dependent effects of phytoestrogens on bone", *Trends in Endocrinology & Metabolism*, Vol. 16/5, pp. 207-213, https://doi.org/10.1016/j.tem.2005.05.001.

Dang, Z.C. and C.W. Lowik (2004), "The balance between concurrent activation of ERs and PPARs determines daidzein-induced osteogenesis and adipogenesis", *Journal of Bone and Mineral Research*, Vol. 19/5, pp. 853-861, https://doi.org/10.1359/jbmr.040120.

Dang, Z.C. et al. (2003), "Peroxisome proliferator-activated receptor y (PPARy) as a molecular target for the soy phytoestrogen genistein", *The Journal of Biological Chemistry*, Vol. 278/2, pp. 962-967, http://dx.doi.org/10.1074/jbc.M209483200.

Dinan, L. et al. (2001), "Screening of environmental contaminants for ecdysteroid agonist and antagonist activity using the *Drosophila melanogaster* B-II cell *in vitro* assay", *Environmental Toxicology and Chemistry*, Vol. 20/9, pp. 2038-2046, https://doi.org/10.1002/etc.5620200924.

DK EPA (2011), "Annex A: Report on criteria for endocrine disrupters", in: "Establishment of criteria for endocrine disrupters and options for regulation", Centre on Endocrine Disrupters, http://eng.mst.dk/media/mst/Attachments/DKEDcriteria110517_finalcorr1.pdf.

Dybdahl, M. et al. (2012), "QSAR model for human pregnane X receptor (PXR) binding: Screening of environmental chemicals and correlations with genotoxicity, endocrine disruption and teratogenicity", *Toxicology and Applied Pharmacology*, Vol. 262/3, pp. 301-309, https://doi.org/10.1016/j.taap.2012.05.008.

EC (2006), "Regulation (EC) No. 1907/2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH)", European Commission, Brussels, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02006R1907-20140410.

EC (1999), "Community strategy for endocrine disrupters: A range of substances suspected of interfering with the hormone systems of humans and wildlife", COM(1999)706 final, Commission of the European Communities, Brussels, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:51999DC0706&from=EN.

ECHA (2016), "Practical guide: How to use alternatives to animal testing to fulfil your information requirements for REACH registration", Version 2.0 July 2016, European Food Safety Authority, Parma, Italy, https://echa.europa.eu/documents/10162/13655/practical_guide_how_to_use_alternatives_en.pdf/148b30c7-c186-463c-a898-522a888a4404.

EFSA (2017a), "Guidance on the use of the weight of evidence approach in scientific assessments", draft May 2017, European Food Safety Authority, Parma, Italy.

EFSA (2017b), "Guidance for the identification of biological relevance of adverse/positive health effects from experimental animal and human studies", draft May 2017, European Food Safety Authority, Parma, Italy.

EFSA (2014), "A systematic procedure for the identification of emerging chemical risks in the food and feed chain", *EFSA Supporting Publications*, Vol. 11/1, http://dx.doi.org/10.2903/sp.efsa.2014.EN-547.

EFSA (2013), "Scientific opinion on the hazard assessment of endocrine disrupters: Scientific criteria for identification of endocrine disrupters and appropriateness of existing test methods for assessing effects mediated by these substances on human health and the environment", *EFSA Journal*, Vol. 11/3, pp. 3132, http://dx.doi.org/10.2903/j.efsa.2013.3132.

EFSA (2011), "Submission of scientific peer-reviewed open literature for the approval of pesticide active substances under Regulation (EC) No. 1107/2009", *EFSA Journal*, Vol. 9/2, p.2092, http://dx.doi.org/10.2903/j.efsa.2011.2092.

EFSA/ECHA, JRC (2017), "Guidance document for the implementation of the hazard-based criteria to identify endocrine disrupters (EDs) in the context of Regulations (ED) No. 1107/2009 and (EU) No. 528/2012", draft April 2017, European Food Safety Authority/European Chemicas Agency and the Joint Research Centre.

Fenner-Crisp, P.A. and V.L. Dellarco (2016), "Key elements for judging the quality of a risk assessment", *Environmental Health Perspectives*, Vol. 124/8, pp. 1127-1135, http://dx.doi.org/10.1289/ehp.1510483.

Fic, A. et al. (2014), "Estrogenic and androgenic activities of TBBA and TBMEPH, metabolites of novel brominated flame retardants, and selected bisphenols, using the XenoScreen XL YES/YAS assay", *Chemosphere*, Vol. 112, pp. 362-369, http://dx.doi.org/10.1016/j.chemosphere.2014.04.080.

Flynn, K. et al. (2017), "Summary of the development the US Environmental Protection Agency's Medaka Extended One-Generation Reproduction Test (MEOGRT) using data from 9 multigenerational medaka tests", *Environmental Toxicology and Chemistry*, Vol. 36/12, pp. 3387-3403, https://doi.org/10.1002/etc.3923.

Fontana, A. and P.D. Delmas (2003), "Selective estrogen receptors modulators in the prevention and treatment of postmenopausal osteoporosis", *Endocrinology & Metabolism Clinics of North America*, Vol. 32/1, pp. 219-232, https://doi.org/10.1016/S0889-8529(02)00057-9.

Gallavan, R.H. Jr et al. (1999), "Interpreting the toxicologic significance of alterations in anogenital distance: Potential for confounding effects of progeny body weights", *Reproductive Toxicology*, Vol. 13/5, pp. 383-390, https://doi.org/10.1016/S0890-6238(99)00036-2.

Gross, M. et al. (2017), "Weight of evidence approaches for the identification of endocrine disrupting properties of chemicals: Review and recommendations for EU regulatory application", *Regulatory Toxicology and Pharmacology*, Vol. 91, pp. 20-28, https://doi.org/10.1016/j.yrtph.2017.10.004.

Hornung, M.W. et al. (2017), "Avoiding false positives and optimizing identification of true negatives in estrogen receptor binding and agonist/antagonist assays", *Applied In Vitro Toxicology*, Vol. 3/2, pp. 163-181, https://doi.org/10.1089/aivt.2016.0021.

Huang, B., M. Warner and J.A. Gustafsson (2015), "Estrogen receptors in breast carcinogenesis and endocrine therapy", *Molecular and Cellular Endocrinology*, Vol. 418/Part 3, pp. 240-244, https://doi.org/10.1016/j.mce.2014.11.015.

Jacobs, M. et al. (2013), "*In vitro* metabolism and bioavailability tests for endocrine active substances: What is needed next for regulatory purposes?", *ALTEX – Alternatives to Animal Experimentation*, Vol. 30/3, pp. 331-351.

Jacobs, M.N. et al. (2008), "The use of metabolising systems for *in vitro* testing of endocrine disrupters", *Current Drug Metabolism*, Vol. 9/8, pp. 796-826.

Jensen, G.E. et al. (2011), "QSAR models for anti-androgenic effect: A preliminary study", *SAR and QSAR in Environmental Research*, Vol. 22/1-2, pp. 35-49, https://doi.org/10.1080/1062936X.2010.528981.

Jónsdóttir, S.O. et al. (2012), "Identification of cytochrome P450 2D6 and 2C9 substrates and inhibitors by QSAR analysis", *Bioorganic & Medicinal Chemistry*, Vol. 20/6, pp. 2042-2053, https://doi.org/10.1016/j.bmc.2012.01.049, https://doi.org/10.1016/j.bmc.2012.01.049.

Jordan, V.C. (1992), "The role of tamoxifen in the treatment and prevention of breast cancer", *Current Problems in Cancer*, Vol. 16/3, pp. 129-176, https://doi.org/10.1016/0147-0272(92)90002-6.

JRC (2013), "Identification of potential endocrine disrupting substances by *in silico* approaches", Joint Research Centre.

Juberg, D.R. et al. (2013), "Chlorpyrifos: Weight of evidence evaluation of potential interaction with the estrogen, androgen, or thyroid pathways", *Regulatory Toxicology and Pharmacology*, Vol. 66/3, pp. 249-263, https://doi.org/10.1016/j.yrtph.2013.03.003.

Kaltenhäuser, J. et al. (2017), "Relevance and reliability of experimental data in human health risk assessment of pesticides", *Regulatory Toxicology and Pharmacology*, Vol. 88, pp. 227-237, https://doi.org/10.1016/j.yrtph.2017.06.010.

Kim, H.S. et al. (2002), "Evaluation of the 20-day pubertal female assay in Sprague-Dawley rats treated with DES, tamoxifen, testosterone, and flutamide", *Toxicological Sciences*, Vol. 67/1, pp. 52-62.

Kjaerstad, M.B. et al. (2010), "Endocrine disrupting effects *in vitro* of conazole antifungals used as pesticides and pharmaceuticals", *Reproductive Toxicology*, Vol. 30/4, pp. 573-582, https://doi.org/10.1016/j.reprotox.2010.07.009.

Kleinstreuer, N.C. et al. (2016), "A curated database of rodent uterotrophic bioactivity", *Environmental Health Perspectives*, Vol. 124/5, pp. 556-562, http://dx.doi.org/10.1289/ehp.1510183.

Klimisch, H.-J., M. Andreae and U. Tillmann (1997), "A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data", *Regulatory Toxicology and Pharmacology*, Vol. 25/1, pp. 1-5, https://doi.org/10.1006/rtph.1996.1076.

Kucheryavenko, O. et al. (2018), "Report from the BfR Expert Hearing on Practicability of Hormonal Measurements", *Archives of Toxicol*ogy (in prep.).

Kuiper, G.G., G.J. van den Bemd and J.P. van Leeuwen (1999), "Estrogen receptor and the SERM concept", *Journal of Endocrinological Investigation*, Vol. 22/8, pp. 594-603, https://doi.org/10.1007/BF03343616.

LaLone C.A. et al. (2017), "Advancing the adverse outcome pathway framework: An international horizon scanning approach", *Environmetal Toxicology and Chemistry*, Vol. 36/6, pp. 1411-1421, https://doi.org/10.1002/etc.3805.

LaLone, C.A. et al. (2016), "Editor's highlight: Sequence alignment to predict across species susceptibility (SeqAPASS): A web-based tool for addressing the challenges of cross-species extrapolation of chemical toxicity", *Toxicological Sciences*, Vol. 153/2, pp. 228-245, https://doi.org/10.1093/toxsci/kfw119.

Larralde, C. et al. (1995), "Sex hormone changes induced by the parasite lead to feminization of the male host in murine *Taenia crassiceps* cysticercosis", *The Journal of Steroid Biochemistry and Molecular Biology*, Vol. 52/6, pp. 575-580, https://doi.org/10.1016/0960-0760(95)00062-5.

Lo Piparo, E. and A. Worth (2010), *Review of QSAR Models and Software Tools for Predicting Developmental and Reproductive Toxicity*, Publications Office of the European Union, http://dx.doi.org/10.2788/9628.

Lufkin, E.G., M. Wong and C. Deal (2001), "The role of selective estrogen receptor modulators in the prevention and treatment of osteoporosis", *Rheumatic Disease Clinics of North America*, Vol. 27/1, pp. 163-85, vii, https://doi.org/10.1016/S0889-857X(05)70192-7.

Lutter, R. et al. (2015), "Improving weight of evidence approaches to chemical evaluations", *Risk Analysis*, Vol. 35/2, pp. 186-192, https://doi.org/10.1111/risa.12277.

Madeira, M. et al. (2013), "Estrogen receptor alpha/beta ratio and estrogen receptor beta as predictors of endocrine therapy responsiveness-a randomized neoadjuvant trial comparison between anastrozole and tamoxifen for the treatment of postmenopausal breast cancer", *BMC Cancer*, Vol. 13/425, https://doi.org/10.1186/1471-2407-13-425.

Manibusan, M.K. and L.W. Touart (2017), "A comprehensive review of regulatory test methods for endocrine adverse health effects", *Critical Reviews in Toxicology*, Vol. 47/6, pp. 433-481, https://doi.org/10.1080/10408444.2016.1272095.

Mansouri, K. et al. (2016), "CERAPP: Collaborative Estrogen Receptor Activity Prediction Project", *Environmental Health Perspectives*, Vol. 124/7, pp. 1023-1033, http://dx.doi.org/10.1289/ehp.1510267.

Marty, M.S. et al. (2017), "Population-relevant endpoints in the evaluation of endocrine-active substances (EAS) for ecotoxicological hazard and risk assessment", *Integrated Environmental Assessment and Management*, Vol. 13/2, pp. 317-330, https://doi.org/10.1002/ieam.1887.

Matthiessen, P. et al. (2017), "Recommended approaches to the scientific evaluation of ecotoxicological hazards and risks of endocrine-active substances", *Integrated Environmental Assessment and Management*, Vol. 13/2, pp. 267-279, https://doi.org/10.1002/ieam.1885.

Meek, M.E. et al. (2014), "New developments in the evolution and application of the WHO/IPCS framework on mode of action/species concordance analysis", *Journal of Applied Toxicology*, Vol. 34/1, pp. 1-18, https://doi.org/10.1002/jat.2949.

Melching-Kollmuss, S. et al. (2017), "Comparing effect levels of regulatory studies with endpoints derived in targeted anti-androgenic studies: Example prochloraz", *Archives of Toxicology*, Vol. 91/1, pp. 143-162, https://doi.org/10.1007/s00204-016-1678-y.

Mihaich, E.M. et al. (2017), "Challenges in assigning endocrine-specific modes of action: Recommendations for researchers and regulators", *Integrated Environmental Assessment and Management*, Vol. 13/2, pp. 280-292, https://doi.org/10.1002/ieam.1883.

Miyagawa, S. et al. (2011), "The role of sonic hedgehog-Gli2 pathway in the masculinization of external genitalia", *Endocrinology*, Vol. 152/7, pp. 2894-2903, https://doi.org/10.1210/en.2011-0263.

Moermond, C.T.A. et al. (2017), "Assessing the reliability of ecotoxicological studies: An overview of current needs and approaches", *Integrated Environmental Assessment and Management*, Vol. 13/4, pp. 640-651, https://doi.org/10.1002/ieam.1870.

Molander, L. et al. (2015), "Science in Risk Assessment and Policy (SciRAP): An online resource for evaluating and reporting *in vivo* (eco) toxicity studies", *Human and Ecological Risk Assessment*, Vol. 21, pp. 753-762, https://doi.org/10.1080/10807039.2014.928104.

Molina-Molina, J.M. et al. (2014), "Assessment of estrogenic and anti-androgenic activities of the mycotoxin zearalenone and its metabolites using *in vitro* receptor-specific bioassays", *Food and Chemical Toxicology*, Vol. 74, pp. 233-239, https://doi.org/10.1016/j.fct.2014.10.008.

Mombelli, E. (2012), "Evaluation of the OECD (Q)SAR application toolbox for the profiling of estrogen receptor binding affinities", *SAR and QSAR in Environmental Research*, Vol. 23/1-2, pp. 37-57, https://doi.org/10.1080/1062936X.2011.623325.

Morley, N.J. (2006), "Parasitism as a source of potential distortion in studies on endocrine disrupting chemicals in molluscs", *Marine Pollution Bulletin*, Vol. 52/11, pp. 1330-1332, https://doi.org/10.1016/j.marpolbul.2006.08.025.

National Academies of Sciences, Engineering, and Medicine (2017), *Application of Systematic Review Methods in an Overall Strategy for Evaluating Low-Dose Toxicity from Endocrine Active Chemicals*, The National Academies Press, Washington, DC, https://doi.org/10.17226/24758.

OECD (2017a), "Detailed review paper on the retinoid system (draft)", OECD, Paris.

OECD (2017b), "Draft guidance document on good *in vitro* method practices (GIVIMP) for the development and implementation of *in vitro* methods for regulatory use in human safety assessment", OECD, Paris.

OECD (2016), "Guidance document for the use of adverse outcome pathways in developing integrated approaches to testing and assessment (IATA)", ENV/JM/MONO(2016)67, OECD, Paris.

OECD (2015), "Feasibility study for minor enhancements of TG 421/422 with ED relevant endpoints", OECD Series on Testing and Assessment, No. 217, OECD, Paris, www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2015)24&doclanguage=en.

OECD (2013), "Guidance document supporting of OECD test guideline on the Extended One-Generation Reproductive Toxicity Study", OECD Series on Testing and Assessment, No. 151, OECD, Paris, www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/JM/MONO(2013)10&doclanguage=en.

OECD (2012a), *Detailed Review Paper on the State of the Science on Novel* In Vitro *and* In Vivo *Screening and Testing Methods and Endpoints for Evaluating Endocrine Disrupters*, OECD Series on Testing and Assessment, No. 178, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264221352-en.

OECD (2012b), "Guidance document (GD) on standardised test guidelines for evaluating chemicals for endocrine disruption: Case studies using example chemicals", OECD Series on Testing and Assessment, No. 181, OECD, Paris, www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2012)34&doclanguage=en.

OECD (2010a), "Guidance document on the diagnosis of endocrine-related histopathology in fish gonads", OECD Series on Testing and Assessment, No. 123, OECD, Paris, https://www.oecd.org/chemicalsafety/testing/42140701.pdf.

OECD (2010b), "Workshop report on OECD countries' activities and management of endocrine disrupters", OECD Series on Testing and Assessment, No. 118, OECD, Paris, www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/JM/MONO(2010)2&doclanguage=en.

OECD (2010c), *Detailed Review Paper (DRP) on Molluscs Life-Cycle Toxicity Testing*, OECD Series on Testing and Assessment, No. 121, OECD Publishing, Paris, https://doi.org/10.1787/9789264221468-en.

OECD (2010d), "Detailed review paper on environmental endocrine disrupter screening: The use of estrogen and androgen receptor binding and transactivation assays in fish", OECD Series on Testing and Assessment, No. 135, OECD, Paris, www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2010)34&doclanguage=en.

OECD (2009), "Guidance document for histologic evaluation of endocrine and reproductive tests in rodents", OECD Series on Testing and Assessment, No. 106, OECD, Paris, www.oecd.org/env/ehs/testing/43411534.pdf.

OECD (2008a), *Detailed Review Paper on the Use of Metabolising Systems for* In Vitro *Testing of Endocrine Disrupters*, OECD Series on Testing and Assessment, No. 97, OECD Publishing, Paris, https://doi.org/10.1787/9789264085497-en.

OECD (2008b), "Guidance document on mammalian reproductive toxicity testing and assessment", OECD Series on Testing and Assessment, No. 43, OECD, Paris, www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/JM/MONO(2008)16&doclanguage=en.

OECD (2006), "Report of the validation of the updated Test Guideline 407: Repeat dose 28-day oral toxicity study in laboratory rats", OECD Series on Testing and Assessment, No. 59, OECD, Paris, www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2006)26&doclanguage=en.

Paris, F. et al. (2002), "Phenylphenols, biphenols, bisphenol-A and 4-tert-octylphenol exhibit alpha and beta estrogen activities and antiandrogen activity in reporter cell lines", *Molecular and Cellular Endocrinology*, Vol. 193/1-2, pp. 43-49, https://doi.org/10.1016/S0303-7207(02)00094-1.

Pettersson, K., F. Delaunay and J.A. Gustafsson (2000), "Estrogen receptor beta acts as a dominant regulator of estrogen signalling", *Oncogene*, Vol. 19/43, pp. 4970-4978, https://doi.org/10.1038/sj.onc.1203828.

de Peyster, A. and E. Mihaich (2014), "Hypothesis-driven weight of evidence analysis to determine potential endocrine activity of MTBE", *Regulatory Toxicology and Pharmacology*, Vol. 69/3, pp. 48-70, https://dx.doi.org/10.1016/j.yrtph.2014.04.017.

Pickford, D.B. (2010), "Screening chemicals for thyroid-disrupting activity: A critical comparison of mammalian and amphibian models", *Critical Reviews in Toxicology*, Vol. 40/10, pp. 845-892, https://doi.org/10.3109/10408444.2010.494250.

Rodgers-Gray, T.P. et al. (2004), "Mechanisms of parasite-induced sex reversal in *Gammarus duebeni*", *International Journal for Parasitology*, Vol. 34/6, pp. 747-753, https://doi.org/10.1016/j.ijpara.2004.01.005.

Rosenberg, S.A. et al. (2017), "QSAR models for thyroperoxidase inhibition and screening of US and EU chemical inventories", *Computational Toxicology*, Vol. 4, pp. 11-21, https://doi.org/10.1016/j.comtox.2017.07.006.

Schneider, K. et al. (2009), "ToxRTool, a new tool to assess the reliability of toxicological data", *Toxicology Letters*, Vol. 189/2, pp. 138-144, https://doi.org/10.1016/j.toxlet.2009.05.013.

Schultz, T. (2010), "Adverse outcome pathways: A way of linking chemical structure to *in vivo* toxicological hazards", Chapter 14 in: Cronin, M.T.D. and J.C. Madden (eds.), *In Silico Toxicology: Principles and Applications*, Royal Society of Chemistry, http://dx.doi.org/10.1039/9781849732093-00346.

Seifert, A.W. et al. (2009), "Multiphasic and tissue-specific roles of sonic hedgehog in cloacal septation and external genitalia development", *Development*, Vol. 136/23, pp. 3949-3957, http://dx.doi.org/10.1242/dev.042291.

Simon, T.W. et al. (2014), "The use of mode of action information in risk assessment: Quantitative key events/dose-response framework for modeling the dose-response for key events", *Critical Reviews in Toxicology*, Vol. 44/Suppl 3, pp. 17-43, https://doi.org/10.3109/10408444.2014.931925.

Sitjà-Bobadilla, A. (2009), "Can myxosporean parasites compromise fish and amphibian reproduction?", *Proceedings of the Royal Society B: Biological Sciences*, Vol. 276/1669, pp. 2861-2870, http://dx.doi.org/10.1098/rspb.2009.0368.

Smagghe, G. et al. (2003), "Cultured mosquito cells *Aedes albopictus* C6/36 (Dip, Culicidae) responsive to 20-hydroxyecdysone and non-steroidal ecdysone antagonists", *Journal of Applied Entomology*, Vol. 127/3, pp. 167-173, https://doi.org/10.1046/j.1439-0418.2003.00727.x.

Solecki, R. et al. (2017), "Scientific principles for the identification of endocrine-disrupting chemicals: A consensus statement", *Archives of Toxicology*, Vol. 91/2, pp. 1001-1006, https://doi.org/10.1007/s00204-016-1866-9.

Song, Y. et al. (2017), "Ecdysone receptor agonism leading to lethal molting disruption in arthropods: Review and adverse outcome pathway development", *Environmental Science & Technology*, Vol. 51/8, pp. 4142-4157, http://dx.doi.org/10.1021/acs.est.7b00480.

Sotoca, A.M. et al. (2008), "Influence of cellular ERalpha/ERbeta ratio on the ERalpha-agonist induced proliferation of human T47D breast cancer cells", *Toxicological Sciences*, Vol. 105/2, pp. 303-311, https://doi.org/10.1093/toxsci/kfn141.

Swevers, L. et al. (2003), "A high-throughput screening system for fast detection of ecdysteroid mimetic and antagonistic substances using transformed *Bombyx mori*-derived cell lines", *The FASEB Journal*, Vol. 18/1, pp. 134-136, http://dx.doi.org/10.1096/fj.03-0627fje.

Thigpen, J.E. et al. (2013), "The estrogenic content of rodent diets, bedding, cages, and water bottles and its effect on bisphenol A studies", *Journal of the American Association for Laboratory Animal Science*, Vol. 52/2, pp. 130-141.

Thorne, N., D.S. Auld and J. Inglese (2010), "Apparent activity in high-throughput screening: Origins of compound-dependent assay interference", *Current Opinion in Chemical Biology*, Vol. 14/3, pp. 315-324, https://doi.org/10.1016/j.cbpa.2010.03.020.

Thorne, N., J. Inglese and D.S. Auld (2010), "Illuminating insights into firefly luciferase and other bioluminescent reporters used in chemical biology", *Chemical Biology*, Vol. 17/6, pp. 646-657, http://dx.doi.org/10.1016/j.chembiol.2010.05.012.

Trubiroha, A. et al. (2010), "Naturally-induced endocrine disruption by the parasite *Ligula intestinalis* (Cestoda) in roach (*Rutilus rutilus*)", *General and Comparative Endocrinology*, Vol. 166/2, pp. 234-240, https://doi.org/10.1016/j.ygcen.2009.08.010.

US EPA (2013), "FIFRA Scientific Advisory Panel Meeting held July 30-August 1, 2013 on Weight-of-Evidence: Evaluatign results of EDSP Tier 1 screening", Meeting Minutes, Environmental Protection Agency, Washington, DC, https://www.epa.gov/sites/production/files/2015-06/documents/073013minutes.pdf.

US EPA (2011), "Endocrine Disrupter Screening Program: Weight-of-evidence: Evaluating results of EDSP Tier 1. Screening to identify the need for Tier 2 testing", Environmental Protection Agency, Washington, DC.

Van Der Kraak, G.J. et al. (2014), "Effects of atrazine in fish, amphibians, and reptiles: An analysis based on quantitative weight of evidence", *Critical Reviews in Toxicology*, Vol. 44/S5, pp. 1-66, https://doi.org/10.3109/10408444.2014.967836.

Vandenberg, L.N. et al. (2016), "A proposed framework for the systematic review and integrated assessment (SYRINA) of endocrine disrupting chemicals", *Environmental Health*, Vol. 15/1, p.74, https://doi.org/10.1186/s12940-016-0156-6.

Vandenberg, L.N. et al. (2012), "Hormones and endocrine-disrupting chemicals: Low-dose effects and nonmonotonic dose responses", *Endocrine Reviews*, Vol. 33/3, pp. 378-455, https://doi.org/10.1210/er.2011-1050.

Vinggaard, A.M. et al. (2005), "Perinatal exposure to the fungicide prochloraz feminizes the male rat offspring", *Toxicological Sciences*, Vol. 85/2, pp. 886-897, https://doi.org/10.1093/toxsci/kfi150.

Vinggaard, A.M. et al. (2002), "Antiandrogenic effects *in vitro* and *in vivo* of the fungicide prochloraz", *Toxicological Sciences*, Vol. 69/2, pp. 344-353, https://doi.org/10.1093/toxsci/69.2.344.

Vuorinen, A. et al. (2015), "*In silico* methods in the discovery of endocrine disrupting chemicals", *The Journal of Steroid Biochemistry and Molecular Biology*, Vol. 153, pp. 93-101. https://dx.doi.org/10.1016/j.jsbmb.2015.08.015.

Weltje, L. et al. (2013), "Refinement of the ECETOC approach to identify endocrine disrupting properties of chemicals in ecotoxicology", *Toxicology Letters*, Vol. 223/3, pp. 291-294, https://doi.org/10.1016/j.toxlet.2013.02.019.

Wheeler, J.R. and K.K. Coady (2016), "Are all chemicals endocrine disrupters?", *Integrated Environmental Assessment and Management*, Vol. 12/2, pp. 402-403, http://dx.doi.org/10.1002/ieam.1747.

WHO (2007), *Harmonization Project Document No. 4, Part 1: IPCS Framework for Analyzing the Relevance of a Cancer Mode of Action for Humans and Case Studies, Part 2: IPCS Framework for Analyzing the Relevance of a Non-Cancer Mode of Action for Humans*, World Health Organizaton, Geneva, www.who.int/ipcs/methods/harmonization/areas/cancer_mode.pdf?ua=1.

WHO/IPCS (2009), *Principles and Methods for the Risk Assessment of Chemicals in Food*, Environmental Health Criteria 240, World Health Organization, Geneva, www.who.int/foodsafety/publications/chemical-food/en.

WHO/IPCS (2002), "Global assessment of the state-of-the-science of endocrine disrupters", Damstra, T. et al. (eds.) WHO/PCS/EDC/02.2, World Health Organization, Geneva, www.who.int/ipcs/publications/new_issues/endocrine_disruptors/en.

Zhu, H. et al. (2016), "Supporting read-across using biological data", *ALTEX – Alternatives to Animal Experimentation*, Vol. 33/2, pp. 167-182, https://doi.org/10.14573/altex.1601252.

Zoeller, R.T. et al. (2015), "A path forward in the debate over health impacts of endocrine disrupting chemicals", *Environmental Health*, Vol. 13/1, p.118, https://doi.org/10.1186/1476-069X-13-118.

C. Specific guidance for the test guidelines addressed

Introduction to specific guidelines

125. This introduction applies to all assays covered by this guidance document (GD), although it should be noted that guidance for test guidelines (TGs) that have not received full validation by the OECD, or are in the process of OECD validation, remains provisional until those assays have been fully validated with endocrine active substances (EASs) and the TG published.

126. As indicated earlier, the information given in Section C is intended to provide guidance on the interpretation of data from individual assays, and on a possible next step for obtaining additional data, if required by a given user. It is important to understand that the guidance should be used flexibly in the light of local regulatory circumstances and available data – it is not a rigid prescription, but should be considered as a decision-support tool. The guidance in Section C adopts a form of weight of evidence (WOE) approach that uses all available data and expert judgement, but it should be noted that other WOE methodologies are available (see Section B.5).

127. Discussion of each assay takes the form of textual guidance which describes the basis of the assay and any special considerations or limitations, when and why the assay is likely to be used, and what broad conclusions may be appropriate when one is in possession of positive, negative or equivocal results. This is followed by a table (known as a "building block") that elaborates that guidance for each of a number of data scenarios. Thus, for each type of assay result, the guidance varies depending on the type and amount of pre-existing data (both *in vitro* and *in vivo*). The intention has been to cover all the major possible scenarios, but the document cannot address all eventualities. Furthermore, it is implicit that expert advice will need to be consulted at many points in these building blocks – they are not recipes which can be followed blindly. Note that some scenarios are much less likely to occur than others – for example, it is unlikely (but still possible) that a higher tier procedure such as a fish life cycle test will have been performed in the absence of various screening assays. A large range of possible scenarios has, therefore, been described for the sake of completeness.

128. When considering a possible "next step" in evidence gathering that could follow from a particular result in an *in vitro* assay, guidance is given in the next section about suitable *in vivo* testing with vertebrate species. It is, of course, important to ensure that an *in vitro* assay has been conducted at realistic exposure levels before concluding on the possible need for *in vivo* testing. Some guidance is also given concerning possible mammalian tests that might be conducted following positive non-mammalian tests, and vice versa. Experience using these assays, particularly in the United States Environmental Protection Agency's (US EPA) Endocrine Disruptor Screening Program (EDSP) Tier 1, has demonstrated a high degree of cross-species sensitivity (Ankley and Gray, 2013). A positive result in an endocrine disrupter-responsive mammalian assay could be interpreted as an alert about possible related effects in non-mammalian wildlife, and the reverse also applies (although mammalian assays will often have been performed before any with non-mammals). Positive effects in mammalian assays should generally be regarded as a trigger for

some non-mammalian testing if the hazards experienced by the non-mammalian group are to be taken into account. On the other hand, insufficient data yet exist to be confident that negative mammalian data imply an absence of effects in non-mammalian wildlife, even when assuming the pathway under investigation is present and relatively well conserved.

129. It will be apparent that the underlying approach when implementing this guidance is to consider the weight of available evidence – situations in which a single assay provides conclusive evidence that a chemical is an ED may not be common, although there will be exceptions. For example, feminised anogenital distance in male offspring (observed in OECD TG 443, TG 421/422, TG 414 or TG 416) may be considered as conclusive evidence of an endocrine disrupting effect. OECD GD 43 (Guidance Document on Mammalian Reproductive Toxicity Testing and Assessment; OECD, 2008) states: "A statistically significant change in [anogenital distance] that cannot be explained by the size of the animal indicates effects of the exposure and should be used for setting the [no observed adverse effect level]." It is vital to consider all relevant data on the test chemical, including their quantity, type and quality. For example, without adequate mechanistic data from (quantitative) structure activity relationships (QSARs), *in vitro* and/or other *in vivo* assays, or from the *in vivo* assay under consideration, it will often not be possible to conclude with confidence that any apical effects have been caused by an endocrine mode of action. Indeed, any linkage between mechanistic data and apical responses will probably have to be assessed according to the weight of evidence and is unlikely to be confirmed absolutely. Another example of the use of WOE concerns *in vivo* screening assays which may indicate that a chemical can interfere with the endocrine system in intact animals, but will sometimes not be able to provide data on apical effects, or supply information which could be used on its own in a full hazard identification/characterisation of endocrine disruption. In such situations, more complete apical data may have to be obtained from a higher tier test, which will then be evaluated in conjunction with the screening data. Note, however, that negative data from a higher tier test should generally be given more weight than positive data from a lower tier screen, assuming the same class of vertebrates has been employed at both tiers, the quality of the data is good, the suspected mechanism or mode of action is adequately covered by apical endpoints, and a sensitive life stage has been used in the higher tier negative test.

130. The guidance in this document is considered reliable for estrogen/androgen/thyroid/steroidogenesis (E,A,T,S) modalities, and in certain invertebrates, juvenile hormone, ecdysone or retinoid-related activity. However, it is recognised that other endocrine modes of action exist and that some assays have not yet been fully validated (e.g. see OECD [2012]). The field of endocrine disruption continues to develop, so for that reason, this is still a "living document" which will be subject to amendment as new data are generated, new modalities are described and new assays are published as test guidelines.

131. Users of this GD should be aware that comparisons of no-effect doses or concentrations from different types of test may be very difficult or impossible. This is obvious if one is trying to compare an oral dose in a mammalian or avian test with an ambient concentration in an aquatic test. However, caution should also be used when making comparisons within these two major types of test if different methods have been used to calculate the no-effect dose or concentration (e.g. if test concentrations in one test were nominal and in the other were measured).

References

Ankley, G.T. and L.E. Gray (2013), "Cross-species conservation of endocrine pathways: A critical analysis of tier 1 fish and rat screening assays with 12 model chemicals", *Environmental Toxicology and Chemistry*, Vol. 32/5, pp. 1084-1087, https://doi.org/10.1002/etc.2151.

OECD (2012), *Detailed Review Paper on the State of the Science on Novel* In Vitro *and* In Vivo *Screening and Testing Methods and Endpoints for Evaluating Endocrine Disrupters*, OECD Series on Testing and Assessment, No. 178, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264221352-en.

OECD (2008), "Guidance document on mammalian reproductive toxicity testing and assessment", OECD Series on Testing and Assessment, No. 43, OECD, Paris, www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/JM/MONO(2008)16&doclanguage=en.

OECD *in vitro* screens (Conceptual Framework Level 2)

C.1.1. Performance-based Test Guideline for Human Recombinant Estrogen Receptor (hrER) *In Vitro* Assays to Detect Chemicals with ER Binding Affinity (OECD TG 493)

Status: Assay validated by the OECD.

Modality detected/endpoints: Binding to estrogen receptor isoforms.

Background to the assay

132. The estrogen receptor (ER) binding assay is an *in vitro* screening assay to detect substances that bind to ERs. The assay has been in use for many years and there are different variations of the protocol. Older versions of the protocol utilise rat uterine cytosol as a source of ER without further purifications of ER isoforms (e.g. US EPA OPPTS 890.1250). Binding therefore occurs to a mixture of ERα and ERβ, although the primary isoform in rat uterine cytosol is ERα. The ER binding assay was chosen to be one of the suite of assays comprising the United States Environmental Protection Agency's (US EPA) "Tier 1" and the rat uterine cytosolic assay (US EPA OPPTS 890.1250) was validated in that context (US EPA, 2009). More recent protocols do not use animals as a source of ER but use human ERα recombinant protein (hrERα). OECD TG 493 (published in July 2015) is a performance-based test guideline (PBTG) that describes two methods using human ERα:

- the Freyberger-Wilson (FW) *In Vitro* Estrogen Receptor (ER) Binding Assay Using a Full Length Human Recombinant ERα
- the Chemical Evaluation and Research Institute (CERI) *In Vitro* Estrogen Receptor Binding Assay Using a Human Recombinant Ligand Binding Domain Protein.

133. The FW *In Vitro* ERα Binding Assay uses a full-length recombinant hERα produced in and isolated from baculovirus-infected insect cells. The CERI *In Vitro* ERα Binding Assay uses a truncated ER that contains only the ligand binding domain of the hERα. Both methods were validated according to OECD principles (OECD, 2005) and are the first two in this PBTG. Performance standards to enable the development and validation of similar test methods have also been published (OECD, 2015). During validation, both methods gave similar results. There was almost 100% agreement between the two test methods, based on the classifications of all the substances up to 10^{-4} M. Each substance was also correctly classified as an ER binder or non-binder. In addition, a comparison with the rat uterine cytosol assay (US EPA OPPTS 890.1250) also showed a high degree of correlation (Laws and Wilson, 2014).

134. Binding assays provide information on the ability of a compound to interact with ERs *in vitro*? but results should not be directly extrapolated to the complex signaling and regulation of the intact endocrine system *in vivo*. Binding assays determine saturation binding and competitive binding. The saturation binding assay is used to confirm the specificity and activity of the receptor preparations, while the competitive binding experiment is used to evaluate the ability of a test chemical to bind to ER and determine IC50 (the half maximal effective concentration of an inhibitory test chemical) if possible. The assay determines the ability of a chemical to displace a radiolabeled ligand

(17β-estradiol) from ER and generally provides a positive, negative or equivocal result for the ability to bind to ER.

135. Chemicals that bind to the ER may induce hormone-dependent transcriptional activity (agonist) or block normal hormone function by preventing the endogenous hormone from binding to the receptor (antagonist). The binding assay does not distinguish between these. The hormone-binding domain of the ER is highly conserved across vertebrate species and therefore represents a simple evaluation of estrogenic potential that is relevant to many taxa. A positive result in this assay requires demonstration of a concentration response curve for the ability of the test chemical to displace radiolabelled 17β-estradiol. The concentration response curve allows the determination of potency (e.g. IC50 and relative binding affinity by comparing the log [IC50] of 17β-estradiol with that of the test chemical). OECD TG 493 provides guidance on data interpretation and criteria for assigning classification based on the competitive binding curve for a test chemical. Final classification of a test chemical is as a binder, non-binder or equivocal.

136. Occasionally, there are test chemicals where additional attention is needed to appropriately analyse and interpret the binding data. Previous studies have shown cases where the analysis and interpretation of competitive receptor binding data can be complicated by an upturn of the per cent specific binding at the highest concentrations of the test chemical. Chemicals showing this characteristic often have limited solubility. The maximum concentration of chemical to be used in the assay is 1mM. The guideline provides detailed guidance on data analysis in these circumstances.

137. The ER binding assay may suffer from variability in response if not performed exactly as stated in the protocols (e.g. if the receptor concentration in the cytosol is too low or too high, or the microtiter plates are not kept cold at all times during the experiment). Performance criteria are therefore specified in order to demonstrate that the assay is functioning correctly. Reference substances are also used on each run to demonstrate the sensitivity of the experiment (reference standard: 17β-estradiol; weak positive control: norethynodrel or norethindrone; and negative control: octyltriethoxysilane). Compliance with the performance criteria should be checked before evaluating results from an assay run to ensure that most have been met. Small deviations are unlikely to have compromised the assay? but judgement should be made on a case-by-case basis. It is recommended that laboratory proficiency be demonstrated by the periodic use of proficiency substances. These are a subset of the substances provided in the performance standards for the ER binding assays (OECD, 2015). They represent the classes of chemicals commonly associated with ER binding activity, exhibiting a suitable range of potency expected for ER binding (i.e. strong to weak) and non-binders (i.e. negatives).

When/why the assay may be used

138. Although the ER binding assay may be used at any stage in the hazard assessment process, the most likely use scenario is during initial assessment of chemicals for their ability to interact with endocrine systems *in vitro* (i.e. E,A,T,S modalities). Assays for interaction with other modalities (e.g. androgen receptor [AR] and steroidogenesis interference), are likely to be conducted at the same time so that all results can be considered together. Thyroid hormone receptor (TR) and other assays concerning mechanisms of thyroid disruption may be conducted, but the methods for these are not in common use and are not validated (see Section A.6). The ER binding assay does not include the use of a xenobiotic metabolising system, but consideration should be given to the inclusion of this (Jacobs et al., 2008; OECD, 2008) depending on the circumstances (e.g.

if the metabolism of a chemical is unknown), although the methods for inclusion of xenobiotic metabolising systems are not yet validated (see Paragraph 50). Alternatively, for a chemical with known metabolites, these could also be tested in the ER binding assay. Another use scenario may be following effects obtained in higher tier tests, for example delayed or accelerated puberty onset in females, which could be indicative of an effect mediated by ER. Selection of the most appropriate tests has to be on a case-by-case basis, but also considering the need to minimise animal testing. A further example could be results obtained in other apical assays, e.g. in OECD TG 408 (90-Day Toxicity Test), where effects on reproductive organs could be investigated further by testing in the ER binding assay in combination with AR- and steroidogenesis-based assays.

Introduction to the table of scenarios

139. Table C.1.1 gives guidance on a further step to take in the event of a positive (+) or negative (-) result in the ER binding assay and in the presence of positive (+), negative (-) or equivocal/absent (Eq/0) existing results. "Existing results" are subdivided into "mechanism" and "effects" data (third and fourth columns). The table is divided horizontally into a series of scenarios that represent all the combinations of these events.

140. The results of the ER binding assay are given in the second column. Criteria for positive, negative and equivocal results are given in OECD TG 493. A test chemical is considered to be a binder if a binding curve can be fit and the lowest point on the response curve within the range of the data is less than 50% and a log IC50 can be obtained. A positive result should be obtained in at least two out of three independent test runs. It is also important that quality and proficiency criteria are demonstrated for both positive and negative results.

141. Equivocal results for the guideline are not included in the table because these data generally require further interrogation about the result itself. In the event of an equivocal result, the considerations mentioned above about control quality and proficiency criteria should be taken into account and further investigations made. Equivocal results at high concentrations may indicate solubility issues.

Existing data to be considered

142. Existing "mechanism" *in vitro* data are assumed to be available from AR-based assays (Level 2) and the Steroidogenesis Assay. Assays may also be available for interference with thyroid modalities. The ER binding assay is most likely to be performed before the ER Stably Transfected Transactivation Assay (STTA – OECD TG 455) and so the ability of the chemical to affect ER-mediated gene expression may not be known. In practice, it is possible that data from some or all of these assays may not be available, so judgement will need to be used to decide which assays to perform. The ER binding assay and ER STTA both provide data about the intrinsic ability of a chemical to interact with ER, but each has their own advantages and disadvantages. The ER binding assay will not distinguish between agonists and antagonists whilst some chemicals testing positive in the ER STTA assay may have affected the reporter gene activity through non-ER related mechanisms. Consistent results in both assays give more confidence in the presence or absence of an ER-related mode of action (MOA).

143. Existing "effects" data refer to *in vivo* effects "of concern" (i.e. data from Level 4 or 5 vertebrate assays). These may come from varied sources and will depend on the type of substance (e.g. new chemicals, high production volume [HPV] chemical, pesticide). Thus,

available data may range from repeated dose toxicity studies (28-day, 90-day), combined repeat dose/reproductive screening assays or fish screening assays, to chronic toxicity studies and multigeneration reproductive tests in vertebrate species. Some studies fail to identify EASs that weakly affect estrogen or androgen receptors, as was demonstrated on the basis of data generated in the validation process of OECD TG 407 assay with endocrine endpoints. In this validation only moderate endocrine disruptors, such as nonylphenol and DDE, and strong EDs such as ethinylestradiol and flutamide (acting via ER and AR, respectively) were detected. Thus OECD TG 407 cannot be regarded as a screening assay for endocrine activity. This means that when a relatively insensitive test is positive for both endocrine-specific and apical endpoints, this should be taken as an indication that the substance is a potential ED. Caution should be exercised, however, because endocrine endpoints may be impacted secondary to non-endocrine toxicity and *in vivo* apical endpoints can be affected by many modes of action, including endocrine modalities. Data may also be available on effects in non-mammalian species, although caution should be used when extrapolating between taxa. A chemical causing endocrine effects in fish or amphibians (for example, OECD TG 240 or TG 241) may also have endocrine effects in mammals, but the physiological consequences of the effects are likely to be different.

144. Data may also be available from Level 3 mammalian assays (Hershberger Bioassay [H] and Uterotrophic Bioassay [UT]) but as the UT assay primarily detects (*in vivo*) the same modality as ER binding, it is unlikely that it would be conducted before ER binding. An Amphibian Metamorphosis Assay (AMA) may also be available, but as this test primarily detects thyroid disruption in amphibians, it is unlikely to provide useful data for E-modalities.

145. When considering the results of the ER binding assay, all available data should be used in order to reach a conclusion and a WOE approach taken. This may include high throughput screening (HTS) data, read-across data from structural analogues and Quantitative structure activity relationship (QSAR). Several QSAR models for ER and AR binding/activation are now available (see Sections B.1.1.1 and B.1.1.2).

Scenarios: Positive and negative results combined with existing data

146. The scenarios (A to R) presented in Table C.1.1 represent all the possibilities of positive or negative results in combination with the presence or absence of existing data. The action taken will also depend on the regulatory environment, but the considerations given here are generally science based. Although OECD TG 493 uses hrER, the well-conserved nature of ER across taxa should be a strong indication that results in this assay are relevant to other vertebrate species. Wherever possible, the recommended "next step which could be taken" avoids unnecessary animal testing. However, sometimes conducting an animal test will be indicated and then the relevance of species, strain, exposure route and species-specific metabolism should always be considered. Further considerations specific to each scenario are given in the table.

147. Scenarios A to C represent positive results in the ER binding assay in the presence of positive *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. A positive result in an ER binding assay is strong evidence for (anti)estrogenic activity that may or may not be supported by the *in vivo* effects data. In the case of positive *in vivo* effects data, there may be sufficient evidence to conclude concern for endocrine disruption and therefore no need for further screening. *In vivo* assays/tests with negative results should be interpreted with caution as they may either indicate that the tests used do not have sufficient power to detect weak effects or, alternatively, that the effects do not present a

concern for endocrine disruption. Generally, a conclusion of lack of concern for endocrine disruption in the presence of positive effects data may only be made given adequate Level 5 assays. Information on some endocrine-related tumours may be detected more comprehensively in carcinogenicity studies (OECD TG 451/453) (Level 4); for example, detection of certain types of thyroid tumors in the absence of reproductive or developmental effects, as well as substances causing tumours in other endocrine-sensitive tissues. MOA data to provide a clear interpretation may be required by some regulatory agencies.

148. Scenarios D to F represent positive results in the ER binding assay in the presence of negative *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. Unless the metabolic profile of the test substance is known, one option may be to conduct these *in vitro* assays with an added metabolising system. If the metabolic profile is known, then an *in vivo* test may be advisable. The choice of tests will depend on the available *in vivo* effects data. As in Scenarios A to C, *in vivo* assays/tests with negative results should be interpreted with caution as they may either indicate that the tests used do not have sufficient power to detect weak effects or, alternatively, that the effects do not present a concern for endocrine disruption. Generally, a conclusion of lack of concern for endocrine disruption in the presence of positive effects data (Scenario E) may only be made given adequate Level 5 assays. Information on some endocrine-related tumours may be detected more comprehensively in carcinogenicity studies (OECD TG 451/453) (Level 4); for example, detection of certain types of thyroid tumours in the absence of reproductive or developmental effects, as well as substances causing tumours in other endocrine-sensitive tissues. MOA data to provide a clear interpretation may be required by some regulatory agencies.

149. Scenarios G to I represent positive results in the ER binding assay in the presence of various combinations of missing or equivocal data. The next step to take in these eventualities will depend on the nature of the other available data and the jurisdiction in which it is being used. In some cases, equivocal data may be viewed as positive whilst in others it may or may not contribute to the weight of evidence. The interpretation may also depend on the MOA in question and why the data are considered equivocal, e.g. a study that is equivocal for thyroid effects may still be of value in evaluating (anti)androgenic effects. In all three scenarios, the recommended first step is to obtain reliable mechanistic (*in vitro*) data rather than proceed directly to *in vivo* testing. Equivocal and missing data are alternative scenarios and two possibilities for the next step are given in most cases, but the nature of equivocal data means that decisions need to be taken on a case-by-case basis. In all cases, the role of metabolism, route of exposure and data from structural analogues should be considered before deciding on the next step. In some cases, it may be necessary to conduct *in vivo* tests and some guidance is given in the final column. As above, generally a conclusion of lack of concern for endocrine disruption in the presence of some positive effects data (Scenario H) may only be made given adequate Level 5 assays. Information on some endocrine-related tumours may be detected more comprehensively in carcinogenicity studies (OECD TG 451/453) (Level 4); for example, detection of certain types of thyroid tumours in the absence of reproductive or developmental effects, as well as substances causing tumours in other endocrine-sensitive tissues. MOA data to provide a clear interpretation may be required by some regulatory agencies.

150. Scenarios J to L represent negative results in the ER binding assay in the presence of positive *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. The limitations of the ER binding assay should be considered first (e.g. lack of metabolic activation, possible involvement of other binding proteins). The positive *in vitro* mechanistic data indicates possible alternative A,T,S mechanisms. To confirm lack of ER-

related activity in the presence of *in vivo* data, an ER STTA could be performed. Otherwise *in vivo* tests will confirm or refute E,A,T,S activity.

151. Scenarios M to O represent negative results in the ER binding assay in the presence of negative *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. Negative results for all tests (Scenario N) may be sufficient to enable a conclusion of no concern for endocrine disruption. This will depend on the weight of evidence and may not be possible in some cases. However, in the presence of negative data from robust Level 4 and 5 assays, further animal testing is probably not justified. The limitations of the ER binding assay should also be considered (as described for Scenarios J to L). To confirm lack of ER-related activity in the presence of *in vivo* data, an ER STTA could be performed. Otherwise *in vivo* tests will confirm or refute E,A,T,S activity (Scenarios M and O).

152. Scenarios P to R represent negative results in the ER binding assay in the presence of various combinations of missing or equivocal data. The limitations of the ER binding assay should be considered first (as described for Scenarios J to L). As with the positive result scenarios above (Paragraph 145), the next step to take for Scenarios P to R will have to be decided on a case-by-case basis. However, the recommended first step is generally to obtain reliable mechanistic (*in vitro*) data rather than proceed directly to *in vivo* testing. In all cases, the role of metabolism, route of exposure and data from structural analogues should be considered before deciding on the next step.

153. In all scenarios (A to R), the next step to take to strengthen weight of evidence will depend on the existing information. Table C.1.1 is meant to provide a succinct guide and may not cover all circumstances or possibilities. The scenarios may also suggest that chemicals have simple or single MOA, when in practice they may have multiple endocrine and non-endocrine MOA. In some cases, for example, two opposite modes of simultaneous action (e.g. estrogenic and anti-estrogenic) could, depending on dose, lead to a minimisation or abolition of effects, while in others two different MOA (e.g. estrogenic and anti-androgenic) could potentially reinforce effects. Endocrine pathways interact, mixed effects are common and there are many pathways that cannot be distinguished with currently available TGs. If multiple MOA are suspected, either from the existing results or based on QSAR/read-across/integrated approaches, this should be investigated further if needed for regulatory decision making.

154. In general, a decision about whether or not to conduct *in vivo* vertebrate tests will depend on the weight of evidence of new and existing data. If most available data (e.g. the results of the ER binding assay, results from an ER STTA assay, predictions from QSARs, "read-across" from data on similar substances and results from mammalian *in vivo* assays) suggest that the substance has the potential to cause endocrine disruption via the estrogen receptor (i.e. the level of suspicion about endocrine disrupting action is high – corresponding to Scenario A), then consideration should be given to the conduct of a higher level test.

155. For non-mammalian wildlife species, higher level tests with fish or amphibians (i.e. OECD TG 234, TG 240, TG 241) are recommended. Choice about which of these tests is most appropriate will be driven *inter alia* by MOA considerations, and by whether multigeneration effects are to be expected. Such tests are unlikely to be needed if exposure of the natural environment is not expected. On the other hand, if available data only raise a low or moderate level of suspicion about endocrine disrupting action (e.g. the data appear to conflict with each other), then consideration should be given to the conduct of a fish or amphibian screen (e.g. juvenile medaka anti-androgen screening assay; EASZY; *Xenopus* embryonic thyroid signalling assay; OECD TG 231, TG 229 or TG 230). There are fewer

options available for invertebrates, but if ecdysteroid or juvenile hormone activity are suspected in arthropods (e.g. from a screening test with short-term juvenile hormone activity screening assay), various higher level tests are available, including OECD GD 201, the *Daphnia* Multigeneration Test and TG 233.

156. For mammals, similar considerations apply, but lower level tests (e.g. Level 3 or 4) should be conducted before higher level tests in order to avoid unnecessary animal usage, unless it is apparent that a Level 5 test will be required anyway or will be needed to establish the evidence to conclude on ED properties. At Level 5, the Extended One-Generation Reproductive Toxicity Study (OECD TG 443) is the most sensitive reproduction assay for detecting endocrine disruption because it includes evaluation of a number of endocrine endpoints not included in the two-generation study (OECD TG 416) adopted in 2001. It is recognised, however, that some jurisdictions may require a two-generation study.

References

Jacobs, M.N. et al. (2008), "The use of metabolising systems for *in vitro* testing of endocrine disrupters", *Current Drug Metabolism*, Vol. 9/8, pp. 796-826.

Laws, S. and V. Wilson (2014), "Integrated summary report: Validation of two binding assays using human recombinant estrogen receptor alpha (hrER alpha)", Environmental Protection Agency, Washington, DC, https://cfpub.epa.gov/si/si_public_record_report.cfm?dirEntryId=307944.

OECD (2015), "Performance standards for the human recombinant estrogen receptor binding assay (intended for the developers of new or modified similar test methods)", OECD Series on Testing and Assessment, No. 222, OECD, Paris, www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/JM/MONO(2015)29&doclanguage=en.

OECD (2008), *Detailed Review Paper on the Use of Metabolising Systems for* In Vitro *Testing of Endocrine Disrupters*, OECD Series on Testing and Assessment, No. 97, OECD Publishing, Paris, https://doi.org/10.1787/9789264085497-en.

OECD (2005), "Guidance document on the validation and international acceptance of new or updated test methods for hazard assessment", OECD Series on Testing and Assessment, No. 34, OECD, Paris, www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?doclanguage=en&cote=env/jm/mono(2005)14.

US EPA (2009), "Integrated summary report for validation of an estrogen receptor binding assay using rat uterine cytosol as source of receptor as a potential screen in the Endocrine Disrupter Screening Program Tier 1 battery", Environmental Protection Agency, Washington, DC.

Table C.1.1. **Performance-Based Test Guideline for Human Recombinant Estrogen Receptor (hrER) *In Vitro* Assays to Detect Chemicals with ER Binding Affinity (OECD TG 493): Guidance for scenarios of combinations of results with existing data**

This table represents possible conclusions to be drawn from assay data, and a next step which could be taken if further evidence is required about possible endocrine disrupting properties and/or effects. The guidance offered is not meant to be prescriptive, but provides science-based considerations. It encourages the use of all available data and expert judgement in a weight of evidence approach. Regional and national interpretation of results and "next steps" may vary.

The conclusions are grouped into a series of scenarios (A-R), each scenario representing a different combination of assay results, existing *in vitro* data and existing *in vivo* data. The symbol "+" indicates that the data in question represent a positive result, "-" indicates a negative result, and "Eq/0" indicates that the data are either equivocal or are not available.

Existing results: * "Mechanism (*in vitro* mechanistic data)" assumes that mechanistic data are available from androgen receptor-based assays and the Steroidogenesis Assay (Level 2). Thyroid hormone receptor and other assays concerning mechanisms of thyroid disruption may be available, but they are not in common use. In practice, data from all assays may not be available and therefore this must be taken into account when deciding on the "next step". Quantitative structure activity relationship (QSAR) predictions of estrogen and androgen binding/activation may be made for some substances. Data from the ER STTA are assumed to be unavailable, but a decision about the next step to be taken will also depend on the availability of this assay and QSAR data.

Existing results: ** "Effects (*in vivo* effects of concern)" assumes various information, such as data from repeat dose oral toxicity studies, reproduction/developmental toxicity screening tests, read-across from analogues, will be available.

Scenarios	Result of hrER binding assay	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
A	+	+	+	Interaction with estrogen receptor (ER) combined with effects on AR/T/S and potential for adverse effects via multiple mechanisms.	Perform ER STTA or assay from Levels 3-5, e.g. Uterotrophic Bioassay (UT) (Level 3) or female Peripubertal (PP) Assay (Level 4) or Extended One-Generation Reproductive Toxicity Study (EOGRTS) or two-generation assays or partial/full non-mammalian wildlife life cycle tests, e.g. TG 241 and TG 240 (Level 4/5).	Binding to hrER indicates strong probability of binding to ERs in other taxa. If existing data are from Level 5, there may be sufficient information to conclude evidence of concern for endocrine disruption (the EOGRTS provides the most information; however, for endocrine disrupting chemicals [EDCs] with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from Level 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism, although some Level 4 assays (e.g. TG 234) may be sufficient for this purpose. If existing data are from UT assays, then Level 4 assays will provide data on multiple modalities. Consider route of exposures for existing effects data and possible transformation products and implications of absorption, distribution, metabolism and excretion (ADME) characteristics of the chemical. The chemical itself may give positive *in vitro* results and *in vivo* results, but may also be metabolised to a metabolite that also has positive results *in vitro* and *in vivo*. However, note that uptake and metabolism of chemicals can be different between vertebrate wildlife species.
B	+	+	–	Interaction with ER combined with effects on AR/T/S but effects not detected in *in vivo* studies. Interaction with ER does not result in adverse effects in the selected species under the conditions of the test. Metabolic differences may explain *in vitro*/*in vivo* differences.	Perform binding assay or ER STTA with added metabolising system or Assay from Levels 3-4, e.g. UT assay or fish screen (OECD TG 229/230) (Level 3) or female PP assay (Level 4).	Binding to hrER indicates strong probability of binding to ERs in other taxa. If existing data are from an adequate Level 5 study there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from Level 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism, although some Level 4 assays (e.g. TG 234) may be sufficient for this purpose. If existing data are from UT assays, then Level 4 assays will provide data on multiple modalities. Consider route of exposures for existing effects data and possible transformation products and implications of ADME characteristics of the chemical. The chemical itself may give positive *in vitro* results but may not be absorbed or may be metabolised to an inactive metabolite *in vivo*. However, note that uptake and metabolism of chemicals can be different between mammalian and non-mammalian wildlife species.
C	+	+	Eq/0	Interaction with ER combined with effects on AR/T/S but no or equivocal data from *in vivo* studies. Interaction with ER may not result in adverse effects in the selected species under the conditions of the test.	Perform ER STTA or Perform assay from Levels 3-4 e.g. UT assay or fish screen (OECD TG 229/230) (Level 3) or female PP assay (Level 4).	Binding to hrER indicates strong probability of binding to ERs in other taxa. Consider route of exposures for equivocal existing effects data and possible implications of ADME characteristics of the chemical. Equivocal results may indicate chemical has multiple modes of action (MOA). Check data on chemical analogues.

Scenarios	Result of hrER binding assay	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
D	+	–	+	Interaction with ER and potential for adverse effects.	Perform ER STTA or Perform assay from Levels 3-4, e.g. UT assay or fish screen (OECD TG 229/230) (Level 3) or female PP assay (Level 4).	Binding to hrER indicates strong probability of binding to ERs in other taxa. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude evidence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from Level 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism, although some Level 4 assays (e.g. TG 234) may be sufficient for this purpose. If existing data are from UT assay, then Level 4 assays will provide data on multiple modalities. A positive result could have arisen from other (E,A,T,S or non-E,A,T,S) mechanisms, e.g. hypothalamic/pituitary/gonadal (HPG) axis.
E	+	–	–	Interaction with ER but effects not detected in *in vivo* studies. Interaction with ER does not result in adverse effects in the selected species under the conditions of the test. Metabolic differences may explain *in vitro*/*in vivo* differences.	Perform binding assay or ER STTA with added metabolising system or Assay from Levels 3-4, e.g. UT assay or fish screen (OECD TG 229/230) (Level 3) or female PP assay (Level 4).	Binding to hrER indicates strong probability of binding to ERs in other taxa. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information); however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from Level 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism, although some Level 4 assays (e.g. TG 234) may be sufficient for this purpose. If existing data are from UT assays, then Level 4 assays will provide data on multiple modalities. Consider route of exposures for existing effects data and possible transformation products and implications of ADME characteristics of the chemical. The chemical itself may give positive *in vitro* results but may not be absorbed or may be metabolised to an inactive metabolite *in vivo*. However, note that uptake and metabolism of chemicals can be different between mammalian and non-mammalian wildlife species. Check data on chemical analogues.
F	+	–	Eq/0	Interaction with ER but no or equivocal data from *in vivo* studies.	Perform ER STTA or assay from Levels 3-4, e.g. UT assay or fish screen (OECD TG 229/230) (Level 3) or female PP assay (Level 4).	Binding to hrER indicates strong probability of binding to ERs in other taxa. Consider route of exposures for equivocal existing effects data and possible implications of ADME characteristics of the chemical. Equivocal results may indicate chemical has multiple MOA. The chemical itself may give positive *in vitro* results but may not be absorbed or may be metabolised to an inactive metabolite *in vivo*. Check data on chemical analogues.

Scenarios	Result of hrER binding assay	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
G	+	Eq/0	+	Interaction with ER and potential for adverse effects via ER. May act via E,A,T,S mechanism and may or may not require metabolic activation.	Perform ER STTA.	Binding to hrER indicates strong probability of binding to ERs in other taxa. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude evidence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from Level 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism, although some Level 4 assays (e.g. TG 234) may be sufficient for this purpose. If existing data are from UT assays, then Level 4 assays will provide data on multiple modalities. Check data on chemical analogues. Further mechanistic studies may help determine MOA. Equivocal results may indicate chemical has multiple MOA. A positive result could have arisen from other (E,A,T,S or non-E,A,T,S) mechanisms, e.g. HPG axis.
H	+	Eq/0	–	Interaction with ER but effects not detected in *in vivo* studies. Interaction with ER does not result in adverse effects in the selected species under the conditions of the test. Metabolic differences may explain *in vitro*/*in vivo* differences.	For the "0" scenario, perform ER STTA. For the "Eq" scenario, perform ER STTA.	Binding to hrER indicates strong probability of binding to ERs in other taxa. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from Level 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism, although some Level 4 assays (e.g. TG 234) may be sufficient for this purpose. If existing data are from UT assays, then Level 4 assay will provide data on multiple modalities. Consider route of exposures for existing effects data and possible transformation products and implications of ADME characteristics of the chemical. The chemical itself may give positive *in vitro* results but may not be absorbed or may be metabolised to an inactive metabolite *in vivo*. However, note that uptake and metabolism of chemicals can be different between mammalian and non-mammalian wildlife species. Check data on chemical analogues. Further mechanistic studies may help determine MOA. Equivocal results may indicate chemical has multiple MOA.

Scenarios	Result of hrER binding assay	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
I	+	Eq/0	Eq/0	Interaction with ER with unknown potential for effects in *in vivo* studies. May act via ER and may or may not require metabolic activation. Unknown potential for adverse effects.	For the "0" scenario, ER STTA with added metabolising system. For the "Eq" scenario, UT assay or fish screen (OECD TG 229/230) (Level 3) if existing data indicate this is needed.	Binding to hrER indicates strong probability of binding to ERs in other taxa. Consider route of exposures for equivocal existing effects data and possible implications of ADME characteristics of the chemical. The chemical itself may give positive *in vitro* results but may not be absorbed or may be metabolised to an inactive metabolite *in vivo*. Check data on chemical analogues. Equivocal results may indicate chemical has multiple MOA.
J	–	+	+	No evidence for interaction with ER. Effects on AR/T/S and potential for adverse effects via E,A,T,S mechanisms.	Perform ER binding assay or ER STTA with added metabolising system or Perform assay from Levels 3-4, e.g. UT assay or fish screen (OECD TG 229/230) (Level 3) or female PP assay (Level 4).	Lack of binding to hrER indicates binding to ERs in other taxa is unlikely. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude concern for endocrine disruption. If existing data are from Level 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism, although some Level 4 assays (e.g. OECD TG 234) may be sufficient for this purpose. If existing data are from UT assays, then Level 4 assays will provide data on multiple modalities. Consider route of exposures for existing effects data and possible transformation products and implications of ADME characteristics of the chemical. Check data on chemical analogues. Further mechanistic studies may help determine MOA.
K	–	+	–	No evidence for interaction with ER. Effects on AR/T/S but effects not detected in *in vivo* studies. Metabolic differences explain *in vitro*/*in vivo* A,T,S differences.	Perform assay from Levels 3-4, e.g. fish screen (OECD TG 229/230) (Level 3) or male or female PP assay (Level 4).	Lack of binding to hrER indicates binding to ERs in other taxa is unlikely. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from Level 4 vertebrate wildlife assays then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism, although some Level 4 assays (e.g. OECD TG 234 [FSDT]) may be sufficient for this purpose. If existing data are from UT assays, then Level 4 assays will provide data on multiple modalities. Metabolic deactivation of chemical may occur *in vivo* so that potential *in vitro* A,T,S activity is not realised. Consider possible routes of exposure, implications of metabolism.
L	–	+	Eq/0	No evidence for interaction with ER. Effects on AR/T/S but effects not detected in *in vivo* studies. Unknown potential for adverse effects. Metabolic differences explain *in vitro*/*in vivo* A,T,S differences.	Perform assay from Levels 3-4, e.g. fish screen (OECD TG 229/230) (Level 3), or male or female PP assay (Level 4).	Lack of binding to hrER indicates binding to ERs in other taxa is unlikely. Metabolic deactivation of chemical may occur *in vivo* so that potential *in vitro* A,T,S activity is not realised. Consider possible routes of exposure, implications of metabolism. Check data on chemical analogues. Equivocal results may indicate chemical has multiple MOA.

Scenarios	Result of hrER binding assay	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
M	–	–	+	No evidence for interaction with ER. Metabolic differences may explain *in vitro*/*in vivo* differences. Effects seen in existing studies are via non-E,A,T,S or non-endocrine mechanisms.	Perform ER STTA with added metabolising system or Perform assay from Levels 3-4, e.g. fish screen (OECD TG 229/230) (Level 3) or male or female PP assay (Level 4).	Lack of binding to hrER indicates binding to ERs in other taxa is unlikely. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude concern for endocrine disruption. If existing data are from Level 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism, although some Level 4 assays (e.g. OECD TG 234 [FSDT]) may be sufficient for this purpose. If existing data are from UT assays, then Level 4 assays will provide data on multiple modalities. Metabolic activation of chemical may occur *in vivo*. Consider possible routes of exposure, implications of metabolism. Check data on chemical analogues. Further mechanistic studies may help determine MOA.
N	–	–	–	No evidence for interaction with ER. No evidence of adverse effects.	Possibly no need for further testing. If there is uncertainty, may perform assay from Levels 3-4, e.g. fish screen (OECD TG 229/230) (Level 3) or male or female PP assay (Level 4).	Lack of binding to hrER indicates binding to ERs in other taxa is unlikely. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude evidence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from Level 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism, although some Level 4 assays (e.g. OECD TG 234 [FSDT]) may be sufficient for this purpose. If existing data are from UT assays, then Level 4 assays will provide data on multiple modalities. Check data on chemical analogues.
O	–	–	Eq/0	No evidence for interaction with ER. Metabolic differences may explain *in vitro*/*in vivo* differences. Unknown potential for adverse effects via other mechanisms.	Perform ER STTA with added metabolising system or Fish screen (OECD TG 229/230) (Level 3) or male or female PP assay (Level 4) if existing data indicate this is needed.	Lack of binding to hrER indicates binding to ERs in other taxa is unlikely. Consider possible routes of exposure, implications of metabolism. Check data on chemical analogues.

Scenarios	Result of hrER binding assay	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
P	–	Eq/0	+	No evidence for interaction with ER. Metabolic differences may explain *in vitro*/*in vivo* differences. Unknown potential for adverse effects via other mechanisms.	Perform ER STTA with added metabolising system.	Lack of binding to hrER indicates binding to ERs in other taxa is unlikely. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude concern for endocrine disruption. If existing data are from Level 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism, although some Level 4 assays (e.g. OECD TG 234 [FSDT]) may be sufficient for this purpose. If existing data are from UT assays, then Level 4 assays will provide data on multiple modalities. Consider possible routes of exposure, implications of metabolism. Check data on chemical analogues. Further mechanistic studies may help determine MOA. Equivocal results may indicate chemical has multiple MOA.
Q	–	Eq/0	–	No evidence for interaction with ER. No evidence of adverse effects.	Perform ER STTA with added metabolising system.	Lack of binding to hrER indicates binding to ERs in other taxa is unlikely. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from Level 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism, although some Level 4 assays (e.g. TG 234 [FSDT]) may be sufficient for this purpose. If existing data are from UT assays, then Level 4 assays will provide data on multiple modalities. Check data on chemical analogues. Further mechanistic studies may help determine MOA.
R	–	Eq/0	Eq/0	No evidence for interaction with ER. Unknown potential for adverse effects via other mechanisms.	For the "0" scenario, perform ER STTA with added metabolising system or Perform assay from Levels 3-4, e.g. UT assay or fish screen (OECD TG 229/230) (Level 3) or female PP assay (Level 4).	Lack of binding to hrER indicates binding to ERs in other taxa is unlikely. Consider possible routes of exposure, implications of metabolism. Check data on chemical analogues. Further mechanistic studies may help determine MOA. Equivocal results may indicate chemical has multiple MOA.

C.1.2. Performance-Based Test Guideline for Stably Transfected Transactivation *In Vitro* Assays to Detect Estrogen Receptor Agonists and Antagonists (ER STTA) (OECD TG 455)

Status: Assay validated by the OECD.

Modality detected/endpoints: Activation of reporter gene linked to ER (agonist assay). Inhibition of activation of reporter gene linked to ER (antagonist assay).

Background to the assay

157. The Stably Transfected hERα Transcriptional Activation Assay (ER STTA) is an *in vitro* screening assay to detect substances that bind to hER and activate the transcription of estrogen responsive genes. It is an *in vitro* tool that provides mechanistic data. Several ER STTA assays in common use can be found in the literature (e.g. Andersen et al. [2002]; Escande et al. [2006]; Takeyoshi et al. [2002]; Du et al. [2010]; Witters et al. [2010]). One of the first versions of this assay used was the "yeast estrogen screen" (Routledge and Sumpter, 1996; Odum et al., 1997; Sheahan et al., 2002) which is still widely used for screening of environmental samples. Some variants of the yeast-based assays (*Saccharomyces cerevisiae* and *Arxula adeninivorans*) carrying the human ERα-receptor have recently been standardised within the ISO 19040 series: Determination of the estrogenic potential of water and waste water, together with human cell line-based transactivation assays (see Paragraph 20 in Section A). The guidance in this building block can be cautiously used for these assays.

The previous version of OECD TG 455 described agonist interaction with hERα utilising the hERα-HeLa-9903 cell line (derived from a human cervical tumor) and a luciferase reporter gene. Antagonist interaction was provided in a separate guidance. Another OECD TG (457) for the ER STTA assay was adopted in October 2012 and described the agonist and antagonist assay using the BG1Luc cell line. However, OECD TG 455 was revised (September 2016) to become a performance-based test guideline (PBTG) and include both methods. OECD TG 457 became redundant in January 2018. The most recent version of OECD TG 455 is a PBTG that describes the two ER STTA methods. Agonism and antagonism assays are included in both test methods. The two methods are:

- the Stably Transfected TA assay using the (h) ERα-HeLa-9903 cell line
- the VM7Luc ER STTA assay using the VM7Luc4E2 cell line which predominately expresses hERα with some contribution from hERβ.

158. Note that the VM7Luc4E2 cell line was originally designated as the BG1Luc cell line. However, in July 2016 in-depth analysis of the cells revealed that the cell line used to develop the assay was not the BG1 human ovarian carcinoma cell line, but was instead a variant of the MCF7 human breast cancer cell line. The designation of the cell line was

then changed accordingly. This does not affect published validation studies nor the utility and application of this assay for screening of estrogenic/antiestrogenic test chemicals.

159. Both test methods use a human cell line stably transfected with ERα, the main difference being that the VM7Luc4E2 cell line also expresses a minor amount of endogenous ERβ. Both assays use a luciferase reporter gene. The two test methods are the first to be included in the PBTG, other test methods are in validation and may be included later. Performance standards to enable the development and validation of similar test methods have also been published (OECD, 2012a; 2012b; 2012c). During validation, both methods gave similar results. There was almost 100% agreement between the two test methods, based on the classifications of all the substances except for one (mifepristone) for the antagonist assay, and each substance was correctly classified as an ER agonist/antagonist or negative.

160. OECD TG 455 provides a positive or negative result for the ability of a chemical to induce hERα-mediated transactivation of luciferase gene expression (agonist assay) compared to a vehicle control. The antagonist assay determines whether a reduction in response occurs when cells are co-exposed to a chemical and a potent estrogen agonist compared to the potent estrogen agonist alone. Any reduction in response must occur in the absence of cytotoxicity. There is currently no universally agreed method for interpreting ER STTA data. However, both qualitative (e.g. positive/negative) and/or quantitative (e.g. EC50, PC50, IC50) assessments of ER-mediated activity should be based on empirical data and sound scientific judgment. Where possible, positive results should be characterised by both the magnitude of the effect as compared to the vehicle (solvent) control or reference estrogen and the concentration at which the effect occurs (e.g. an EC50, PC50, RPCMax, IC50, etc.).

161. Consistent results should be achieved in at least two out of two or three runs of the assay. To be acceptable, the results should also meet the performance standards given in the assay. Small deviations are unlikely to have compromised the assay, but judgement should be made on a case-by-case basis.

162. Both test methods showed a high degree of sensitivity and specificity for both estrogenic and antiestrogenic responses in the validation studies when compared with the ER binding, UT assays and published reports determining the ability of a chemical to elicit an equivalent response *in vivo*. OECD TG 455 requires strict control of assay conditions in order to maintain the accuracy and reliability of response. Demonstration of laboratory proficiency with proficiency chemicals is required at the outset: 14 for the agonist assay and 10 for the antagonist assay. These chemicals are a subset of the substances provided in the performance standards for the ER STTA (OECD, 2012b; 2012c), represent the classes of chemicals commonly associated with ER agonist or antagonist activity, exhibit a suitable range of potency expected for ER agonists/antagonists (i.e. strong to weak), and include negatives. Periodic testing with proficiency chemicals should also be carried out. In addition, each experiment requires reference chemicals. For example, the ERα-HeLa-9903 cell line test method requires for the agonist assay: a strong estrogen (E2), a weak estrogen (17α-estradiol), a very weak agonist (17α-methyltestosterone) and a negative substance (corticosterone); and for the antagonist assay: a positive substance (tamoxifen) and a negative substance (flutamide). In the assay, each plate requires positive and vehicle controls. Criteria for the degree of response with these chemicals are given in the TG. The assay also requires a minimum of 80% cell viability; this is critical for the antagonist assay where positive results can only be demonstrated in the absence of cytotoxicity. Compliance

with the quality control criteria and with the performance criteria should be accepted before evaluating results from this assay.

163. A limitation of OECD TG 455, related to the reporter gene luciferase, is the potential for chemicals to increase chemiluminescence via non-ERα mechanisms, thus possibly giving a false positive response. This has been reported for certain phytoestrogens such as genistein and daidzein but not for industrial chemicals (Kuiper et al., 1998; Escande et al., 2006). This may be recognised by incomplete or unusual dose response curves and can be tested by performing a specific antagonist assay (provided as an Appendix 2 to OECD TG 455). Other ER STTAs that do not use luciferase as a reporter gene may not have this drawback (Escande et al., 2006). Confirmation of the results (both positive and negative) could be obtained by using a cell system relying on a different reporter/read-out. For a review, see Thorne, Inglese and Auld (2010).

164. The ER STTA will not detect substances that act by other mechanisms (e.g. AR, TR and steroidogenesis interference). These chemicals will, however, be detected in AR-, TR- and steroidogenesis-specific assays and therefore results from a suite of *in vitro* tests should be considered together. The assay will not detect substances that act by affecting the hypothalamic/pituitary/gonadal (HPG) system as an *in vivo* intact axis is required for this.

When/why the assay may be used

165. Although the ER STTA may be used at any stage in the hazard assessment process, the most likely use scenario is during initial assessment of chemicals for their ability to interact with endocrine systems *in vitro* (i.e. E,A,T,S modalities). The ER STTA is frequently conducted following a positive result in the ER binding assay. Assays for interaction with other modalities (e.g. AR, ER and steroidogenesis), are likely to be conducted at the same time so that all results can be considered together. TR and other assays concerning mechanisms of thyroid disruption may be available, but they are not in common use. OECD TG 445 does not include the use of a xenobiotic metabolising system, but consideration should be given to the inclusion of this (Jacobs et al., 2008, 2013; OECD, 2008) depending on the circumstances, e.g. if the metabolism of a chemical is unknown, although the methods for inclusion of xenobiotic metabolising systems are not yet validated (see Paragraph 50). Alternatively, for a chemical with known metabolites, these could also be tested in the ER STTA.

166. Another use scenario may be following effects obtained in higher tier tests, for example accelerated puberty onset in females, but which are not exclusively indicative of an effect on ER. Selection of the most appropriate tests has to be on a case-by-case basis, but also considering the need to minimise animal testing. A further example could be results obtained in other apical assays, e.g. OECD TG 408 (90-day toxicity test); effects on reproductive organs may be investigated further by testing in the ER STTA in combination with AR- and steroidogenesis-based assays.

Introduction to the table of scenarios

167. Table C.1.2 gives guidance on a further step to take in the event of a positive (+) or negative (-) result in the ER STTA and in the presence of positive (+), negative (-) or equivocal/absent (Eq/0) existing results. “Existing results” are subdivided into “mechanism” and “effects” data (third and fourth columns). The table is divided horizontally into a series of scenarios that represent all the combinations of these events.

168. The results of the ER STTA are given in the second column. Criteria for positive and negative results in OECD TG 455 for both test methods are given in the test guideline. Reproducible results in at least two runs are required. If two runs do not give reproducible results (e.g. a test chemical is positive in one run and negative in the other run), or if a higher degree of certainty is required regarding the outcome of the assay, at least three independent runs should be conducted. In this case the classification is based on the two concordant results out of the three. It is important that quality and proficiency criteria are demonstrated for both positive and negative results. The concentrations tested should remain within the solubility range of the test chemicals and not demonstrate cytotoxicity.

169. Equivocal results for the guideline are not included in the table because these data generally require further interrogation about the result itself. This assay is a screen and therefore a clear positive or negative result should be obtained. In the event of an equivocal result, the considerations mentioned above about control quality and proficiency criteria should be taken into account and further investigations made.

Existing data to be considered

170. Existing "mechanism" *in vitro* data are assumed to be available from AR- and steroidogenesis-based assays (Level 2). Assays may also be available for interference with thyroid modalities. In practice, it is possible that data from all of these assays may not be available, so judgement will need to be used to decide which assays to perform.

171. Existing "effects" data refer to *in vivo* effects "of concern" (i.e. data from Level 4 or 5 vertebrate wildlife assays). These may come from varied sources and will depend on the type of substance (e.g. new chemicals, high production volume [HPV] chemical, pesticide). Thus, available data may range from repeated dose toxicity studies (28-day, 90-day), combined repeat dose/reproductive screening assays or fish screening assays, to chronic toxicity studies and multigeneration reproductive tests in vertebrate wildlife species. Some studies fail to identify EDs that weakly affect estrogen or androgen receptors as was demonstrated on the basis of data generated in the validation process of OECD TG 407 assay with endocrine endpoints. In this validation, only moderate EDs such as nonylphenol and DDE, and strong EDs such as ethinylestradiol and flutamide (acting via ER and AR respectively) were detected. Thus, OECD TG 407 cannot be regarded as a screening assay for endocrine activity. This means that when a relatively insensitive test is positive for both endocrine-specific and apical endpoints, this should be taken as an indication that the substance is a potential ED. Caution should be exercised, however, because endocrine endpoints may be impacted secondary to non-endocrine toxicity and *in vivo* apical endpoints can be affected by many modes of action, including endocrine modalities. Data may also be available on effects in mammalian and non-mammalian wildlife species, although caution should be used when extrapolating between taxa. A chemical causing endocrine effects in fish or amphibians (for example, OECD TG 240 or TG 241), may also have endocrine effects in mammals, but the physiological consequences of the effects are likely to be different.

172. Data may also be available from Level 3 tests (H and UT assays), but as the UT assay primarily detects (*in vivo*) the same modality as the ER STTA, it is unlikely that it would be conducted prior to this. An AMA may also be available, but as this test primarily detects thyroid disruption in amphibians it is unlikely to provide useful data for E-modalities.

173. When considering the results of the ER STTA, all available data should be used in order to reach a conclusion and a weight of evidence approach taken. This may include

HTS data, read-across data from structural analogues and QSAR. Several QSAR models for ER and AR binding/activation are now available (see Sections B.1.1.1 and B.1.1.2).

Scenarios: Positive and negative results combined with existing data

174. The scenarios (A to R) presented in Table C.1.2 represent all the possibilities of positive or negative results in combination with the presence or absence of existing data. The action taken will also depend on the regulatory environment, but the considerations given here are generally science based. Although OECD TG 455 uses hrER, the well-conserved nature of ER across taxa should be a strong indication that results in this assay are relevant to other vertebrate species. Wherever possible, the recommended "next step which could be taken" avoids unnecessary animal testing. However, sometimes conducting an animal test will be indicated and then the relevance of species, strain, exposure route and species-specific metabolism should always be considered. Further considerations specific to each scenario are given in the table.

175. Scenarios A to C represent positive results in the ER STTA assay in the presence of positive *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. A positive result in an ER STTA assay is strong evidence for (anti)estrogenic activity that may or may not be supported by the *in vivo* effects data. In the case of positive *in vivo* effects data, there may be sufficient evidence to conclude concern for endocrine disruption and therefore no need for further screening. *In vivo* assays/tests with negative results should be interpreted with caution as they may either indicate that the tests used do not have sufficient power to detect weak effects or, alternatively, that the effects do not present a concern for endocrine disruption. Generally, a conclusion of lack of concern for endocrine disruption in the presence of positive effects data may only be made given adequate Level 5 assays. Information on some endocrine-related tumours may be detected more comprehensively in carcinogenicity studies (OECD TG 451/453) (Level 4); for example, detection of certain types of thyroid tumors without causing reproductive or developmental effects, as well as substances causing tumors in other endocrine-sensitive tissues. Mode of action (MOA) data to provide a clear interpretation may be required by some regulatory agencies.

176. Scenarios D to F represent positive results in the ER STTA assay in the presence of negative *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. Unless the metabolic profile of the test substance is known, one option may be to conduct these *in vitro* assays with an added metabolising system. If the metabolic profile is known, then an *in vivo* test may be advisable. The choice of tests will depend on the available *in vivo* effects data. As in Scenarios A to C, *in vivo* assays/tests with negative results should be interpreted with caution as they may either indicate that the tests used do not have sufficient power to detect weak effects or, alternatively, that the effects do not present a concern for endocrine disruption. Generally, a conclusion of lack of concern for endocrine disruption in the presence of positive effects data (Scenario E) may only be made given adequate Level 5 assays. Information on some endocrine-related tumours may be detected more comprehensively in carcinogenicity studies (OECD TG 451/453) (Level 4); for example, detection of certain types of thyroid tumors in the absence of reproductive or developmental effects, as well as substances causing tumors in other endocrine-sensitive tissues. MOA data to provide a clear interpretation may be required by some regulatory agencies.

177. Scenarios G to I represent positive results in the ER STTA assay in the presence of various combinations of missing or equivocal data. The next step to take in these

eventualities will depend on the nature of the other available data and the jurisdiction in which it is being used. In some cases, equivocal data may be viewed as positive whilst in others it may or may not contribute to the weight of evidence. The interpretation may also depend on the MOA in question and why the data are considered equivocal, e.g. a study that is equivocal for thyroid effects may still be of value in evaluating (anti)estrogenic effects. In all three scenarios, the recommended first step is to obtain reliable mechanistic (*in vitro*) data rather than proceed directly to *in vivo* testing. Equivocal and missing data are alternative scenarios and two possibilities for the next step are given in most cases, but the nature of equivocal data means that decisions need to be taken on a case-by-case basis. In all cases, the role of metabolism, route of exposure and data from structural analogues should be considered before deciding on the next step. As above, generally a conclusion of lack of concern for endocrine disruption in the presence of some positive effects data (Scenario H) may only be made given adequate Level 5 assays. Information on some endocrine-related tumours may be detected more comprehensively in carcinogenicity studies (OECD TG 451/453) (Level 4); for example, detection of certain types of thyroid tumors in the absence of reproductive or developmental effects, as well as substances causing tumors in other endocrine-sensitive tissues. MOA data to provide a clear interpretation may be required by some regulatory agencies.

178. Scenarios J to L represent negative results in the ER STTA assay in the presence of positive *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. The limitations of the ER STTA assay should be considered first (e.g. lack of metabolic activation, possible involvement of other factors). The positive *in vitro* mechanistic data indicates possible alternative A,T,S mechanisms. To confirm lack of ER-related activity in the presence of *in vivo* data, an ER STTA with added metabolising capability could be performed. Otherwise, *in vivo* tests will confirm or refute E,A,T,S activity.

179. Scenarios M to O represent negative results in the ER STTA assay in the presence of negative *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. Negative results for all tests (Scenario N) may be sufficient to enable a conclusion of no concern for endocrine disruption. This will depend on the weight of evidence and may not be possible in some cases. However, in the presence of negative data from robust Level 4 and 5 assays, further animal testing is probably not justified. The limitations of the ER STTA assay should also be considered (as described for Scenarios J to L). To confirm lack of ER-related activity in the presence of *in vivo* data, an ER STTA with added metabolising capability could be performed. Otherwise, *in vivo* tests will confirm or refute E,A,T,S activity (Scenarios M and O).

180. Scenarios P to R represent negative results in the ER STTA assay in the presence of various combinations of missing or equivocal data. The limitations of the ER STTA binding assay should be considered first (as described for Scenarios J to L). As with the positive result scenarios above, the next step to take for Scenarios P to R will have to be decided on a case-by-case basis. However, the recommended first step is generally to obtain reliable mechanistic (*in vitro*) data rather than proceed directly to *in vivo* testing. In all cases, the role of metabolism, route of exposure and data from structural analogues should be considered before deciding on the next step.

181. In all scenarios (A to R), the next step to take to strengthen weight of evidence will depend on the existing information. Table C.1.2 is meant to provide a succinct guide and may not cover all circumstances or possibilities. The scenarios may also suggest that chemicals have simple or single MOA, when in practice they may have multiple endocrine and non-endocrine MOA. In some cases, for example, two opposite modes of simultaneous

action (e.g. estrogenic and anti-estrogenic) could, depending on dose, lead to a minimisation or abolition of effects, while in others two different MOA (e.g. estrogenic and anti-androgenic) could potentially reinforce effects. Endocrine pathways interact, mixed effects are common and there are many pathways that cannot be distinguished with currently available TGs. If multiple MOA are suspected, either from the existing results or based on QSAR/read-across/integrated approaches, this should be investigated further if needed for regulatory decision making.

182. In general, a decision about whether or not to conduct *in vivo* vertebrate wildlife tests will depend on the weight of evidence of new and existing data. If most available data (e.g. the results of the ER STTA assay, predictions from QSARs, "read-across" from data on similar substances and results from mammalian *in vivo* assays) suggest that the substance has the potential to cause endocrine disruption via the estrogen receptor (i.e. the level of suspicion about endocrine disrupting action is high – corresponding to Scenario A), then consideration should be given to the conduct of a higher level test.

183. For non-mammalian wildlife species, higher level tests with fish or amphibians (i.e. OECD TG 234, TG 240 or TG 241) are recommended. Choice about which of these tests is most appropriate will be driven *inter alia* by mode of action considerations, and by whether multigeneration effects are to be expected. If available data only raise a low or moderate level of suspicion about endocrine disrupting action (e.g. the data appear to conflict with each other), then consideration should be given to the conduct of a fish screen (i.e. EASZY, OECD TG 229 or TG 230).

184. Potency of any interaction with ER should also be considered in relation to cross-species effects. Ankley et al. (2016) showed that chemicals with moderate to high estrogenic potency in mammalian systems should be priority chemicals in non-mammalian vertebrates. However, applicablity to invertebrates was uncertain because of a lack of knowledge of the biological role(s) of possible ERα orthologs found in phyla such as annelids. For low-affinity chemicals, comparative analysis of *in vitro* data for low-affinity chemicals suggested that mammalian-based assays may not effectively capture ERα interactions for fish and reptiles.

185. For mammals, similar considerations apply but lower level tests (e.g. Level 3 or 4) should be conducted before higher level tests in order to avoid unnecessary animal usage, unless it is apparent that a Level 5 test will be required anyway or will be needed to establish the evidence to conclude on ED properties. At Level 5, the EOGRTS (OECD TG 443) is the most sensitive reproduction assay for detecting endocrine disruption because it includes evaluation of a number of endocrine endpoints not included in the two-generation study (OECD TG 416) adopted in 2001. It is recognised, however, that some jurisdictions may require a two-generation study.

References

Andersen, H.R. et al. (2002), "Effects of currently used pesticides in assays for estrogenicity, androgenicity, and aromatase activity *in vitro*", *Toxicology and Applied Pharmacology.*, Vol. 179/1, pp. 1-12, https://doi.org/10.1006/taap.2001.9347.

Ankley, G.T. et al. (2016), "Evaluation of the scientific underpinnings for identifying estrogenic chemicals in nonmammalian taxa using mammalian test systems", *Environmental Toxicology and Chemistry*, Vol. 35/11, pp. 2806-2816, https://doi.org/10.1002/etc.3456.

Du, G. et al. (2010), "Assessing hormone receptor activities of pyrethroid insecticides and their metabolites in reporter gene assays", *Toxicological Sciences*, Vol. 116/1, pp. 58-66, https://doi.org/10.1093/toxsci/kfq120.

Escande, A. et al. (2006), "Evaluation of ligand selectivity using reporter cell lines stably expressing estrogen receptor alpha or beta", *Biochemical Pharmacology*, Vol. 71/10, pp. 1459-1469, https://doi.org/10.1016/j.bcp.2006.02.002.

Jacobs, M. et al. (2013), "*In vitro* metabolism and bioavailability tests for endocrine active substances: What is needed next for regulatory purposes?", *ALTEX – Alternatives to Animal Experimentation*, Vol. 30/3, pp. 331-351.

Jacobs, M.N. et al. (2008), "The use of metabolising systems for *in vitro* testing of endocrine disrupters", *Current Drug Metabolism*, Vol. 9/8, pp. 796-826.

Kuiper, G.G. et al. (1998), "Interaction of estrogenic chemicals and phytoestrogens with estrogen receptor beta", *Endocrinology*, Vol. 139/10, pp. 4252-3263, https://doi.org/10.1210/endo.139.10.6216.

Odum, J. et al. (1997), "The rodent uterotrophic assay: Critical protocol features, studies with nonyl phenols, and comparison with a yeast estrogenicity assay", *Regulatory Toxicology and Pharmacology*, Vol. 25/2, pp. 176-188, https://doi.org/10.1006/rtph.1997.1100.

OECD (2012a), *Detailed Review Paper on the State of the Science on Novel* In Vitro *and* In Vivo *Screening and Testing Methods and Endpoints for Evaluating Endocrine Disrupters*, OECD Series on Testing and Assessment, No. 178, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264221352-en.

OECD (2012b), "Performance standards for stably transfected transactivation *in vitro* assays to detect estrogen agonists for TG 455", OECD Series on Testing and Assessment, No. 173, OECD, Paris, www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2012)18&doclanguage=en.

OECD (2012c), "Performance standards for stably transfected transactivation *in vitro* assays to detect estrogen receptor antagonists", OECD Series on Testing and Assessment, No. 174, OECD, Paris, www.oecd.org/env/ehs/testing/19_REV1.pdf.

OECD (2008), *Detailed Review Paper on the Use of Metabolising Systems for* In Vitro *Testing of Endocrine Disrupters*, OECD Series on Testing and Assessment, No. 97, OECD Publishing, Paris, https://doi.org/10.1787/9789264085497-en.

Routledge, E.J. and J.P. Sumpter (1996), "Estrogenic activity of surfactants and some of their degradation products assessed using a recombinant yeast screen", *Environmental Toxicology and Chemistry*, Vol. 15/3, pp. 241-248, https://doi.org/10.1002/etc.5620150303.

Sheahan, D.A. et al. (2002), "Reduction in the estrogenic activity of a treated sewage effluent discharge to an English river as a result of a decrease in the concentration of industrially derived surfactants", *Environmental Toxicology and Chemistry*, Vol. 21, pp. 515-419, https://doi.org/10.1002/etc.5620210307.

Takeyoshi, M. et al. (2002), "The efficacy of endocrine disrupter screening tests in detecting anti-estrogenic effects downstream of receptor-ligand interactions", *Toxicology Letters*, Vol. 126/2, pp. 91-98, https://doi.org/10.1016/S0378-4274(01)00446-5.

Thorne, N., J. Inglese and D.S. Auld (2010), "Illuminating insights into firefly luciferase and other bioluminescent reporters used in chemical biology", *Chemical Biology*, Vol. 17/6, pp. 646-657, http://dx.doi.org/10.1016/j.chembiol.2010.05.012.

Witters, H. et al. (2010), "The assessment of estrogenic or anti-estrogenic activity of chemicals by the human stably transfected estrogen sensitive MELN cell line: Results of test performance and transferability", *Reproductive Toxicology*, Vol. 30/1, pp. 60-72, https://doi.org/10.1016/j.reprotox.2010.02.008.

Table C.1.2. **Performance-Based Test Guideline for Stably Transfected Transactivation *In Vitro* Assays to Detect Estrogen Receptor Agonists and Antagonists (ER STTA) (OECD TG 455): Guidance for scenarios of combinations of results with existing data**

This table represents possible conclusions to be drawn from assay data, and a next step which could be taken if further evidence is required about possible endocrine disrupting properties and/or effects. The guidance offered is not meant to be prescriptive, but provides science-based considerations. It encourages the use of all available data and expert judgement in a weight of evidence approach. Regional and national interpretation of results and "next steps" may vary.

The conclusions are grouped into a series of scenarios (A-R), each scenario representing a different combination of assay results, existing *in vitro* data and existing *in vivo* data. The symbol "+" indicates that the data in question represent a positive result, "–" indicates a negative result, and "Eq/0" indicates that the data are either equivocal or are not available.

Existing results: * "Mechanism (*in vitro* mechanistic data)" assumes that mechanistic data are available from estrogen receptor (ER-), androgen receptor (AR-) and steroidogenesis (S-) based assays (Level 2). The ER binding assay is likely to be performed prior to the Stably Transfected Human Estrogen Receptor-alpha Transactivation Assay for Detection of Estrogenic Agonist-Activity of Chemicals (ER STTA). Thyroid hormone receptor (TR) and other assays concerning mechanisms of thyroid disruption may be available, but they are not in common use. In practice, data from all assays may not be available and therefore this must be taken into account when deciding on the "next step". Quantitative structure activity relationship (QSAR) predictions of estrogen and androgen binding/activation may be made for some substances.

Existing results: ** "Effects (*in vivo* effects of concern)" assumes various information, such as data from repeat dose oral toxicity studies, reproduction/developmental toxicity screen tests, read-across from analogues, will be available.

Scenarios	Result of OECD TG 455 (ER STTA)	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
A	+	+	+	Estrogen receptor (ER) (ant)agonism combined with effects on AR/T/S and potential for adverse effects via multiple mechanisms.	Perform assay from Levels 3-5, e.g. Uterotrophic Bioassay (UT) assay (Level 3) or female Peripubertal (PP) Assay (Level 4) or Extended One-Generation Reproductive Toxicity Study (EOGRTS) or two-generation assays (Level 5) or partial/full non-mammalian wildlife life cycle tests, e.g. OECD TG 241 and TG 240 (Level 4/5).	A positive result indicates strong probability of interaction with ERs in other taxa. If existing data are from a Level 5 assay, there may be sufficient information to conclude evidence of concern for endocrine disruption (the EOGRTS provides the most information; however, for endocrine disrupting chemicals (EDCs)s with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from Level 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism, although some Level 4 assays (e.g. TG 234 [FSDT]) may be sufficient for this purpose. If existing data are from UT assays, then Level 4 assays will provide data on multiple modalities. Consider route of exposures for existing effects data and possible transformation products and implications of absorption, distribution, metabolism and excretion (ADME) characteristics of the chemical. The chemical itself may give positive *in vitro* results and *in vivo* results, but may also be metabolised to a metabolite that also has positive results *in vitro* and *in vivo*. However, note that uptake and metabolism of chemicals can be different between mammalian and non-mammalian wildlife species.
B	+	+	–	ER (ant)agonism combined with effects on AR/T/S but effects not detected in *in vivo* studies. Weak ER (ant)agonism does not result in adverse effects in the selected species under the conditions of the test. Metabolic differences may explain *in vitro*/*in vivo* differences.	Perform Stably Transfected Human Estrogen Receptor-alpha Transactivation Assay for Detection of Estrogenic Agonist-Activity of Chemicals (ER STTA) with added metabolising system or Assay from Levels 3-4, e.g. UT assay or fish screen (OECD TG 229/230) (Level 3) or female PP assay (Level 4).	A positive result indicates strong probability of interaction with ERs in other taxa. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from Level 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism, although some Level 4 assays (e.g. TG 234 [FSDT]) may be sufficient for this purpose. If existing data are from UT assays, then Level 4 assays will provide data on multiple modalities. Consider route of exposures for existing effects data and possible transformation products and implications of ADME characteristics of the chemical. The chemical itself may give positive *in vitro* results but may not be absorbed or may be metabolised to an inactive metabolite *in vivo*. However, note that uptake and metabolism of chemicals can be different between mammalian and non-mammalian wildlife species.

Scenarios	Result of OECD TG 455 (ER STTA)	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
C	+	+	Eq/0	ER (ant)agonism combined with effects on AR/T/S but no or equivocal data from *in vivo* studies. Weak ER (ant)agonism may not result in adverse effects in the selected species under the conditions of the test.	Perform assay from Levels 3-4, e.g. UT assay or fish screen (OECD TG 229/230) (Level 3) or female PP assay (Level 4).	A positive result indicates strong probability of interaction with ERs in other taxa. Consider route of exposures for equivocal existing effects data and possible implications of ADME characteristics of the chemical. Equivocal results may indicate chemical has multiple modes of action (MOA). Check data on chemical analogues.
D	+	–	+	ER (ant)agonism and potential for adverse effects.	Perform assay from Levels 3-4, e.g. UT assay or fish screen (OECD TG 229/230) (Level 3) or female PP assay (Level 4).	A positive result indicates strong probability of interaction with ERs in other taxa. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude evidence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from Level 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism, although some Level 4 assays (e.g. TG 234 [FSDT]) may be sufficient for this purpose. If existing data are from UT assays, then Level 4 assays will provide data on multiple modalities. A positive result could have arisen from other (E,A,T,S or non-E,A,T,S) mechanisms, e.g. hypothalamic/pituitary/gonadal (HPG) axis.
E	+	–	–	ER (ant)agonism but effects not detected in *in vivo* studies. Weak ER (ant)agonism does not result in adverse effects in the selected species under the conditions of the test. Metabolic differences may explain *in vitro*/*in vivo* differences.	Perform ER STTA with added metabolising system or Assay from Levels 3-4, e.g. UT assay or fish screen (OECD TG 229/230) (Level 3) or female PP assay (Level 4).	A positive result indicates strong probability of interaction with ERs in other taxa. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from Level 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism, although some Level 4 assays (e.g. TG 234 [FSDT]) may be sufficient for this purpose. If existing data are from UT assays, then Level 4 assays will provide data on multiple modalities. Consider route of exposures for existing effects data and possible transformation products and implications of ADME characteristics of the chemical. The chemical itself may give positive *in vitro* results but may not be absorbed or may be metabolised to an inactive metabolite *in vivo*. However, note that uptake and metabolism of chemicals can be different between mammalian and non-mammalian wildlife species. Check data on chemical analogues.

Scenarios	Result of OECD TG 455 (ER STTA)	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
F	+	–	Eq/0	ER (ant)agonism but no or equivocal data from *in vivo* studies. Weak ER (ant)agonism does not result in adverse effects in the selected species under the conditions of the test.	Perform assay from Levels 3-4, e.g. UT assay or fish screen (OECD TG 229/230) (Level 3) or female PP assay (Level 4).	A positive result indicates strong probability of interaction with ERs in other taxa. Consider route of exposures for equivocal existing effects data and possible implications of ADME characteristics of the chemical. Check data on chemical analogues. Equivocal results may indicate chemical has multiple MOA.
G	+	Eq/0	+	ER (ant)agonism and potential for adverse effects via ER (ant)agonism or other A,T,S mechanisms. May act via E,A,T,S mechanism and may or may not require metabolic activation.	Perform assay from Levels 3-4, e.g. UT assay or fish screen (OECD TG 229/230) (Level 3) or female PP assay (Level 4).	A positive result indicates strong probability of interaction with ERs in other taxa. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude evidence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from Level 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism, although some Level 4 assays (e.g. TG 234 [FSDT]) may be sufficient for this purpose. If existing data are from UT assays, then Level 4 assays will provide data on multiple modalities. Check data on chemical analogues. Further mechanistic studies may help determine MOA. Equivocal results may indicate chemical has multiple MOA. A positive result could have arisen from other (E,A,T,S or non-E,A,T,S) mechanisms (e.g. HPG axis).
H	+	Eq/0	–	ER (ant)agonism but effects not detected in *in vivo* studies. Weak ER (ant)agonism does not result in adverse effects in the selected species under the conditions of the test. Metabolic differences may explain *in vitro*/*in vivo* differences.	Perform ER STTA with added metabolising system.	A positive result indicates strong probability of interaction with ERs in other taxa. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from Level 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism, although some Level 4 assays (e.g. TG 234 [FSDT]) may be sufficient for this purpose. If existing data are from UT assays, then Level 4 assays will provide data on multiple modalities. Consider route of exposures for existing effects data and possible transformation products and implications of ADME characteristics of the chemical. The chemical itself may give positive *in vitro* results but may not be absorbed or may be metabolised to an inactive metabolite *in vivo*. However, note that uptake and metabolism of chemicals can be different between mammalian and non-mammalian wildlife species. Check data on chemical analogues. Further mechanistic studies may help determine MOA. Equivocal results may indicate chemical has multiple MOA.

Scenarios	Result of OECD TG 455 (ER STTA)	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
I	+	Eq/0	Eq/0	ER (ant)agonism with unknown potential for effects in *in vivo* studies. May act via ER mechanism and may or may not require metabolic activation. Unknown potential for adverse effects.	Perform ER STTA with added metabolising system or UT assay or fish screen (OECD TG 229/230) (Level 3), if existing data indicate this is needed.	A positive result indicates strong probability of interaction with ERs in other taxa. Consider route of exposures for equivocal existing effects data and possible implications of ADME characteristics of the chemical. Check data on chemical analogues. Equivocal results may indicate chemical has multiple MOA.
J	–	+	+	No evidence for ER (ant)agonism. Effects on AR/T/S and potential for adverse effects via E,A,T,S mechanisms.	Perform ER STTA with added metabolising system or Perform assay from Levels 3-4, e.g. UT assay or fish screen (OECD TG 229/230) (Level 3) or female PP assay (Level 4).	A negative result indicates interaction with ERs in other taxa is unlikely. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude concern for endocrine disruption. If existing data are from Level 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism, although some Level 4 assays (e.g. TG 234 [FSDT]) may be sufficient for this purpose. If existing data are from UT assays, then Level 4 assays will provide data on multiple modalities. Consider route of exposures for existing effects data and possible transformation products and implications of ADME characteristics of the chemical. Check data on chemical analogues. Further mechanistic studies may help determine MOA.
K	–	+	–	No evidence for ER (ant)agonism. Effects on AR/T/S but effects not detected in *in vivo* studies.	Perform assay from Levels 3-4, e.g. fish screen (OECD TG 229/230) (Level 3) or male or female PP assay (Level 4).	A negative result indicates interaction with ERs in other taxa is unlikely. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from Level 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism, although some Level 4 assays (e.g. TG 234 [FSDT]) may be sufficient for this purpose. If existing data are from UT assays, then Level 4 assays will provide data on multiple modalities. Metabolic deactivation of chemical may occur *in vivo* so that potential *in vitro* E,A,T,S activity is not realised. Consider possible routes of exposure, implications of metabolism.
L	–	+	Eq/0	No evidence for ER (ant)agonism. Effects on AR/T/S but effects not detected in *in vivo* studies. Unknown potential for adverse effects. Metabolic differences explain *in vitro*/*in vivo* A,T,S differences.	Perform assay from Levels 3-4, e.g. fish screen (OECD TG 229/230) (Level 3) or male or female PP assay (Level 4).	A negative result indicates interaction with ERs in other taxa is unlikely. Metabolic deactivation of chemical may occur *in vivo* so that potential *in vitro* E,A,T,S activity is not realised. Consider possible routes of exposure, implications of metabolism. Check data on chemical analogues. Equivocal results may indicate chemical has multiple MOA.

Scenarios	Result of OECD TG 455 (ER STTA)	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
M	–	–	+	No evidence for ER (ant)agonism. Metabolic differences or route of exposure explain *in vitro*/*in vivo* differences. Effects seen in existing studies are via non-E,A,T,S or non-endocrine mechanisms.	Perform ER STTA with added metabolising system or Perform assay from Levels 3-4, e.g. fish screen (OECD TG 229/230) (Level 3) or male or female PP assay (Level 4).	A negative result indicates interaction with ERs in other taxa is unlikely. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude concern for endocrine disruption. If existing data are from Level 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism, although some Level 4 assays (e.g. TG 234 [FSDT]) may be sufficient for this purpose. If existing data are from UT assays, then Level 4 assays will provide data on multiple modalities. Metabolic activation of chemical may occur *in vivo*. Consider possible routes of exposure, implications of metabolism. Check data on chemical analogues. Further mechanistic studies may help determine MOA.
N	–	–	–	No evidence for ER (ant)agonism. No evidence of adverse effects.	Possibly no need for further testing. If there is uncertainty, may perform assay from Levels 3-4, e.g. fish screen (OECD TG 229/230) (Level 3) or male or female PP assay (Level 4).	A negative result indicates interaction with ERs in other taxa is unlikely. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from Level 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism, although some Level 4 assays (e.g. TG 234 [FSDT]) may be sufficient for this purpose. If existing data are from UT assays, then Level 4 assays will provide data on multiple modalities. Check data on chemical analogues.
O	–	–	Eq/0	No evidence for ER (ant)agonism. Unknown potential for adverse effects via other mechanisms.	Perform ER STTA with added metabolising system or Fish screen (OECD TG 229/230) (Level 3) or male or female PP assay (Level 4) if existing data indicate this is needed.	A negative result indicates interaction with ERs in other taxa is unlikely. Consider possible routes of exposure, implications of metabolism. Check data on chemical analogues.

Scenarios	Result of OECD TG 455 (ER STTA)	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
P	–	Eq/0	+	No evidence for ER (ant)agonism. Unknown potential for adverse effects via other mechanisms.	Perform ER STTA with added metabolising system.	A negative result indicates interaction with ERs in other taxa is unlikely. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude concern for endocrine disruption. If existing data are from Level 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism, although some Level 4 assays (e.g. TG 234 [FSDT]) may be sufficient for this purpose. If existing data are from UT assays, then Level 4 assays will provide data on multiple modalities. Consider possible routes of exposure, implications of metabolism. Check data on chemical analogues. Further mechanistic studies may help determine MOA. Equivocal results may indicate chemical has multiple MOA.
Q	–	Eq/0	–	No evidence for ER (ant)agonism. No evidence of adverse effects.	Perform ER STTA with added metabolising system.	A negative result indicates interaction with ERs in other taxa is unlikely. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from Level 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism, although some Level 4 assays (e.g. TG 234 [FSDT]) may be sufficient for this purpose. If existing data are from UT assays, then Level 4 assays will provide data on multiple modalities. Check data on chemical analogues. Further mechanistic studies may help determine MOA.
R	–	Eq/0	Eq/0	No evidence for ER (ant)agonism. Unknown potential for adverse effects via other mechanisms.	For the "0" scenario, perform ER STTA with added metabolising system or Perform UT assay or fish screen (OECD TG 229/230) (Level 3), if existing data indicate this is needed.	A negative result indicates interaction with ERs in other taxa is unlikely. Consider possible routes of exposure, implications of metabolism. Check data on chemical analogues. Further mechanistic studies may help determine MOA. Equivocal results may indicate chemical has multiple MOA.

C.1.3. Stably Transfected Human Androgen Receptor Transcriptional Activation Assay for Detection of Androgenic Agonist and Antagonist Activity of Chemicals (AR STTA) (OECD TG 458)

Status: Assay validated by the OECD.

Modality detected/endpoints: Activation of reporter gene linked to AR (agonist assay). Inhibition of activation of reporter gene linked to AR.

Background to the assay

186. The Stably Transfected AR Transcriptional Activation Assay (AR STTA) is an *in vitro* screening assay to detect substances that bind to androgen receptors (AR) and activate the transcription of androgen responsive genes. It is an *in vitro* tool that provides mechanistic data. Several AR STTA assays in common use can be found in the literature (Hartig et al., 2002; Birkhøj et al., 2004; Araki et al., 2005a, 2005b), one of the first versions of this assay used was the "yeast androgen screen" (Sohoni and Sumpter, 1998), which is still widely used for screening of environmental samples. The guidance in this building block can be cautiously used for the yeast assays. OECD TG 458 was published in July 2016, following validation of this assay using the AR-EcoScreen™ cell line. Other AR STTA assays are being validated via OECD initiatives and a performance-based test guideline for this assay will be developed in the future.

187. The AR-EcoScreen™ cell line is derived from a Chinese hamster ovary cell line (CHO-K1) stably transfected with human (h) AR and uses a firefly luciferase reporter gene resulting in increased cellular expression of the luciferase enzyme in the presence of AR agonists. This cell line has also been constructed to stably express a non-inducible renilla luciferase reporter gene, the activity of which decreases in the presence of cytotoxic agents. The luciferase enzymes differ in their substrate and cofactor requirements and emit light at different wavelengths. Cell viability can therefore be determined in the same cells as those used for AR (ant)agonism. This enables pure antagonisms to be distinguished from a cytotoxicity-related decrease of luciferase activity. Other AR STTA assays utilise different viability assessments.

188. The AR STTA assay provides a positive or negative result for the ability of a chemical to induce AR-mediated transactivation of gene expression (agonist assay) compared to a vehicle control. The antagonist assay determines whether a reduction in response occurs when cells are co-exposed to chemicals and a potent androgen agonist compared to the potent androgen agonist alone. 5α- Dihydrotestosterone (DHT) is used as the co-administered agonist in OECD TG 458. R1881 is also commonly used. Any reduction in response must occur in the absence of cytotoxicity.

189. OECD TG 458 gives a positive or negative result for a test chemical when reporter gene activity is compared to controls. A measure of potency is also provided by the magnitude of the effect and the concentration at which it occurs. An AR agonistic effect is

based on the maximum response level induced by a test chemical. If this response equals or exceeds 10% of the response induced by DHT (the positive AR agonist control) (i.e. the log PC10), the test chemical is considered positive. An AR antagonistic effect is based on a cut-off of a 30% inhibitory response against DHT (i.e. the log IC30). If the response exceeds this 30% AR inhibition, then the chemical is considered a positive AR antagonist.

190. OECD TG 458 requires strict control of assay conditions in order to maintain the accuracy and reliability of response. Demonstration of laboratory proficiency with proficiency chemicals is required at the outset, ten for each of the agonist and antagonist assays. These chemicals were used in the validation of this assay (OECD, 2011), represent the classes of chemicals commonly associated with AR agonist or antagonist activity, exhibit a suitable range of potency expected for AR agonists/antagonists (i.e. strong to weak), and include negatives. Periodic testing with proficiency chemicals should also be carried out. In addition, each experiment requires reference chemicals: for the agonist assay, DHT (a strong agonist), mestanolone (a weak agonist) and (2-ethylhexyl)phthalate (DEHP) (negative); for the antagonist assay: hydroxyl flutamide (a strong antagonist), bisphenol A (a weak antagonist) and DEHP (negative) should be used. In the assay, each plate requires positive and vehicle controls. A positive control for cytotoxicity (cycloheximide) is also required for each plate. Criteria for the degree of response with these chemicals are given in the test guidance (TG). The assay requires a minimum of 80% cell viability, demonstrated by renilla luciferase activity. This is critical for the antagonist assay where positive results can only be demonstrated in the absence of cytotoxicity. Compliance with the quality control criteria and with the performance criteria should be accepted before evaluating results from this assay. The response with positive control chemicals (e.g. hydroxy-flutamide for antagonism and dihydrotestosterone for agonism) should be robust and cell viability should be above 80%.

191. Some cell lines used for the AR STTA also express the glucocorticoid receptor (GR), which may cause cross-talk interference with AR (Hartig et al., 2002). This is due to the fact that the receptor can act on the same responsive elements (androgen response elements). The level of GR expression in the cell line and therefore potential for interference should be known.

192. The AR STTA assay will not detect substances that act by other mechanisms (e.g. estrogen receptor [ER], thyroid hormone receptor [TR] and steroidogenesis interference). These chemicals will, however, be detected in ER-, TR- and steroidogenesis-specific assays and therefore results from a suite of *in vitro* tests should be considered together. The assay will not detect substances that act by affecting the hypothalamic/pituitary/ gonadal (HPG) as an *in vivo* intact axis is required for this.

When/why the assay may be used

193. Although the AR STTA assay may be used at any stage in the hazard assessment process, the most likely use scenario is during initial assessment of chemicals for their ability to interact with endocrine systems *in vitro*, i.e. estrogen/androgen/thyroid/steroidogenesis (E,A,T,S) modalities. The AR STTA assay is frequently conducted following a positive result in the AR Binding Assay. Assays for interaction with other modalities (e.g. AR, ER and steroidogenesis) are likely to be conducted at the same time so that all results can be considered together. TR and other assays concerning mechanisms of thyroid disruption may be available, but they are not in common use. AR STTAs do not include the use of a xenobiotic metabolising system, but consideration should be given to the inclusion of this (OECD, 2008; Jacobs et al., 2008, 2013) depending on the circumstances, e.g. if the

metabolism of a chemical is unknown, although the methods for inclusion of xenobiotic metabolising systems are not yet validated (see Paragraph 50). Alternatively, for a chemical with known metabolites, these could also be tested in the AR STTA assay.

194. Another use scenario may be following effects obtained in higher tier tests, for example accelerated puberty onset in males, but which are not exclusively indicative of an effect on AR. Selection of the most appropriate tests has to be on a case-by-case basis, but also considering the need to minimise animal testing.

Introduction to the table of scenarios

195. Table C.1.3 gives guidance on a further step to take in the event of a positive (+) or negative (-) result in the AR STTA assay and in the presence of positive (+), negative (-) or equivocal/absent (Eq/0) existing results. "Existing results" are subdivided into "mechanism" and "effects" data (third and fourth columns). The table is divided horizontally into a series of scenarios that represent all the combinations of these events.

196. The results of the AR STTA assay are given in the second column. Criteria for positive and negative results in OECD TG 458 are given in the test guideline. Reproducible results in at least two runs are required. If two runs do not give reproducible results (e.g. a test chemical is positive in one run and negative in the other run), at least three independent runs should be conducted. In this case, the classification is based on the two concordant results out of the three. It is important that quality and proficiency criteria are demonstrated for both positive and negative results. The concentrations tested should remain within the solubility range of the test chemicals and not demonstrate cytotoxicity.

197. Equivocal results for the AR STTA assay are not included in the table because these data require further interrogation about the result itself. This assay is a screen and therefore a clear positive or negative result should be obtained. In the event of an equivocal result, the considerations mentioned above about control quality and proficiency criteria should be taken into account and further investigations made. Equivocal results at high concentrations may may indicate solubility issues.

Existing data to be considered

198. Existing "mechanism" *in vitro* data are assumed to be available from AR-, ER- and steroidogenesis-based assays (Level 2). Assays may also be available for interference with thyroid modalities. In practice, it is possible that data from all of these assays may not be available, so judgement will need to be used to decide which assays to perform.

199. Existing "effects" data refer to *in vivo* effects "of concern" (i.e. data from Level 3, 4 or 5 vertebrate wildlife assays/tests). These may come from varied sources and will depend on the type of substance (e.g. new chemicals, high production volume [HPV] chemical, pesticide). Thus, available data may range from repeated dose toxicity studies (28-day, 90-day), or combined repeat dose/reproductive screening assays or fish screening assays, to chronic toxicity studies and multigeneration reproductive tests in vertebrate wildlife species. Some studies fail to identify endocrine disrupters (EDs) that weakly affect estrogen or androgen receptors as was demonstrated on the basis of data generated in the validation process of the OECD TG 407 assay with endocrine endpoints. In this validation only moderate EDs such as nonylphenol and DDE, and strong EDs such as ethinylestradiol and flutamide (acting via ER and AR respectively), were detected. Thus OECD TG 407 cannot be regarded as a screening assay for endocrine activity. This means that when a

relatively insensitive test is positive for both endocrine-specific and apical endpoints, this should be taken as an indication that the substance is a potential ED. Caution should be exercised, however, because endocrine endpoints may be impacted secondary to non-endocrine toxicity and *in vivo* apical endpoints can be affected by many modes of action (MOA), including endocrine modalities. Data may also be available on effects in mammalian and non-mammalian wildlife species, although caution should be used when extrapolating between taxa. A chemical causing endocrine effects in non-mammalian vertebrates species may also have endocrine effects in mammals, but the physiological consequences of the effects are likely to be different.

200. Data may also be available from Level 3 tests (Hershberger [H] and Uterotrophic [UT] Assays) although these tests may not give rise to "concern" as they are hazard screening tests only. The H assay is, however, more likely to be conducted **after** the AR STTA assay (to test whether a chemical that is positive *in vitro* is also positive *in vivo*) rather than before. An Amphibian Metamorphosis Assay may also be available, but as this test primarily detects thyroid disruption in amphibians it is unlikely to provide useful data for A-modalities.

201. When considering the results of the AR STTA assay, all available data should be used in order to reach a conclusion and a weight of evidence approach taken. This may include read-across data from structural analogues and quantitative structure activity relationships (QSARs). Several QSAR models for ER and AR binding/activation are now available (see Sections B.1.1.1 and B.1.1.2).

Scenarios: Positive and negative results combined with existing data

202. The scenarios (A to R) presented in Table C.1.3 represent all the possibilities of positive or negative results in combination with the presence or absence of existing data. The action taken will also depend on the regulatory environment, but the considerations given here are generally science based. Although OECD TG 458 uses hAR (human androgen receptor), the well-conserved nature of AR across taxa is assumed be a strong indication that results in this assay are relevant to other vertebrate species. Wherever possible, the recommended "next step which could be taken" avoids unnecessary animal testing. However, sometimes conducting an animal test will be indicated and then the relevance of species, strain, exposure route and species-specific metabolism should always be considered. Further considerations specific to each scenario are given in the table.

203. Scenarios A to C represent positive results in the AR STTA assay in the presence of positive *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. A positive result in an AR STTA assay is strong evidence for (anti)androgenic activity that may or may not be supported by the *in vivo* effects data. In the case of positive *in vivo* effects data, there may be sufficient evidence to conclude concern for endocrine disruption and therefore no need for further screening. *In vivo* assays/tests with negative results should be interpreted with caution as they may either indicate that the tests used do not have sufficient power to detect weak effects or, alternatively, that the effects do not present a concern for endocrine disruption. Generally, a conclusion of lack of concern for endocrine disruption in the presence of positive effects data may only be made given adequate Level 5 assays. Information on some endocrine-related tumours may be detected more comprehensively in carcinogenicity studies (OECD TG 451/453) (Level 4); for example, detection of certain types of thyroid tumors in the absence of reproductive or developmental effects, as well as substances causing tumors in other endocrine-sensitive tissues. MOA data to provide a clear interpretation may be required by some regulatory agencies.

204. Scenarios D to F represent positive results in the AR STTA assay in the presence of negative *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. Unless the metabolic profile of the test substance is known, one option may be to conduct these *in vitro* assays with an added metabolising system. If the metabolic profile is known, then an *in vivo* test may be advisable. The choice of tests will depend on the available *in vivo* effects data. As in Scenarios A to C, *in vivo* assays/tests with negative results should be interpreted with caution as they may either indicate that the tests used do not have sufficient power to detect weak effects or, alternatively, that the effects do not present a concern for endocrine disruption. Generally, a conclusion of lack of concern for endocrine disruption in the presence of positive effects data (Scenario E) may only be made given adequate Level 5 assays. Information on some endocrine-related tumours may be detected more comprehensively in carcinogenicity studies (OECD TG 451/453) (Level 4); for example, detection of certain types of thyroid tumors in the absence of reproductive or developmental effects, as well as substances causing tumors in other endocrine-sensitive tissues. MOA data to provide a clear interpretation may be required by some regulatory agencies.

205. Scenarios G to I represent positive results in the AR STTA assay in the presence of various combinations of missing or equivocal data. The next step to take in these eventualities will depend on the nature of the other available data and the jurisdiction in which it is being used. In some cases, equivocal data may be viewed as positive whilst in others it may or may not contribute to the weight of evidence. The interpretation may also depend on the MOA in question and why the data are considered equivocal, e.g. a study that is equivocal for thyroid effects may still be of value in evaluating (anti)androgenic effects. In all three scenarios, the recommended first step is to obtain reliable mechanistic (*in vitro*) data rather than proceed directly to *in vivo* testing. Equivocal and missing data are alternative scenarios and two possibilities for the next step are given in most cases, but the nature of equivocal data means that decisions need to be taken on a case-by-case basis. In all cases, the role of metabolism, route of exposure and data from structural analogues should be considered before deciding on the next step. As above, generally a conclusion of lack of concern for endocrine disruption in the presence of some positive effects data (Scenario H) may only be made given adequate Level 5 assays. Information on some endocrine-related tumours may be detected more comprehensively in carcinogenicity studies (OECD TG 451/453) (Level 4); for example, detection of certain types of thyroid tumors in the absence of reproductive or developmental effects, as well as substances causing tumors in other endocrine-sensitive tissues. MOA data to provide a clear interpretation may be required by some regulatory agencies.

206. Scenarios J to L represent negative results in the AR STTA assay in the presence of positive *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. The limitations of the AR STTA assay should be considered first (e.g. lack of metabolic activation, possible involvement of other factors). The positive *in vitro* mechanistic data indicate possible alternative E,T,S mechanisms. To confirm lack of AR-related activity in the presence of *in vivo* data, an AR STTA with added metabolising capability could be performed. Otherwise *in vivo* tests will confirm or refute E,A,T,S activity.

207. Scenarios M to O represent negative results in the AR STTA assay in the presence of negative *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. Negative results for all tests (Scenario N) may be sufficient to enable a conclusion of no concern for endocrine disruption. This will depend on the weight of evidence and may not be possible in some cases. However, in the presence of negative data from robust Level 4 and 5 assays, further animal testing is probably not justified. The limitations of the AR

STTA assay should also be considered (as described for Scenarios J to L). To confirm lack of AR-related activity in the presence of *in vivo* data, an AR STTA with added metabolising capability could be performed. Otherwise, *in vivo* tests will confirm or refute E,A,T,S activity (Scenarios M and O).

208. Scenarios P to R represent negative results in the AR STTA assay in the presence of various combinations of missing or equivocal data. The limitations of the AR STTA binding assay should be considered first (as described for Scenarios J to L). As with the positive result scenarios above (see Paragraph 203), the next step to take for Scenarios P to R will have to be decided on a case-by-case basis. However, the recommended first step is generally to obtain reliable mechanistic (*in vitro*) data rather than proceed directly to *in vivo* testing. In all cases, the role of metabolism, route of exposure and data from structural analogues should be considered before deciding on the next step.

209. In all scenarios (A to R), the next step to take to strengthen weight of evidence will depend on the existing information. Table C.1.3 is meant to provide a succinct guide and may not cover all circumstances or possibilities. The scenarios may also suggest that chemicals have simple or single MOA, when in practice they may have multiple endocrine and non-endocrine MOA. In some cases, for example, two opposite modes of simultaneous action (e.g. estrogenic and anti-estrogenic) could, depending on dose, lead to a minimisation or abolition of effects, while in others two different MOA (e.g. estrogenic and anti-androgenic) could potentially reinforce effects. Endocrine pathways interact, mixed effects are common and there are many pathways that cannot be distinguished with currently available TGs. If multiple MOA are suspected, either from the existing results or based on QSAR/read-across/integrated approaches, this should be investigated further if needed for regulatory decision making.

210. In general, a decision about whether or not to conduct *in vivo* vertebrate wildlife tests will depend on the weight of evidence of new and existing data. If most available data (e.g. the results of the AR STTA assay, predictions from QSARs, "read-across" from data on similar substances and results from mammalian *in vivo* assays) suggest that the substance has the potential to cause endocrine disruption via the AR (i.e. the level of suspicion about endocrine disrupting action is high – corresponding to Scenario A), then consideration should be given to the conduct of a higher level test.

211. For non-mammalian wildlife species, higher level tests with fish or amphibians (i.e. the TG 234 [FSDT], TG 240, TG 241) are recommended. Choice about which of these tests is most appropriate will be driven *inter alia* by MOA considerations, and by whether multigeneration effects are to be expected. Such tests are unlikely to be needed if exposure of the natural environment is not expected. On the other hand, if available data only raise a low or moderate level of suspicion about endocrine disrupting action (e.g. the data appear to conflict with each other), then consideration should be given to the conduct of a fish screen (i.e. JMASA, OECD TG 229 or TG 230).

212. For mammals, similar considerations apply but lower level tests (e.g. Level 3 or 4) should be conducted before higher level tests in order to avoid unnecessary animal usage, unless it is apparent that a Level 5 test will be required anyway or will be needed to establish the evidence to conclude on ED properties. At Level 5, the EOGRTS (OECD TG 443) is the most sensitive reproduction assay for detecting endocrine disruption because it includes evaluation of a number of endocrine endpoints not included in the two-generation study (OECD TG 416) adopted in 2001. It is recognised, however, that some jurisdictions may require a two-generation study.

References

Araki, N. et al. (2005a), "Screening for androgen receptor activities in 253 industrial chemicals by *in vitro* reporter gene assays using AR-EcoScreen cells", *Toxicology In Vitro*, Vol. 19/6, pp. 831-842, https://doi.org/10.1016/j.tiv.2005.04.009.

Araki, N. et al. (2005b), "Evaluation of a rapid *in vitro* androgen receptor transcriptional activation assay using AR-EcoScreen cells", *Toxicology In Vitro*, Vol. 19/3, pp. 335-352, https://doi.org/10.1016/j.tiv.2004.10.008.

Birkhøj, M. et al. (2004), "The combined antiandrogenic effects of five commonly used pesticides", *Toxicology and Applied Pharmacology*, Vol. 201/1, pp. 10-20, https://doi.org/10.1016/j.taap.2004.04.016.

Hartig, P.C. et al. (2002), "Development of two androgen receptor assays using adenoviral transduction of MMTV-luc reporter and/or hAR for endocrine screening", *Toxicological Sciences*, Vol. 66/1, pp. 82-90.

Jacobs, M. et al. (2013), "*In vitro* metabolism and bioavailability tests for endocrine active substances: What is needed next for regulatory purposes?", *ALTEX – Alternatives to Animal Experimentation*, Vol. 30/3, pp. 331-351.

Jacobs, M.N. et al. (2008), "The use of metabolising systems for *in vitro* testing of endocrine disrupters", *Current Drug Metabolism*, Vol. 9/8, pp. 796-826.

OECD (2011), "Peer review report for the validation of the Stably Transfected Transcriptional Activation Assay for the detection androgenic and anti-androgenic activity of chemicals", OECD Series on Testing and Assessment, No. 161, OECD, Paris, www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/JM/MONO(2011)46&doclanguage=en.

OECD (2008), *Detailed Review Paper on the Use of Metabolising Systems for* In Vitro *Testing of Endocrine Disrupters*, OECD Series on Testing and Assessment, No. 97, OECD Publishing, Paris, https://doi.org/10.1787/9789264085497-en.

Sohoni, P. and J.P. Sumpter (1998), "Several environmental oestrogens are also anti-androgens", *Journal of Endocrinology*, Vol. 158/3, pp. 327-339.

Table C.1.3. **Stably Transfected Human Androgen Receptor Transcriptional Activation Assay for Detection of Androgenic Agonist and Antagonist Activity of Chemicals (AR STTA) (OECD TG 458): Guidance for scenarios of combinations of results with existing data**

This table represents possible conclusions to be drawn from assay data, and a next step which could be taken if further evidence is required about possible endocrine disrupting properties and/or effects. The guidance offered is not meant to be prescriptive, but provides science-based considerations. It encourages the use of all available data and expert judgement in a weight of evidence approach. Regional and national interpretation of results and "next steps" may vary.

The conclusions are grouped into a series of scenarios (A-R), each scenario representing a different combination of assay results, existing *in vitro* data and existing *in vivo* data. The symbol "+" indicates that the data in question represent a positive result, "-" indicates a negative result, and "Eq/0" indicates that the data are either equivocal or are not available.

Existing results: * "Mechanism (*in vitro* mechanistic data)" assumes that mechanistic data are available from estrogen receptor (ER-), androgen receptor (AR-) and steroidogenesis (S-) based assays (Level 2). The AR Binding Assay is likely to be performed prior to the AR STTA assay. TR and other assays concerning mechanisms of thyroid disruption may be available, but they are not in common use. In practice, data from all assays may not be available and therefore this must be taken into account when deciding on the "next step". Quantitative structure activity relationship (QSAR) predictions of estrogen and androgen binding/activation may be made for some substances.

Existing results: ** "Effects (*in vivo* effects of concern)" assumes various information, such as data from repeat dose oral toxicity studies, reproduction/developmental toxicity screen tests, read-across from analogues, will be available.

Scenarios	Result of AR STTA	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
A	+	+	+	Androgen receptor (AR) (ant)agonism combined with effects on estrogen receptor/thyroid/steroidogenesis (ER/T/S) and potential for adverse effects via multiple mechanisms.	Perform assay from Levels 3-5, e.g. Hershberger (H) Assay or fish screen (AFSS or JMASA) (Level 3) or male Peripubertal (PP) Assay (Level 4) or EOGRTS or two-generation assays or partial/full non-mammalian wildlife life cycle tests, e.g. OECD TG 241 and TG 240 (Level 4/5).	A positive result indicates strong probability of interaction with AR in other taxa. If existing data are from an adequate Level 5 assay, then there may be sufficient information to conclude evidence of concern for endocrine disruption (the EOGRTS provides the most information; however, for endocrine disrupting chemicals [EDCs] with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from Level 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism. If existing data are from H assays or AFSS or JMASA, then Level 4 mammalian assays or Level 3 or 4 fish screens (OECD TG 229/230/234) will provide data on multiple modalities. Consider route of exposures for existing effects data and possible transformation products and implications of absorption, distribution, metabolism and excretion (ADME) characteristics of the chemical. The chemical itself may give positive *in vitro* results and *in vivo* results but may also be metabolised to a metabolite that also has positive results *in vitro* and *in vivo*. However, note that uptake and metabolism of chemicals can be different between mammalian and non-mammalian wildlife species.
B	+	+	–	AR (ant)agonism combined with effects on ER/T/S but effects not detected in *in vivo* studies. Weak AR (ant)agonism does not result in adverse effects in the selected species under the conditions of the test. Metabolic differences may explain *in vitro*/*in vivo* differences.	Perform AR STTA with added metabolising system or Assay from Levels 3-4, e.g. H assay or fish screen (AFSS) (Level 3) or male PP assay (Level 4).	A positive result indicates strong probability of interaction with AR in other taxa. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from Level 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism. If existing data are from H assays or AFSS or JMASA, then Level 4 mammalian assays or Level 3 or 4 fish screens (OECD TG 229/230/234) will provide data on multiple modalities. Consider route of exposures for existing effects data and possible transformation products and implications of ADME characteristics of the chemical. The chemical itself may give positive *in vitro* results but may not be absorbed or may be metabolised to an inactive metabolite *in vivo*. However, note that uptake and metabolism of chemicals can be different between mammalian and non-mammalian wildlife species.
C	+	+	Eq/0	AR (ant)agonism combined with effects on ER/T/S but no or equivocal data from *in vivo* studies. Weak AR (ant)agonism may not result in adverse effects in the selected species under the conditions of the test.	Perform assay from Levels 3-4, e.g. H assay or fish screen (AFSS) (Level 3) or male PP assay (Level 4).	A positive result indicates strong probability of interaction with AR in other taxa. Consider route of exposures for equivocal existing effects data and possible implications of ADME characteristics of the chemical. Equivocal results may indicate chemical has multiple modes of action (MOA). Check data on chemical analogues.

Scenarios	Result of AR STTA	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
D	+	–	+	AR (ant)agonism and potential for adverse effects.	Perform assay from Levels 3-4, e.g. H assay or fish screen (AFSS) (Level 3) or male PP assay (Level 4).	A positive result indicates strong probability of interaction with AR in other taxa. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude evidence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from Level 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism. If existing data are from H assays or AFSS or JMASA, then Level 4 mammalian assays or Level 3 or 4 fish screens (OECD TG 229/230/234) will provide data on multiple modalities. A positive result could have arisen from other (E,A,T,S or non-E,A,T,S) mechanisms, e.g. hypothalamic/pituitary/gonadal (HPG) axis.
E	+	–	–	AR (ant)agonism but effects not detected in *in vivo* studies. Weak AR (ant)agonism does not result in adverse effects in the selected species under the conditions of the test. Metabolic differences may explain *in vitro*/*in vivo* differences.	Perform AR STTA with added metabolising system or Assay from Levels 3-4, e.g. H assay or fish screen (AFSS) (Level 3) or male PP assay (Level 4).	A positive result indicates strong probability of interaction with AR in other taxa. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive) If existing data are from Level 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism. If existing data are from H assays or AFSS or JMASA, then Level 4 mammalian assays or Level 3 or 4 fish screens (OECD TG 229/230/234) will provide data on multiple modalities. Consider route of exposures for existing effects data and possible transformation products and implications of ADME characteristics of the chemical. The chemical itself may give positive *in vitro* results but may not be absorbed or may be metabolised to an inactive metabolite *in vivo*. However, note that uptake and metabolism of chemicals can be different between mammalian and non-mammalian wildlife species. Check data on chemical analogues.
F	+	–	Eq/0	AR (ant)agonism but no or equivocal data from *in vivo* studies. Weak AR (ant)agonism does not result in adverse effects in the selected species under the conditions of the test.	Perform assay from Levels 3-4, e.g. H assay or fish screen (AFSS) (Level 3), male PP assay (Level 4).	A positive result indicates strong probability of interaction with AR in other taxa. Consider route of exposures for equivocal existing effects data and possible implications of ADME characteristics of the chemical. Equivocal results may indicate chemical has multiple MOA. Check data on chemical analogues.

Scenarios	Result of AR STTA	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
G	+	Eq/0	+	AR (ant)agonism and potential for adverse effects via AR (ant)agonism or other E,T,S mechanisms. May act via E,A,T,S mechanism and may or may not require metabolic activation.	Perform assay from Levels 3-4, e.g. H assay or fish screen (AFSS) (Level 3) or male PP assay (Level 4).	A positive result indicates strong probability of interaction with AR in other taxa. If existing data are from an adequate Level 5 assay, then there may be sufficient information to conclude evidence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from Level 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism. If existing data are from H assays or AFSS or JMASA, then Level 4 mammalian assays or Level 3 or 4 fish screens (OECD TG 229/230/234) will provide data on multiple modalities. Check data on chemical analogues. Further mechanistic studies may help determine MOA. Equivocal results may indicate chemical has multiple MOA. A positive result could have arisen from other (E,A,T,S or non-E,A,T,S) mechanisms (e.g. HPG axis).
H	+	Eq/0	–	AR (ant)agonism but effects not detected in *in vivo* studies. Weak AR (ant)agonism does not result in adverse effects in the selected species under the conditions of the test. Metabolic differences may explain *in vitro*/*in vivo* differences.	Perform AR STTA with added metabolising system.	A positive result indicates strong probability of interaction with AR in other taxa. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential OECD TG 451-3 may be more sensitive). If existing data are from Level 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism. If existing data are from H assays or AFSS or JMASA, then Level 4 mammalian assays or Level 3 or 4 fish screens (OECD TG 229/230/234) will provide data on multiple modalities. Consider route of exposures for existing effects data and possible transformation products and implications of ADME characteristics of the chemical. The chemical itself may give positive *in vitro* results but may not be absorbed or may be metabolised to an inactive metabolite *in vivo*. However, note that uptake and metabolism of chemicals can be different between mammalian and non-mammalian wildlife species. Check data on chemical analogues. Further mechanistic studies may help determine MOA. Equivocal results may indicate chemical has multiple MOA.
I	+	Eq/0	Eq/0	AR (ant)agonism with unknown potential for effects in *in vivo* studies. May act via AR mechanism and may or may not require metabolic activation. Unknown potential for adverse effects.	Perform AR STTA with added metabolising system or H assay or fish screen (AFSS) (Level 3) if existing data indicate this is needed.	A positive result indicates strong probability of interaction with AR in other taxa. Consider route of exposures for equivocal existing effects data and possible implications of ADME characteristics of the chemical. Check data on chemical analogues. Equivocal results may indicate chemical has multiple MOA.

Scenarios	Result of AR STTA	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
J	–	+	+	No evidence for AR (ant)agonism. Effects on ER/T/S and potential for adverse effects via E,A,T,S mechanisms.	Perform AR STTA with added metabolising system or Perform assay from Levels 3-4, e.g. H assay or fish screen AFSS (Level 3) or male PP assay (Level 4).	A negative result indicates interaction with AR in other taxa is unlikely. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from Level 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism. If existing data are from H assay or AFSS or JMASA, then Level 4 mammalian assays or Level 3 or 4 fish screens (OECD TG 229/230/234) will provide data on multiple modalities. Consider route of exposures for existing effects data and possible transformation products and implications of ADME characteristics of the chemical. Check data on chemical analogues. Further mechanistic studies may help determine MOA.
K	–	+	–	No evidence for AR (ant)agonism. Effects on ER/T/S but effects not detected in *in vivo* studies.	Perform assay from Levels 3-4, e.g. fish screen (AFSS) (Level 3) or male or female PP assay (Level 4).	A negative result indicates interaction with AR in other taxa is unlikely. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive) If existing data are from Level 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism. If existing data are from H assays or AFSS or JMASA, then Level 4 mammalian assays or Level 3 or 4 fish screens (OECD TG 229/230/234) will provide data on multiple modalities. Metabolic deactivation of chemical may occur *in vivo* so that possible *in vitro* E,A,T,S activity is not realised. Consider possible routes of exposure, implications of metabolism.
L	–	+	Eq/0	No evidence for AR (ant)agonism. Effects on ER/T/S but effects not detected in *in vivo* studies. Unknown potential for adverse effects. Metabolic differences explain *in vitro*/*in vivo* E,A,T,S differences.	Perform assay from Levels 3-4, e.g. fish screen (AFSS) (Level 3) or male or female PP assay (Level 4).	A negative result indicates interaction with AR in other taxa is unlikely. Metabolic deactivation of chemical may occur *in vivo* so that possible *in vitro* E,A,T,S activity is not realised. Consider possible routes of exposure, implications of metabolism. Check data on chemical analogues. Equivocal results may indicate chemical has multiple MOA.

Scenarios	Result of AR STTA	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
M	–	–	+	No evidence for AR (ant)agonism. Metabolic differences or route of exposure explain *in vitro*/*in vivo* differences. Effects seen in existing studies are via non-E,A,T,S or non-endocrine mechanisms.	Perform AR STTA with added metabolising system or Perform assay from Levels 3-4, e.g. fish screen (AFSS) (Level 3) or male or female PP assay (Level 4).	A negative result indicates interaction with AR in other taxa is unlikely. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from Level 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism. If existing data are from H assays or AFSS or JMASA, then Level 4 mammalian assays or Level 3 or 4 fish screens (OECD TG 229/230/234) will provide data on multiple modalities. Metabolic activation of chemical may occur *in vivo*. Consider possible routes of exposure, implications of metabolism. Check data on chemical analogues. Further mechanistic studies may help determine MOA.
N	–	–	–	No evidence for AR (ant)agonism. No evidence of adverse effects.	Possibly no need for further testing. If there is uncertainty, may perform assay from Levels 3-4, e.g. fish screen (AFSS) (Level 3) or male or female PP assay (Level 4).	A negative result indicates interaction with AR in other taxa is unlikely. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from Level 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism. If existing data are from H assays or AFSS or JMASA, then Level 4 mammalian assays or Level 3 or 4 fish screens (OECD TG 229/230/234) will provide data on multiple modalities. Check data on chemical analogues.
O	–	–	Eq/0	No evidence for AR (ant)agonism. Unknown potential for adverse effects via other mechanisms.	Perform AR STTA with added metabolising system or Fish screen (AFSS) (Level 3) or male or female PP assay (Level 4) if existing data indicate this is needed.	A negative result indicates interaction with AR in other taxa is unlikely. Consider possible routes of exposure, implications of metabolism. Check data on chemical analogues.

Scenarios	Result of AR STTA	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
P	–	Eq/0	+	No evidence for AR (ant)agonism. Unknown potential for adverse effects via other mechanisms.	Perform AR STTA with added metabolising system.	A negative result indicates interaction with AR in other taxa is unlikely. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from Level 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism. If existing data are from H assays or AFSS or JMASA, then Level 4 mammalian assays or Level 3 or 4 fish screens (OECD TG 229/230/234) will provide data on multiple modalities. Consider possible routes of exposure, implications of metabolism. Check data on chemical analogues. Further mechanistic studies may help determine MOA. Equivocal results may indicate chemical has multiple MOA.
Q	–	Eq/0	–	No evidence for AR (ant)agonism. No evidence of adverse effects.	Perform AR STTA with added metabolising system.	A negative result indicates interaction with AR in other taxa is unlikely. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from Level 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism. If existing data are from H assays or AFSS or JMASA, then Level 4 mammalian assays or Level 3 or 4 fish screens (OECD TG 229/230/234) will provide data on multiple modalities. Check data on chemical analogues. Further mechanistic studies may help determine MOA.
R	–	Eq/0	Eq/0	No evidence for AR (ant)agonism. Unknown potential for adverse effects via other mechanisms.	For the "0" scenario, perform AR STTA with added metabolising system or Perform H assay or fish screen (AFSS) (Level 3) if existing data indicate this is needed.	A negative result indicates interaction with AR in other taxa is unlikely. Consider possible routes of exposure, implications of metabolism. Check data on chemical analogues. Further mechanistic studies may help determine MOA. Equivocal results may indicate chemical has multiple MOA.

C.1.4. H295R Steroidogenesis Assay (OECD TG 456)

Status: Assay validated by the OECD.

Modality detected/endpoints: Interference with steroidogenesis/inhibition and induction of estradiol and testosterone synthesis.

Background to the assay

213. The H295R Steroidogenesis Assay is an *in vitro* screening assay to detect substances that affect production of estradiol and testosterone. OECD TG 456 was published in July 2011. It provides a positive or negative result for the ability of a chemical to induce or inhibit the production of estradiol and testosterone. The assay utilises a human adrenocarcinoma cell line (NCI-H295R cells) that have the characteristics of undifferentiated human fetal adrenal cells. This cell line expresses all the key enzymes involved in steroidogenesis, from cholesterol to estradiol and testosterone. This expression would allow for the detection of other hormones. An "enhanced" Steroidogenesis Assay using H295R cells, where many other hormones are analysed, has been published (Wang et al., 2014). However, the OECD assay validation only included estradiol and testosterone. The cells represent a unique *in vitro* system because *in vivo*, expression of these enzymes is developmental stage specific with no one tissue expressing all the enzymes at once.

214. Chemicals may induce steroidogenesis; this can be determined by increased production of estradiol and testosterone. Alternatively, chemicals may inhibit steroidogenesis; this can be determined by decreased production of estradiol and testosterone. Results are expressed as fold change compared with the negative control. In the validation of the assay, forskolin induced estradiol and testosterone production whilst prochloraz inhibited estradiol and testosterone production. The validation of the Steroidogenesis Assay demonstrated that whilst not always directly predictive of a specific type of response *in vivo*, the chemicals chosen in the validation studies would always be flagged as a disrupter of steroidogenesis or a reproductive toxicant (OECD, 2010). The assay is therefore used somewhat as a "black box" where a positive result indicates that a chemical is a possible disrupter of steroidogenesis but without defining the exact mechanism of action.

215. An adequate response with positive control chemicals (forskolin and prochloraz), and other proficiency chemicals, is required in the OECD test guidance (TG) to demonstrate laboratory proficiency. The assay also requires the assessment of the cytotoxic effect of a chemical, as measurement of cell viability is an important feature of the TG. A minimum of 80% cell viability is needed for the hormone production assessment to be considered adequate. Limitations of the assay are that xenobiotic metabolising capability is unknown, but likely to be limited and production of other hormones (e.g. gluco- and mineralocorticoids) by the cells may affect estradiol and testosterone levels. The current assay does not detect 5-alpha reductase inhibitors (e.g. finasteride) that inhibit the conversion of testosterone to dihydrotestosterone. Although 5-alpha reductase is present in H295R cells, dihydrotestosterone is not a validated endpoint and therefore these chemicals

will not be identified. 5-alpha reductase inhibitors are detected by OECD TG 441 (Hershberger [H] assay).

216. The assay will not detect substances that act by affecting the hypothalamic/ pituitary/gonadal (HPG) as an *in vivo* intact axis is required for this. The effect of androgen receptor (AR), estrogen receptor (ER) and thyroid hormone receptor (TR) ligands on this assay is also not clear, although the Steroidogenesis Assay is not designed to detect these substances, it is not known whether they affect steroidogenesis. These chemicals will, however, be detected in AR-, ER- and TR-specific assays and therefore results from a suite of *in vitro* tests should be considered together.

217. The Steroidogenesis Assay requires that strict control is made of the age at which the cells are used. The capacity of the cells to produce estradiol changes with increasing number of cell passages. In addition, chemicals and cell matrices may interfere with hormone measurements. The TG includes quality control measures to ensure the accuracy and reliability of results. It is recommended that compliance with the quality control criteria and with the performance criteria for the positive control substances forskolin and prochloraz and with the other proficiency chemicals is demonstrated before evaluating results from this assay. Small deviations are unlikely to have compromised the assay, but judgement should be made on a case-by-case basis.

When/why the assay may be used

218. Although the Steroidogenesis Assay may be used at any stage in the hazard assessment process, the most likely use scenario is during initial assessment of chemicals for their ability to interact with endocrine systems *in vitro*, i.e. estrogen/androgen/thyroid/ steroidogenesis (E,A,T,S) modalities. Assays for interaction with other modalities (e.g. AR and ER), are likely to be conducted at the same time so that all results can be considered together. TR and other assays concerning mechanisms of thyroid disruption may be available, but they are not in common use. Data from the aromatase assay may also be available, chemicals testing positive in this assay are likely to also give positive results in the Steroidogenesis Assay as aromatase is one of the key enzymes in the steroidogenesis pathway. The steroidogenesis TG does not include the use of a xenobiotic metabolising system, but consideration should be given to the inclusion of this (Jacobs et al., 2008, 2013; OECD, 2008) depending on the circumstances (e.g. if the metabolism of a chemical is unknown), although the methods for inclusion of xenobiotic metabolising systems are not yet validated (see Paragraph 50). Alternatively, for a chemical with known metabolites, these could also be tested in the Steroidogenesis Assay. Another use scenario may be following effects obtained in higher tier tests, for example delayed puberty onset in females, but which are not exclusively indicative of an effect on ER. Selection of the most appropriate tests has to be on a case-by-case basis, but also considering the need to minimise animal testing. A further example could be results obtained in other apical assays, e.g. OECD TG 408 (90-day toxicity test), where effects on reproductive organs may be investigated further by testing in the Steroidogenesis Assay in combination with AR- and ER-based assays.

Introduction to the table of scenarios

219. Table C.1.4 gives guidance on a further step to take in the event of a positive (+) or negative (-) result in the Steroidogenesis Assay and in the presence of positive (+), negative (-) or equivocal/absent (Eq/0) existing results. “Existing results” are subdivided into “mechanism” and “effects” data (third and fourth columns). The table is divided horizontally into a series of scenarios that represent all the combinations of these events.

220. The results of the Steroidogenesis Assay are given in the second column. Criteria for positive results are given in the draft test guideline. A result is judged positive if the fold difference is statistically significant from the solvent control at two adjacent concentrations in at least two tests, or when a single concentration data point is significantly different from the solvent control, and this can be confirmed by being significantly different in at least one more run within a +/- 1 concentration increment of the respective experiment. The latter allows for effects that may be seen close to the maximum concentration (1mM). It is important that quality and proficiency criteria are demonstrated for both positive and negative results.

221. Equivocal results for the guideline are not included in the table because these data generally require further interrogation about the result itself. This assay is a screen and therefore a clear positive or negative result should be obtained. In the event of an equivocal result, the considerations mentioned above about control quality and proficiency criteria should be taken into account and further investigations made. Equivocal results at high concentrations may may indicate solubility issues.

Existing data to be considered

222. Existing "mechanism" *in vitro* data are assumed to be available from ER- and AR-based assays and the aromatase assay (Level 2). Assays may also be available for interference with thyroid modalities. In practice, it is possible that data from all of these assays may not be available, so judgement will need to be used to decide which assays to perform.

223. Existing "effects" data refer to *in vivo* effects "of concern" (i.e. data from Level 4 or 5 vertebrate wildlife assays). These may come from varied sources and will depend on the type of substance (e.g. new chemicals, high production volume [HPV] chemical, pesticide). Thus, available data may range from repeated dose toxicity studies (28-day, 90-day), combined repeat dose/reproductive screening assays or fish screening assays, to chronic toxicity studies and multigeneration reproductive tests in vertebrate wildlife species. Some studies fail to identify endocrine disruptors (EDs) that weakly affect estrogen or androgen receptors, as was demonstrated on the basis of data generated in the validation process of the OECD TG 407 assay with endocrine endpoints. In this validation, only moderate EDs such as nonylphenol and DDE, and strong EDs such as ethinylestradiol and flutamide (acting via ER and AR respectively) were detected. The aromatase inhibitor CGS 18320B was detected by the OECD TG 407 assay, but this chemical was developed as a pharmaceutical aromatase inhibitor and therefore is a strong ED. The ability to detect chemicals that weakly interfere with steroidogenesis is not known. Thus, OECD TG 407 cannot be regarded as a screening assay for endocrine activity. This means that when a relatively insensitive test is positive for both endocrine-specific and apical endpoints, this should be taken as an indication that the substance is a potential ED. Caution should be exercised, however, because endocrine endpoints may be impacted secondary to non-endocrine toxicity and *in vivo* apical endpoints can be affected by many modes of action, including endocrine modalities. Data may also be available on effects in mammalian and non-mammalian wildlife species, although caution should be used when extrapolating between taxa. A chemical causing endocrine effects in non-mammalian vertebrates may also have endocrine effects in mammals, but the physiological consequences of the effects are likely to be different.

224. Data may also be available from Hershberger (H) and Uterotrophic (UT) Assays (Level 3), but as these assays do not generally detect steroidogenesis interference, they are only useful in these cases for purposes of elimination.

225. When considering the results of the Steroidogenesis Assay, all available data should be used in order to reach a conclusion and a weight of evidence approach taken. This may include high throughput screening (HTS) data, read-across data from structural analogues and quantitative structure activity relationship (QSAR). Several QSAR models for ER and AR binding/activation are now available (see Sections B.1.1.1 and B.1.1.2).

Scenarios: Positive and negative results combined with existing data

226. The scenarios (A to R) presented in Table C.1.4 represent all the possibilities of positive or negative results in combination with the presence or absence of existing data. The action taken will also depend on the regulatory environment, but the considerations given here are generally science based. Although OECD TG 456 uses a human cell line, steroidogenic pathways relevant for androgen and estrogen synthesis are well conserved across taxa and therefore results results in this assay are likely to be relevant to other vertebrate species. Differences in steroidogenesis pathways, however, exist across species/cell/stages of development (for reviews see Scott, Mason and Sharpe [2009]; and Payne and Hales [2004]) and this should also be taken in account. Wherever possible, the recommended "next step which could be taken" avoids unnecessary animal testing. However, sometimes conducting an animal test will be indicated and then the relevance of species, strain, exposure route and species-specific metabolism should always be considered. Further considerations specific to each scenario are given in the table.

227. Scenarios A to C represent positive results in the Steroidogenesis Assay in the presence of positive *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. A positive result in a Steroidogenesis Assay is strong evidence for disruption of steroidogenesis that may or may not be supported by the *in vivo* effects data. Inhibition of steroidogenesis (but not induction) could be followed up by a confirmatory aromatase assay if this is not already available. In the case of positive *in vivo* effects data, there may be sufficient evidence to conclude concern for endocrine disruption and therefore no need for further screening. *In vivo* assays/tests with negative results should be interpreted with caution as they may either indicate that the tests used do not have sufficient power to detect weak effects or, alternatively, that the effects do not present a concern for endocrine disruption. Generally, a conclusion of lack of concern for endocrine disruption in the presence of positive effects data may only be made given adequate Level 5 assays. Information on some endocrine-related tumours may be detected more comprehensively in carcinogenicity studies (OECD TG 451/453) (Level 4); for example, detection of certain types of thyroid tumors in the absence of reproductive or developmental effects, as well as substances causing tumors in other endocrine-sensitive tissues. MOA data to provide a clear interpretation may be required by some regulatory agencies.

228. Scenarios D to F represent positive results in the Steroidogenesis Assay in the presence of negative *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. As above, inhibition of steroidogenesis could be followed up by a confirmatory aromatase assay if this is not already available. Unless the metabolic profile of the test substance is known, one option may be to conduct these *in vitro* assays with an added metabolising system. If the metabolic profile is known, then an *in vivo* test may be advisable. The choice of tests will depend on the available *in vivo* effects data. As in Scenarios A to C, *in vivo* assays/tests with negative results should be interpreted with

caution as they may either indicate that the tests used do not have sufficient power to detect weak effects or, alternatively, that the effects do not present a concern for endocrine disruption. Generally, a conclusion of lack of concern for endocrine disruption in the presence of positive effects data (Scenario E) may only be made given adequate Level 5 assays. Information on some endocrine-related tumours may be detected more comprehensively in carcinogenicity studies (OECD TG 451/453) (Level 4); for example, detection of certain types of thyroid tumors in the absence of reproductive or developmental effects, as well as substances causing tumors in other endocrine-sensitive tissues.

229. Scenarios G to I represent positive results in the Steroidogenesis Assay in the presence of various combinations of missing or equivocal data. As above, inhibition of steroidogenesis could be followed up by a confirmatory aromatase assay if this is not already available. The next step to take for missing or equivocal data will depend on the nature of the other available data and the jurisdiction in which it is being used. In some cases, equivocal data may be viewed as positive whilst in others it may or may not contribute to the weight of evidence. The interpretation may also depend on the MOA in question and why the data are considered equivocal, e.g. a study that is equivocal for thyroid effects may still be of value in evaluating (anti)androgenic effects. In all three scenarios, the recommended first step is to obtain reliable mechanistic (*in vitro*) data rather than proceed directly to *in vivo* testing. Equivocal and missing data are alternative scenarios and two possibilities for the next step are given in most cases, but the nature of equivocal data means that decisions need to be taken on a case-by-case basis. In all cases, the role of metabolism, route of exposure and data from structural analogues should be considered before deciding on the next step. As above, generally a conclusion of lack of concern for endocrine disruption in the presence of some positive effects data (Scenario H) may only be made given adequate Level 5 assays. Information on some endocrine-related tumours may be detected more comprehensively in carcinogenicity studies (OECD TG 451/453) (Level 4); for example, detection of certain types of thyroid tumors in the absence of reproductive or developmental effects, as well as substances causing tumors in other endocrine-sensitive tissues. MOA data to provide a clear interpretation may be required by some regulatory agencies.

230. Scenarios J to L represent negative results in the Steroidogenesis Assay in the presence of positive *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. The limitations of the Steroidogenesis Assay should be considered first (e.g. lack of metabolic activation, possible involvement of other factors). The positive *in vitro* mechanistic data indicates possible alternative EAT mechanisms. To confirm lack of steroidogenesis activity in the presence of *in vivo* data, a steroidogenesis with added metabolising capability could be performed. Otherwise *in vivo* tests will confirm or refute E,A,T,S activity.

231. Scenarios M to O represent negative results in the Steroidogenesis Assay in the presence of negative *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. Negative results for all tests (Scenario N) may be sufficient to enable a conclusion of no concern for endocrine disruption. This will depend on the weight of evidence and may not be possible in some cases. However, in the presence of negative data from robust Level 4 and 5 assays, further animal testing is probably not justified. The limitations of the Steroidogenesis Assay should also be considered (as described for Scenarios J to L). To confirm lack of steroidogenesis-related activity in the presence of *in vivo* data, a Steroidogenesis Assay with added metabolising capability could be performed. Otherwise, *in vivo* tests will confirm or refute E,A,T,S activity (Scenarios M to O).

232. Scenarios P to R represent negative results in the Steroidogenesis Assay in the presence of various combinations of missing or equivocal data. The limitations of the Steroidogenesis Assay should be considered first (as described for Scenarios J to L). As with the positive result scenarios above (see Paragraph 229), the next step to take for Scenarios P to R will have to be decided on a case-by-case basis. However, the recommended first step is generally to obtain reliable mechanistic (*in vitro*) data rather than proceed directly to *in vivo* testing. In all cases, the role of metabolism, route of exposure and data from structural analogues should be considered before deciding on the next step.

233. In all scenarios (A to R), the next step to take to strengthen weight of evidence will depend on the existing information. Table C.1.4 is meant to provide a succinct guide and may not cover all circumstances or possibilities. The scenarios may also suggest that chemicals have simple or single MOA, when in practice they may have multiple endocrine and non-endocrine MOA. In some cases, for example, two opposite modes of simultaneous action (e.g. estrogenic and anti-estrogenic) could, depending on dose, lead to a minimisation or abolition of effects, while in others two different MOA (e.g. estrogenic and anti-androgenic) could potentially reinforce effects. Endocrine pathways interact and there are many for which no TGs yet exist. If multiple MOA are suspected, either from the existing results or based on QSAR/read-across/integrated approaches, this should be investigated further, if needed for regulatory decision making.

234. In general, a decision about whether or not to conduct *in vivo* vertebrate wildlife tests will depend on the weight of evidence of new and existing data. If most available data (e.g. the results of the Steroidogenesis Assay, predictions from QSARs, "read-across" from data on similar substances and results from mammalian *in vivo* assays) suggest that the substance has the potential to cause endocrine disruption via interference with steroidogenesis (i.e. the level of suspicion about endocrine disrupting action is high – corresponding to Scenario A), then consideration should be given to the conduct of a higher level test.

235. For non-mammalian wildlife species, higher level tests with fish or amphibians (i.e TG 234 [FSDT], TG 240, TG 241) are recommended. Choice about which of these tests is most appropriate will be driven *inter alia* by mode of action considerations, and by whether multigeneration effects are to be expected. Such tests are unlikely to be needed if exposure of the natural environment is not expected. On the other hand, if available data only raise a low or moderate level of suspicion about endocrine disrupting action (e.g. the data appear to conflict with each other), then consideration should be given to the conduct of a fish or amphibian screen (i.e. OECD TG 229 or TG 230). There are fewer options available for invertebrates, but if ecdysteroid or juvenile hormone activity are suspected in arthropods (e.g. from a screening test with SJHASA), various higher level tests are available, including OECD GD 201, the *Daphnia* Multigeneration Test, and TG 233.

236. For mammals, similar considerations apply, but lower level tests (e.g. Level 3 or 4) should be conducted before higher level tests in order to avoid unnecessary animal usage, unless it is apparent that a Level 5 test will be required anyway or will be needed to establish the evidence to conclude on ED properties. At Level 5, the EOGRTS (OECD TG 443) is the most sensitive reproduction assay for detecting endocrine disruption because it includes evaluation of a number of endocrine endpoints not included in the two-generation study (OECD TG 416) adopted in 2001. It is recognised, however, that some jurisdictions may require a two-generation study.

References

Jacobs, M. et al. (2013), "*In vitro* metabolism and bioavailability tests for endocrine active substances: What is needed next for regulatory purposes?", *ALTEX – Alternatives to Animal Experimentation*, Vol. 30/3, pp. 331-351.

Jacobs, M.N. et al. (2008), "The use of metabolising systems for *in vitro* testing of endocrine disrupters", *Current Drug Metabolism*, Vol. 9/8, pp. 796-826.

OECD (2010), "Peer review report for the H295R Cell-Based Assay for Steroidogenesis", OECD Series on Testing and Assessment, No. 133, OECD, Paris, www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2010)32&doclanguage=en.

OECD (2008), *Detailed Review Paper on the Use of Metabolising Systems for* In Vitro *Testing of Endocrine Disrupters*, OECD Series on Testing and Assessment, No. 97, OECD Publishing, Paris, https://doi.org/10.1787/9789264085497-en.

Payne, A.H. and D.B. Hales (2004), "Overview of steroidogenic enzymes in the pathway from cholesterol to active steroid hormones", *Endocrine Reviews*, Vol. 25/6, pp. 947-970, https://doi.org/10.1210/er.2003-0030.

Scott, H.M., J.I. Mason and R.M. Sharpe (2009), "Steroidogenesis in the fetal testis and its susceptibility to disruption by exogenous compounds", *Endocrine Reviews*, Vol. 30/7, pp. 883-925, https://doi.org/10.1210/er.2009-0016.

Wang, S. et al. (2014), "Extending an *in vitro* panel for estrogenicity testing: The added value of bioassays for measuring antiandrogenic activities and effects on steroidogenesis", *Toxicological Sciences*, Vol. 141/1, pp. 78-89, https://doi.org/10.1093/toxsci/kfu103.

Table C.1.4. **H295R Steroidogenesis Assay (OECD TG 456): Guidance for scenarios of combinations of results with existing data**

This table represents possible conclusions to be drawn from assay data, and a next step which could be taken if further evidence is required about possible endocrine disrupting properties and/or effects. The guidance offered is not meant to be prescriptive, but provides science-based considerations. It encourages the use of all available data and expert judgement in a weight of evidence approach. Regional and national interpretation of results and "next steps" may vary.

The conclusions are grouped into a series of scenarios (A-R), each scenario representing a different combination of assay results, existing *in vitro* data and existing *in vivo* data. The symbol "+" indicates that the data in question represent a positive result, "-" indicates a negative result, and "Eq/0" indicates that the data are either equivocal or are not available.

Existing results: * "Mechanism (*in vitro* mechanistic data)" assumes that mechanistic data are available from estrogen receptor (ER-) and androgen receptor (AR-) based assays (Level 2). Data on aromatase inhibition may also be available. Thyroid hormone receptor (TR) and other assays concerning mechanisms of thyroid disruption may be available, but they are not in common use. In practice, data from all assays may not be available and therefore this must be taken into account when deciding on the "next step". Quantitative structure activity relationship (QSAR) predictions of estrogen and androgen binding/activation may be made for some substances.

Existing results: ** "Effects (*in vivo* effects of concern)" assumes various information, such as data from repeat dose oral toxicity studies, reproduction/developmental toxicity screen tests, read-across from analogues, will be available.

Scenarios	Result of steroid-ogenesis assay	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
A	+	+	+	Inhibition/induction of steroidogenesis combined with effects on ER/AR/T/S and potential for adverse effects via multiple mechanisms.	Perform assay from Levels 3-5, e.g. male or female pubertal assay (Level 4) or EOGRTS or two-generation assays or partial/full non-mammalian wildlife life cycle tests, e.g. OECD TG 241 and TG 240 (Level 4/5).	A positive result indicates a possibility of interference with steroidogenesis in other taxa. If existing data are from a Level 5 assay, there may be sufficient information to conclude evidence of concern for endocrine disruption (the EOGRTS provides the most information; however, for endocrine disrupting chemicals [EDCs] with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from Level 3 or 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism, although some Level 4 assays (e.g. TG 234 [FSDT]) may be sufficient for this purpose. Compare Steroidogenesis Assay results with other *in vitro* results to help discern mechanism. Consider route of exposures for existing effects data and possible transformation products and implications of absorption, distribution, metabolism and excretion (ADME) characteristics of the chemical. The chemical itself may give positive *in vitro* results but may not be absorbed or may be metabolised to an inactive metabolite *in vivo*. However, note that uptake and metabolism of chemicals can be different between mammalian and non-mammalian wildlife species.
B	+	+	–	Inhibition/induction of steroidogenesis combined with effects on ER/AR/T but effects not detected in *in vivo* studies. Weak steroidogenesis inhibition/induction does not result in adverse effects in the selected species under the conditions of the test. Metabolic differences may explain *in vitro*/*in vivo* differences.	Perform Steroidogenesis Assay with added metabolising system or Assay from Levels 3-4, e.g. fish screen (OECD TG 229/230/234) (Level 3) or male or female pubertal assay (Level 4).	A positive result indicates a possibility of interference with steroidogenesis in other taxa. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from Level 3 or 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism, although some Level 4 assays (e.g. TG 234 [FSDT]) may be sufficient for this purpose. Compare Steroidogenesis Assay results with other *in vitro* results to help discern mechanism. Consider route of exposures for existing effects data and possible transformation products and implications of ADME characteristics of the chemical. The chemical itself may give positive *in vitro* results but may not be absorbed or may be metabolised to an inactive metabolite *in vivo*. However, note that uptake and metabolism of chemicals can be different between mammalian and non-mammalian wildlife species.

Scenarios	Result of steroid-ogenesis assay	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
C	+	+	Eq/0	Inhibition/induction of steroidogenesis combined with effects on ER/AR/T but no or equivocal data from *in vivo* studies. Weak steroidogenesis inhibition/induction may not result in adverse effects in the selected species under the conditions of the test.	Perform assay Levels 3-4, e.g. fish screen (OECD TG 229/230/234) (Level 3) or male or female pubertal assay (Level 4).	A positive result indicates a possibility of interference with steroidogenesis in other taxa. Compare Steroidogenesis Assay results with other *in vitro* results to help discern mechanism. Consider route of exposures for equivocal existing effects data and possible implications of ADME characteristics of the chemical. Equivocal results may indicate chemical has multiple modes of action (MOA). Check data on chemical analogues.
D	+	–	+	Inhibition/induction of steroidogenesis and potential for adverse effects.	Perform assay from Levels 3-4, e.g. fish screen (OECD TG 229/230/234) (Level 3) or male or female pubertal assay (Level 4).	A positive result indicates a possibility of interference with steroidogenesis in other taxa. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude evidence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from Level 3 or 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism, although some Level 4 assays (e.g. TG 234 [FSDT]) may be sufficient for this purpose. Further mechanistic studies may help determine MOA. A positive result could have arisen from other (E,A,T,S or non-E,A,T,S) mechanisms, e.g. hypothalamic/pituitary/gonadal (HPG) axis.
E	+	–	–	Inhibition/induction of steroidogenesis but effects not detected in *in vivo* studies. Weak steroidogenesis inhibition/induction does not result in adverse effects in the selected species under the conditions of the test. Metabolic differences may explain *in vitro*/*in vivo* differences.	Perform Steroidogenesis Assay with added metabolising system or Assay from Levels 3-4, e.g. fish screen (OECD TG 229/230/234) (Level 3) or male or female pubertal assay (Level 4).	A positive result indicates a possibility of interference with steroidogenesis in other taxa. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from Level 3 or 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism, although some Level 4 assays (e.g. TG 234 [FSDT]) may be sufficient for this purpose. Consider route of exposures for existing effects data and possible transformation products and implications of ADME characteristics of the chemical. The chemical itself may give positive *in vitro* results but may not be absorbed or may be metabolised to an inactive metabolite *in vivo*. However, note that uptake and metabolism of chemicals can be different between mammalian and non-mammalian wildlife species. Check data on chemical analogues.

Scenarios	Result of steroid-ogenesis assay	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
F	+	–	Eq/0	Inhibition/induction of steroidogenesis but no or equivocal data from *in vivo* studies. Weak steroidogenesis inhibition/induction may not result in adverse effects in the selected species under the conditions of the test.	Perform assay Levels 3-4, e.g. fish screen (OECD TG 229/230/234) (Level 3) or male or female pubertal assay (Level 4).	A positive result indicates a possibility of interference with steroidogenesis in other taxa. Consider route of exposures for equivocal existing effects data and possible implications of ADME characteristics of the chemical. Equivocal results may indicate chemical has multiple MOA. Check data on chemical analogues.
G	+	Eq/0	+	Inhibition/induction of steroidogenesis and potential for adverse effects via steroidogenesis interference or other EAT mechanisms. May act via non-steroidogenesis interference mechanism and may or may not require metabolic activation.	Perform assay from Levels 3-4, e.g. fish screen (OECD TG 229/230/234) (Level 3) or male or female pubertal assay (Level 4).	A positive result indicates a possibility of interference with steroidogenesis in other taxa. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude evidence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from Level 3 or 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism, although some Level 4 assays (e.g. TG 234 [FSDT]) may be sufficient for this purpose. Check data on chemical analogues. Further mechanistic studies may help determine MOA. A positive result could have arisen from other (E,A,T,S or non-E,A,T,S) mechanisms, e.g. HPG axis.
H	+	Eq/0	–	Inhibition/induction of steroidogenesis but effects not detected in *in vivo* studies. Weak steroidogenesis inhibition/induction does not result in adverse effects in the selected species under the conditions of the test. Metabolic differences may explain *in vitro*/*in vivo* differences.	Perform Steroidogenesis Assay with added metabolising system.	A positive result indicates a possibility of interference with steroidogenesis in other taxa. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from Level 3 or 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism. Consider route of exposures for existing effects data and possible transformation products and implications of ADME characteristics of the chemical. The chemical itself may give positive *in vitro* results but may not be absorbed or may be metabolised to an inactive metabolite *in vivo*. However, note that uptake and metabolism of chemicals can be different between non-mammalian wildlife species. Check data on chemical analogues. Further mechanistic studies may help determine MOA.

Scenarios	Result of steroid-ogenesis assay	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
I	+	Eq/0	Eq/0	Steroidogenesis inhibition/induction with unknown potential for effects in *in vivo* studies. May act via non-steroidogenesis interference mechanism and may or may not require metabolic activation. Unknown potential for adverse effects.	Perform Steroidogenesis Assay with added metabolising system or assay from Levels 3-4, e.g. fish screen (OECD TG 229/230/234) (Level 3) or male or female pubertal assay (Level 4) if existing data indicate this is needed.	A positive result indicates a possibility of interference with steroidogenesis in other taxa. Consider route of exposures for equivocal existing effects data and possible implications of ADME characteristics of the chemical. Check data on chemical analogues. Equivocal results may indicate chemical has multiple MOA.
J	–	+	+	No evidence for steroidogenesis interference. Effects on ER/AR/T and potential for adverse effects via EAT mechanisms.	Perform Steroidogenesis Assay with added metabolising system or Perform assay from Levels 3-4, e.g. fish screen (OECD TG 229/230/234) (Level 3) or male or female pubertal assay (Level 4).	A negative result indicates that interference with steroidogenesis in other taxa is unlikely. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude concern for endocrine disruption. If existing data are from Level 3 or 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism, although some Level 4 assays (e.g. TG 234 [FSDT]) may be sufficient for this purpose. Consider route of exposures for existing effects data and possible transformation products and implications of ADME characteristics of the chemical. Check data on chemical analogues. Further mechanistic studies may help determine MOA.
K	–	+	–	No evidence for steroidogenesis interference. Effects on ER/AR/T but effects not detected in *in vivo* studies. Metabolic differences explain *in vitro*/*in vivo* E,A,T,S differences.	Perform assay from Levels 3-4, e.g. fish screen (OECD TG 229/230/234) (Level 3) or male or female pubertal assay (Level 4).	A negative result indicates that interference with steroidogenesis in other taxa is unlikely. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from Level 3 or 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism, although some Level 4 assays (e.g. TG 234 [FSDT]) may be sufficient for this purpose. Metabolic deactivation of chemical may occur *in vivo* so that potential *in vitro* EAT activity is not realised. Consider possible routes of exposure, implications of metabolism.

Scenarios	Result of steroid-ogenesis assay	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
L	–	+	Eq/0	No evidence for steroidogenesis interference. Effects on ER/AR/T but effects not detected in *in vivo* studies. Unknown potential for adverse effects. Metabolic differences explain *in vitro*/*in vivo* EAT differences.	Perform assay from Levels 3-4, e.g. fish screen (OECD TG 229/230/234) (Level 3) or male or female pubertal assay (Level 4).	A negative result indicates that interference with steroidogenesis in other taxa is unlikely. Metabolic deactivation of chemical may occur *in vivo* so that potential *in vitro* EAT activity is not realised. Consider possible routes of exposure, implications of metabolism. Check data on chemical analogues. Equivocal results may indicate chemical has multiple MOA.
M	–	–	+	No evidence for steroidogenesis interference. Metabolic differences or route of exposure explain *in vitro*/*in vivo* differences. Effects seen in existing studies are via non-E,A,T,S or non-endocrine mechanisms.	Perform Steroidogenesis Assay with added metabolising system or Perform assay from Levels 3-4, e.g. fish screen (OECD TG 229/230/234) (Level 3) or male or female pubertal assay (Level 4).	A negative result indicates that interference with steroidogenesis in other taxa is unlikely. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude concern for endocrine disruption. If existing data are from Level 3 or 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism, although some Level 4 assays (e.g. TG 234 [FSDT]) may be sufficient for this purpose. Metabolic activation of chemical may occur *in vivo*. Consider possible routes of exposure, implications of metabolism. Check data on chemical analogues. Further mechanistic studies may help determine MOA.
N	–	–	–	No evidence for steroidogenesis interference. No evidence of adverse effects.	Possibly no need for further testing. If there is uncertainty, may perform assay from Levels 3-4, e.g. fish screen (OECD TG 229/230/234) (Level 3) or male or female pubertal assay (Level 4).	A negative result indicates that interference with steroidogenesis in other taxa is unlikely. If existing data are from adequate Level 4 or 5 assays, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from Level 3 or 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism, although some Level 4 assays (e.g. TG 234 [FSDT]) may be sufficient for this purpose. Check data on chemical analogues.
O	–	–	Eq/0	No evidence for steroidogenesis interference. Unknown potential for adverse effects via other mechanisms.	Perform Steroidogenesis Assay with added metabolising system or assay from Levels 3-4, e.g. fish screen (OECD TG 229/230/234) (Level 3) or male or female pubertal assay (Level 4) if existing data indicate this is needed.	A negative result indicates that interference with steroidogenesis in other taxa is unlikely. Consider possible routes of exposure, implications of metabolism. Check data on chemical analogues.

Scenarios	Result of steroid-ogenesis assay	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
P	–	Eq/0	+	No evidence for steroidogenesis interference. Unknown potential for adverse effects via other mechanisms.	Perform Steroidogenesis Assay with added metabolising system.	A negative result indicates that interference with steroidogenesis in other taxa is unlikely. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude concern for endocrine disruption, although some Level 4 assays (e.g. TG 234 [FSDT]) may be sufficient for this purpose. Consider possible routes of exposure, implications of metabolism. Check data on chemical analogues. Further mechanistic studies may help determine MOA. Equivocal results may indicate chemical has multiple MOA.
Q	–	Eq/0	–	No evidence for steroidogenesis interference. No evidence of adverse effects.	Perform Steroidogenesis Assay with added metabolising system.	A negative result indicates that interference with steroidogenesis in other taxa is unlikely. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from Level 3 or 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism, although some Level 4 assays (e.g. TG 234 [FSDT]) may be sufficient for this purpose. Check data on chemical analogues. Further mechanistic studies may help determine MOA.
R	–	Eq/0	Eq/0	No evidence for steroidogenesis interference. Unknown potential for adverse effects via other mechanisms.	For the "0" scenario, perform Steroidogenesis Assay with added metabolising system or Perform assay from Levels 3-4, e.g. fish screen (OECD TG 229/230/234) (Level 3) or male or female pubertal (Level 4) if existing data indicate this is needed.	A negative result indicates that interference with steroidogenesis in other taxa is unlikely. Consider possible routes of exposure, implications of metabolism. Check data on chemical analogues. Further mechanistic studies may help determine MOA. Equivocal results may indicate chemical has multiple MOA.

Non-OECD *in vitro* screens (Conceptual Framework Level 2)

C.1.5. Androgen Receptor Binding (US EPA OPPTS 890.1150)

Status: Assay validated at national level.

Modality detected/endpoints: Binding to androgen receptor.

Background to the assay

237. The AR Binding Assay is an *in vitro* screening assay to detect substances that bind to androgen receptors (AR). The assay has been in use for a number of years and there are different variations of the protocol. The most commonly used protocol utilises rat prostate cytosol as a source of AR without further purification. Human AR is also available as a recombinant proteind. The AR Binding Assay was chosen to be one of the suite of assays comprising the United States Environmental Protection Agency's (US EPA) "Tier 1" and has been validated in that context (US EPA, 2007). There is no OECD test guideline for the assay, but the US EPA (OPPTS) guideline is available (published in October 2009) (US EPA, 2009). In this context, the assay provides information on the ability of a compound to interact with AR but is not intended to be used to show that the interaction is, specifically, one-site competitive binding, or to characterise precisely the strength of the binding. The assay determines the ability of a chemical to displace a radiolabeled ligand (R1881) from AR (in a rat ventral prostate tissue homogenate) and provides a positive or negative result for the ability to bind to AR.

238. Chemicals that bind to AR may induce hormone-dependent transcriptional activity (agonist) or block normal hormone function by preventing the endogenous hormone from binding to the receptor (antagonist). The binding assay does not distinguish between these. The AR ligand binding domain among vertebrate species is well conserved, so that substances that bind to AR derived from one species are expected to bind to the AR from other vertebrate species. The results from this assay are therefore relevant to many taxa. A positive result in guideline OPPTS 890.1150 requires demonstration of a concentration response curve for the ability of the test chemical to displace radiolabelled R1881. The concentration response curve allows the determination of potency, i.e. IC50 (concentration at which 50% of radioligand is displaced by the test chemical) and relative binding affinity by comparing the log (IC50) of R1881with that of the test chemical.

239. Performance criteria are specified for the assay in order to demonstrate that the assay is functioning correctly. Proficiency chemicals are also used on each run to demonstrate the sensitivity of the experiment (reference standard: R1881 and weak positive control: dexamethasone). Compliance with the performance criteria should be checked before evaluating results from this assay. Small deviations are unlikely to have compromised the assay, but judgement should be made on a case-by-case basis.

When/why the assay may be used

240. Although the AR Binding Assay may be used at any stage in the hazard assessment process, the most likely use scenario is during initial assessment of chemicals for their ability to interact with endocrine systems *in vitro*, i.e. estrogen/androgen/thyroid/ steroidogenesis (E,A,T,S) modalities. Assays for interaction with other modalities (e.g. estrogen receptor [ER] and steroidogenesis interference), are likely to be conducted at the same time so that all results can be considered together. Thyroid hormone receptor (TR)

and other assays concerning mechanisms of thyroid disruption may be available, but they are not in common use. The AR Binding Assay does not include the use of a xenobiotic metabolising system, but consideration should be given to the inclusion of this (Jacobs et al., 2008; OECD, 2008) depending on the circumstances (e.g. if the metabolism of a chemical is unknown), although the methods for inclusion of xenobiotic metabolising systems are not yet validated (see Paragraph 50). Alternatively, for a chemical with known metabolites, these could also be tested in the AR Binding Assay. Another use scenario may be following effects obtained in higher tier tests, for example delayed or accelerated puberty onset in males, which could be indicative of an effect mediated by AR. Selection of the most appropriate tests has to be on a case-by-case basis, but also considering the need to minimise animal testing. A further example could be results obtained in other apical assays, e.g. OECD TG 408 (90-day toxicity test), where effects on reproductive organs may be investigated further by testing the AR Binding Assay in combination with ER- and steroidogenesis-based assays.

Introduction to the table of scenarios

241. Table C.1.5 gives guidance on a further step to take in the event of a positive (+) or negative (-) result in the AR Binding Assay and in the presence of positive (+), negative (-) or equivocal/absent (Eq/0) existing results. "Existing results" are subdivided into "mechanism" and "effects" data (third and fourth columns). The table is divided horizontally into a series of scenarios that represent all the combinations of these events.

242. The results of the AR Binding Assay are given in the second column. Criteria for positive, negative and equivocal results are given in the OPPTS guideline. A result is judged positive if the lowest point on the fitted response curve, within the range of data, is less than 50%. This means that more than 50% of radiolabeled R1881 has been displaced from the receptor and a log IC50 can be obtained. A positive result should be obtained in at least two out of three independent test runs. Chemicals with limited solubility may be problematic in this assay if some binding is seen at high concentrations. The maximum concentration of chemical to be used in the assay is 1mM. The guideline provides detailed guidance on classification of a chemical as "binder", "equivocal", "non-binder" or "untestable" (does not reach 50% reduction in binding and is not soluble above 10^{-6} M). It is important that quality and proficiency criteria are demonstrated for both positive and negative results.

243. Equivocal results for the guideline are not included in the table because these data generally require further interrogation about the result itself. In the event of an equivocal result, the considerations mentioned above about control quality and proficiency criteria should be taken into account and further investigations made. Equivocal results at high concentrations may indicate solubility issues.

Existing data to be considered

244. Existing "mechanism" *in vitro* data are assumed to be available from ER-based assays (Level 2) and the Steroidogenesis Assay. Assays may also be available for interference with thyroid modalities. In practice, it is possible that data from some or all of these assays may not be available, so judgement will need to be used to decide which assays to perform. The AR Binding Assay and AR transactivation assays both provide data about the intrinsic ability of a chemical to interact with AR, but the binding assay will not distinguish between agonists and antagonists whilst some chemicals testing positive in the

transactivation assays may have affected the reporter gene activity through non-AR related mechanisms. Consistent results in both assays give more confidence about the presence or absence of an AR-related mode of action (MOA).

245. Existing "effects" data refer to *in vivo* effects "of concern" (i.e. data from Level 4 or 5 tests). These may come from varied sources and will depend on the type of substance (e.g. new chemicals, high production volume [HPV] chemical, pesticide). Thus, available data may range from repeated dose toxicity studies (28-day, 90-day) or combined repeat dose/reproductive screening assays to chronic toxicity studies and multigeneration reproductive tests. Some studies fail to identify endocrine disrupters (EDs) that weakly affect estrogen or androgen receptors as was demonstrated on the basis of data generated in the validation process of the OECD TG 407 assay with endocrine endpoints. In this validation, only moderate EDs such as nonylphenol and DDE, and strong EDs such as ethinylestradiol and flutamide (acting via ER and AR respectively) were detected. Thus OECD TG 407 cannot be regarded as a screening assay for endocrine activity. This means that when a relatively insensitive test is positive for both endocrine-specific and apical endpoints, this should be taken as an indication that the substance is a potential ED. Caution should be exercised, however, because endocrine endpoints may be impacted secondary to non-endocrine toxicity and *in vivo* apical endpoints can be affected by many MOA, including endocrine modalities. Data may also be available on effects in mammalian and non-mammalian wildlife species, although caution should be used when extrapolating between taxa. A chemical causing endocrine effects in non-mammalian vertebrates may also have endocrine effects in mammals, but the physiological consequences of the effects are likely to be different.

246. Data may also be available from Level 3 tests (Hershberger [H] and Uterotrophic [UT] Assays), but as the H assay primarily detects (*in vivo*) the same modality as AR binding, it is unlikely that it would be conducted before AR binding. An Amphibian Metamorphosis Assay (AMA) may also be available, but as this test primarily detects thyroid disruption in amphibians it is unlikely to provide useful data for A-modalities.

247. When considering the results of the AR Binding Assay, all available data should be used in order to reach a conclusion and a weight of evidence approach taken. This may include high throughput screening (HTS) data, read-across data from structural analogues and quantitative structure activity relationship (QSAR). Several QSAR models for ER and AR binding/activation are now available (see Sections B.1.1.1 and B.1.1.2).

Scenarios: Positive and negative results combined with existing data

248. The scenarios (A to R) presented in Table C.1.5 represent all the possibilities of positive or negative results in combination with the presence or absence of existing data. The action taken will also depend on the regulatory environment, but the considerations given here are generally science based. Although this assay uses rat AR, the well-conserved nature of AR across taxa should be a strong indication that results in this assay are relevant to other vertebrate species. Wherever possible, the recommended "next step which could be taken" avoids unnecessary animal testing. However, sometimes conducting an animal test will be indicated and then the relevance of species, strain, exposure route and species-specific metabolism should always be considered. Further considerations specific to each scenario are given in the table.

249. Scenarios A to C represent positive results in the AR Binding Assay in the presence of positive *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. A positive result in an AR Binding Assay is strong evidence for (anti)androgenic activity that

may or may not be supported by the *in vivo* effects data. In the case of positive *in vivo* effects data, there may be sufficient evidence to conclude concern for endocrine disruption and therefore no need for further screening. *In vivo* assays/tests with negative results should be interpreted with caution as they may either indicate that the tests used do not have sufficient power to detect weak effects or, alternatively, that the effects do not present a concern for endocrine disruption. Generally, a conclusion of lack of concern for endocrine disruption in the presence of positive effects data may only be made given adequate Level 5 assays. Information on some endocrine-related tumours may be detected more comprehensively in carcinogenicity studies (OECD TG 451/453) (Level 4); for example, detection of certain types of thyroid tumors in the absence of reproductive or developmental effects, as well as substances causing tumors in other endocrine-sensitive tissues. MOA data to provide a clear interpretation may be required by some regulatory agencies.

250. Scenarios D to F represent positive results in the AR Binding Assay in the presence of negative *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. Unless the metabolic profile of the test substance is known, one option may be to conduct these *in vitro* assays with an added metabolising system. If the metabolic profile is known, then an *in vivo* test may be advisable. The choice of tests will depend on the available *in vivo* effects data. As in Scenarios A to C, *in vivo* assays/tests with negative results should be interpreted with caution as they may either indicate that the tests used do not have sufficient power to detect weak effects or, alternatively, that the effects do not present a concern for endocrine disruption. Generally, a conclusion of lack of concern for endocrine disruption in the presence of positive effects data (Scenario E) may only be made given adequate Level 5 assays. Information on some endocrine-related tumours may be detected more comprehensively in carcinogenicity studies (OECD TG 451/453) (Level 4); for example, detection of certain types of thyroid tumors in the absence of reproductive or developmental effects, as well as substances causing tumors in other endocrine-sensitive tissues. MOA data to provide a clear interpretation may be required by some regulatory agencies.

251. Scenarios G to I represent positive results in the AR Binding Assay in the presence of various combinations of missing or equivocal data. The next step to take in these eventualities will depend on the nature of the other available data and the jurisdiction in which it is being used. In some cases, equivocal data may be viewed as positive whilst in others it may or may not contribute to the weight of evidence. The interpretation may also depend on the MOA in question and why the data are considered equivocal, e.g. a study that is equivocal for thyroid effects may still be of value in evaluating (anti)androgenic effects. In all three scenarios, the recommended first step is to obtain reliable mechanistic (*in vitro*) data rather than proceed directly to *in vivo* testing. Equivocal and missing data are alternative scenarios and two possibilities for the next step are given in most cases, but the nature of equivocal data means that decisions need to be taken on a case-by-case basis. In all cases, the role of metabolism, route of exposure and data from structural analogues should be considered before deciding on the next step. As above, generally a conclusion of lack of concern for endocrine disruption in the presence of some positive effects data (Scenario H) may only be made given adequate Level 5 assays. Information on some endocrine-related tumours may be detected more comprehensively in carcinogenicity studies (OECD TG 451/453) (Level 4); for example, detection of certain types of thyroid tumors in the absence of reproductive or developmental effects, as well as substances causing tumors in other endocrine-sensitive tissues. MOA data to provide a clear interpretation may be required by some regulatory agencies.

252. Scenarios J to L represent negative results in the AR Binding Assay in the presence of positive *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. The limitations of the AR Binding Assay should be considered first (e.g. lack of metabolic activation, possible involvement of other binding proteins). The positive *in vitro* mechanistic data indicate possible alternative E,T,S mechanisms. To confirm lack of AR-related activity in the presence of *in vivo* data, a Stably Transfected Human Androgen Receptor Transactivation Assay for detection of androgenic (ant)agonist-activity of chemicals (AR STTA) could be performed. Otherwise *in vivo* tests will confirm or refute E,A,T,S activity.

253. Scenarios M to O represent negative results in the AR Binding Assay in the presence of negative *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. Negative results for all tests (Scenario N) may be sufficient to enable a conclusion of no concern for endocrine disruption. This will depend on the weight of evidence and may not be possible in some cases. However, in the presence of negative data from robust Level 4 and 5 assays, further animal testing is probably not justified. The limitations of the AR Binding Assay should also be considered (as described for Scenarios J to L). To confirm lack of AR-related activity in the presence of *in vivo* data, an AR STTA could be performed. Otherwise, *in vivo* tests will confirm or refute E,A,T,S activity (Scenarios M and O).

254. Scenarios P to R represent negative results in the AR Binding Assay in the presence of various combinations of missing or equivocal data. The limitations of the AR Binding Assay should be considered first (as described for Scenarios J to L). As with the positive result scenarios above (see Paragraph 171), the next step to take for Scenarios P to R will have to be decided on a case-by-case basis. However, the recommended first step is generally to obtain reliable mechanistic (*in vitro*) data rather than proceed directly to *in vivo* testing. In all cases, the role of metabolism, route of exposure and data from structural analogues should be considered before deciding on the next step.

255. In all scenarios (A to R), the next step to take to strengthen weight of evidence will depend on the existing information. Table C.1.5 is meant to provide a succinct guide and may not cover all circumstances or possibilities. The scenarios may also suggest that chemicals have simple or single MOA, when in practice they may have multiple endocrine and non-endocrine MOA. In some cases, for example, two opposite modes of simultaneous action (e.g. estrogenic and anti-estrogenic) could, depending on dose, lead to a minimisation or abolition of effects, while in others two different MOA (e.g. estrogenic and anti-androgenic) could potentially reinforce effects. Endocrine pathways interact, mixed effects are common and there are many pathways that cannot be distinguished with currently available test guidelines (TGs). If multiple MOA are suspected, either from the existing results or based on QSAR/read-across/integrated approaches, this should be investigated further if needed for regulatory decision making.

256. In general, a decision about whether or not to conduct *in vivo* vertebrate wildlife tests will depend on the weight of evidence of new and existing data. If most available data (e.g. the results of the AR Binding Assay, results from an AR transcription activation assay, predictions from QSARs, "read-across" from data on similar substances and results from mammalian *in vivo* assays) suggest that the substance has the potential to cause endocrine disruption via the androgen receptor (i.e. the level of suspicion about endocrine disrupting action is high – corresponding to Scenario A), then consideration should be given to the conduct of a higher level test.

257. For non-mammalian wildlife species, higher level tests with fish or amphibians (i.e TG 234 [FSDT], TG 240, TG 241) are recommended. Choice about which of these tests is most appropriate will be driven *inter alia* by MOA considerations, and by whether multigeneration effects are to be expected. Such tests are unlikely to be needed if exposure of the natural environment is not expected. On the other hand, if available data only raise a low or moderate level of suspicion about endocrine disrupting action (e.g. the data appear to conflict with each other), then consideration should be given to the conduct of a fish or amphibian screen (e.g. JMASA, EASZY, XETA, OECD TG 231, TG 229 or TG 230 or the AFSS). There are fewer options available for invertebrates, but if ecdysteroid or juvenile hormone activity are suspected in arthropods (e.g. from a screening test with SJHASA), various higher level tests are available, including OECD GD 201, the DMGT and TG 233.

258. For mammals, similar considerations apply, but lower level tests (e.g. Level 3 or 4) should be conducted before higher level tests in order to avoid unnecessary animal usage, unless it is apparent that a Level 5 test will be required anyway or will be needed to establish the evidence to conclude on ED properties. At Level 5, the Extended One-Generation Reproductive Toxicity Study (EOGRTS – OECD TG 443) is the most sensitive reproduction assay for detecting endocrine disruption because it includes evaluation of a number of endocrine endpoints not included in the two-generation study (OECD TG 416) adopted in 2001. It is recognised, however, that some jurisdictions may require a two-generation study.

References

Jacobs, M.N. et al. (2008), "The use of metabolising systems for *in vitro* testing of endocrine disrupters", *Current Drug Metabolism*, Vol. 9/8, pp. 796-826.

OECD (2008), *Detailed Review Paper on the Use of Metabolising Systems for* In Vitro *Testing of Endocrine Disrupters*, OECD Series on Testing and Assessment, No. 97, OECD Publishing, Paris, https://doi.org/10.1787/9789264085497-en.

US EPA (2009), "Endocrine Disrupter Screening Program Test Guidelines OPPTS 890.1150: Androgen receptor binding (rat prostate cytosol)", Environmental Protection Agency, Washington, DC, https://www.regulations.gov/document?D=EPA-HQ-OPPT-2009-0576-0003.

US EPA (2007), "Integrated summary report for the validation of an androgen receptor binding assay as a potential screen in the Endocrine Disrupter Screening Program", Environmental Protection Agency, Washington, DC.

Table C.1.5. **Androgen Receptor Binding Assay (US EPA OPPTS 890.1150): Guidance for scenarios of combinations of results with existing data**

This table represents possible conclusions to be drawn from assay data, and a next step which could be taken if further evidence is required about possible endocrine disrupting properties and/or effects. The guidance offered is not meant to be prescriptive, but provides science-based considerations. It encourages the use of all available data and expert judgement in a weight of evidence approach. Regional and national interpretation of results and "next steps" may vary.

The conclusions are grouped into a series of scenarios (A-R), each scenario representing a different combination of assay results, existing *in vitro* data and existing *in vivo* data. The symbol "+" indicates that the data in question represent a positive result, «-" indicates a negative result, and "Eq/0" indicates that the data are either equivocal or are not available.

Existing results: * "Mechanism (*in vitro* mechanistic data)" assumes that mechanistic data are available from estrogen receptor (ER-) based assays and the Steroidogenesis Assay (Level 2). Thyroid hormone receptor (TR) and other assays concerning mechanisms of thyroid disruption may be available, but they are not in common use. In practice, data from all assays may not be available and therefore this must be taken into account when deciding on the "next step". Data from the Stably Transfected Human Andogen Receptor Transactivation Assay for detection of androgenic (ant)agonist-activity of chemicals (AR STTA) are assumed to be unavailable, but a decision about the next step to be taken will depend on the availability of this assay. Quantitative structure activity relationship (QSAR) predictions of androgen and estrogen binding/activation may be made for some substances.

Existing results: ** "Effects (*in vivo* effects of concern)" assumes various information, such as data from repeat dose oral toxicity studies, reproduction/developmental toxicity screen tests, read-across from analogues, will be available.

Scenarios	Result of AR Binding Assay	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
A	+	+	+	Interaction with androgen receptor(AR) combined with effects on ER/T/S and potential for adverse effects via multiple mechanisms.	Perform assay AR STTA or Assay from Levels 3-4, e.g. Hershberger (H) assay (Level 3) or fish screen (AFSS or JMASA) (Level 3) or male Perpubertal (PP) assay (Level 4) or EOGRTS or two-generation assays or partial/full non-mammalian wildlife life cycle tests, e.g. OECD TG 241 and TG 240 (Level 4/5).	Binding to mammalian AR indicates strong probability of binding to AR in other taxa. If existing data are from Level 5, there may be sufficient information to conclude evidence of concern for endocrine disruption (the EOGRTS provides the most information; however, for endocrine disrupting chemicals [EDCs] with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from Level 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism, although some Level 4 assays (e.g. TG 234 [Fish Sexual Development Test – FSDT]) may be sufficient for this purpose. If existing data are from an H assay or AFSS, then Level 4 mammalian assays or fish screens (OECD TG 229/230/234) will provide data on multiple modalities. Consider route of exposures for existing effects data and possible transformation products and implications of absorption, distribution, metabolism and excretion (ADME) characteristics of the chemical. The chemical itself may give positive *in vitro* results but may not be absorbed or may be metabolised to an inactive metabolite *in vivo*. However, note that uptake and metabolism of chemicals can be different between mammalian and non-mammalian wildlife species.
B	+	+	–	Interaction with AR combined with effects on AR/T/S but effects not detected in *in vivo* studies. Weak interaction with AR does not result in adverse effects in the selected species under the conditions of the test. Metabolic differences may explain *in vitro*/*in vivo* differences.	Perform binding assay or AR STTA with added metabolising system or Assay from Levels 3-4, e.g. H assay or fish screen (OECD TG 229/230 or AFSS or JMASA) (Level 3) or male PP assay (Level 4).	Binding to mammalian AR indicates strong probability of binding to AR in other taxa. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from Level 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism, although some Level 4 assays (e.g. TG 234 [FSDT]) may be sufficient for this purpose. If existing data are from an H assay or AFSS or JMASA, then Level 4 mammalian assays or fish screens (OECD TG 229/230/234) will provide data on multiple modalities. Consider route of exposures for existing effects data and possible transformation products and implications of ADME characteristics of the chemical. The chemical itself may give positive *in vitro* results but may not be absorbed or may be metabolised to an inactive metabolite *in vivo*. However, note that uptake and metabolism of chemicals can be different between mammalian and non-mammalian wildlife species.

Scenarios	Result of AR Binding Assay	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
C	+	+	Eq/0	Interaction with AR combined with effects on ER/T/S but no or equivocal data from *in vivo* studies. Weak interaction with AR may not result in adverse effects in the selected species under the conditions of the test.	Perform assay from Levels 3-4, e.g H assay or fish screen (OECD TG 229/230/234 or AFSS or JMASA) (Level 3) or male PP assay (Level 4).	Binding to mammalian AR indicates strong probability of binding to AR in other taxa. Consider route of exposures for equivocal existing effects data and possible implications of ADME characteristics of the chemical. Equivocal results may indicate chemical has multiple modes of action (MOA). Check data on chemical analogues.
D	+	–	+	Interaction with AR and potential for adverse effects.	Perform AR STTA or Perform assay from Levels 3-4, e.g. H assay or fish screen (OECD TG 229/230/234 or AFSS) (Level 3) or male PP assay (Level 4).	Binding to mammalian AR indicates strong probability of binding to AR in other taxa. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude evidence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from Level 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism, although some Level 4 assays (e.g. TG 234 [FSDT]) may be sufficient for this purpose. If existing data are from H assay or AFSS or JMASA, then Level 4 mammalian assays or fish screens (OECD TG 229/230) will provide data on multiple modalities. A positive result could have arisen from other (E,A,T,S or non-E,A,T,S) mechanisms, e.g. hypothalamic/pituitary/gonadal (HPG) axis.
E	+	–	–	Interaction with AR but effects not detected in *in vivo* studies. Weak interaction with AR does not result in adverse effects in the selected species under the conditions of the test. Metabolic differences may explain *in vitro*/*in vivo* differences.	Perform binding assay or AR STTA with added metabolising system or Assay from Levels 3-4, e.g H assay or fish screen (OECD TG 229/230/234 or AFSS or JMASA) (Level 3) or male PP assay (Level 4).	Binding to mammalian AR indicates strong probability of binding to AR in other taxa. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from Level 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism, although some Level 4 assays (e.g. TG 234 [FSDT]) may be sufficient for this purpose. If existing data are from an H assay or AFSS or JMASA, then Level 4 mammalian assays or fish screens (OECD TG 229/230/234) will provide data on multiple modalities. Consider route of exposures for existing effects data and possible transformation products and implications of ADME characteristics of the chemical. Equivocal results may occur if chemical has multiple MOA. The chemical itself may give positive *in vitro* results but may not be absorbed or may be metabolised to an inactive metabolite *in vivo*. However, note that uptake and metabolism of chemicals can be different between non-mammalian wildlife species. Check data on chemical analogues.

Scenarios	Result of AR Binding Assay	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
F	+	–	Eq/0	Interaction with AR but no or equivocal data from *in vivo* studies. Weak interaction with AR does not result in adverse effects in the selected species under the conditions of the test.	Perform AR STTA or perform assay from Levels 3-4, e.g. H assay or fish screen (OECD TG 229/230/234 or AFSS or JMASA) (Level 3) or male PP assay (Level 4).	Binding to mammalian AR indicates strong probability of binding to AR in other taxa. AR transactivation assay results will indicate whether AR binding affects transcription. Consider route of exposures for equivocal existing effects data and possible implications of ADME characteristics of the chemical. Check data on chemical analogues. Equivocal results may indicate chemical has multiple MOA.
G	+	Eq/0	+	Interaction with AR and potential for adverse effects via AR or other E,T,S mechanisms. May act via E,A,T,S mechanisms and may or may not require metabolic activation.	Perform AR STTA.	Binding to mammalian AR indicates strong probability of binding to AR in other taxa. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude evidence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from Level 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism, although some Level 4 assays (e.g. TG 234 [FSDT]) may be sufficient for this purpose. If existing data are from an H assay or AFSS or JMASA, then Level 4 mammalian assays or fish screens (OECD TG 229/230/234) will provide data on multiple modalities. Check data on chemical analogues. Further mechanistic studies may help determine MOA. Equivocal results may occur if chemical has multiple MOA A positive result could have arisen from other (E,A,T,S or non-E,A,T,S) mechanisms, e.g. HPG axis.
H	+	Eq/0	–	Interaction with AR but effects not detected in *in vivo* studies. Weak interaction with AR does not result in adverse effects in the selected species under the conditions of the test. Metabolic differences may explain *in vitro*/*in vivo* differences.	For the "0" scenario, perform AR STTA. For the "Eq" scenario perform AR STTA with added metabolising system.	Binding to mammalian AR indicates strong probability of binding to AR in other taxa. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from Level 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism, although some Level 4 assays (e.g. TG 234 [FSDT]) may be sufficient for this purpose. If existing data are from an H assay or AFSS or JMASA, then Level 4 mammalian assays or fish screens (OECD TG 229/230/234) will provide data on multiple modalities. Consider route of exposures for existing effects data and possible transformation products and implications of ADME characteristics of the chemical. The chemical itself may give positive *in vitro* results but may not be absorbed or may be metabolised to an inactive metabolite *in vivo*. However, note that uptake and metabolism of chemicals can be different between mammalian and non-mammalian wildlife species. Check data on chemical analogues. Further mechanistic studies may help determine MOA. Equivocal results may occur if chemical has multiple MOA.

Scenarios	Result of AR Binding Assay	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
I	+	Eq/0	Eq/0	Interaction with AR with unknown potential for effects in *in vivo* studies. May act via AR and may or may not require metabolic activation. Unknown potential for adverse effects.	For the "0" scenario, AR STTA with added metabolising system. For the "Eq" scenario, H assay or fish screen (OECD TG 229/230/234 or AFSS or JMASA) (Level 3) if existing data indicate this is needed.	Binding to mammalian AR indicates strong probability of binding to AR in other taxa. Consider route of exposures for equivocal existing effects data and possible implications of ADME characteristics of the chemical. The chemical itself may give positive *in vitro* results but may not be absorbed or may be metabolised to an inactive metabolite *in vivo*. Check data on chemical analogues. Equivocal results may indicate chemical has multiple MOA.
J	–	+	+	No evidence for interaction with AR. Effects on ER/T/S and potential for adverse effects via E,A,T,S mechanisms.	Perform AR Binding Assay or AR transactivation assay with added metabolising system or Perform assay from Levels 3-4, e.g. Uterotrophic (UT) Assay or fish screen OECD TG 229/230/234 or TG 231) (Level 3) or male PP assay (Level 4).	Lack of binding to mammalian AR indicates binding to AR in other taxa is unlikely. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude concern for endocrine disruption. If existing data are from Level 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism, although some Level 4 assays (e.g. TG 234 [FSDT]) may be sufficient for this purpose. If existing data are from an H Assay or AFSS or JMASA, then a Level 4 mammalian assay or fish screen (OECD TG 229/230/234) will provide data on multiple modalities. Consider route of exposures for existing effects data and possible transformation products and implications of ADME characteristics of the chemical. Check data on chemical analogues. Further mechanistic studies may help determine MOA.
K	–	+	–	No evidence for interaction with AR. Effects on ER/T/S but effects not detected in *in vivo* studies. Metabolic differences explain *in vitro*/*in vivo* A,T,S differences.	Perform assay from Levels 3-4, e.g. fish screen (OECD TG 229/230/234) (Level 3) or male or female PP assay (Level 4).	Lack of binding to mammalian AR indicates binding to AR in other taxa is unlikely. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from a Level 4 vertebrate wildlife assay, then a Level 5 assay should provide more data on adverse effects on endocrine relevant endpoints over more extensive parts of the life cycle of the organism, although some Level 4 assays (e.g. TG 234 [FSDT]) may be sufficient for this purpose. If existing data are from an H assay or AFSS or JMASA, then a Level 4 mammalian assay or fish screen (OECD TG 229/230/234) will provide data on multiple modalities. Metabolic deactivation of chemical may occur *in vivo* so that potential *in vitro* E,T,S activity is not realised. Consider possible routes of exposure, implications of metabolism.

Scenarios	Result of AR Binding Assay	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
L	–	+	Eq/0	No evidence for interaction with AR. Effects on ER/T/S but effects not detected in *in vivo* studies. Unknown potential for adverse effects. Metabolic differences explain *in vitro*/*in vivo* E,T,S differences.	Perform assay from Levels 3-4, e.g. fish screen (OECD TG 229/230/234) (Level 3), or male or female PP (Level 4).	Lack of binding to mammalian AR indicates binding to AR in other taxa is unlikely. Metabolic deactivation of chemical may occur *in vivo* so that potential *in vitro* E,T,S activity is not realised. Consider possible routes of exposure, implications of metabolism. Check data on chemical analogues. Equivocal results may occur if chemical has multiple MOA.
M	–	–	+	No evidence for interaction with AR. Metabolic differences or route of exposure explain *in vitro*/*in vivo* differences. Effects seen in existing studies are via non-E,A,T,S or non-endocrine mechanisms.	Perform AR STTA with added metabolising system or Perform assay from Levels 3-4, e.g. fish screen (OECD TG 229/230/234) (Level 3) or male or female PP assay (Level 4).	Lack of binding to mammalian AR indicates binding to AR in other taxa is unlikely. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude concern for endocrine disruption. If existing data are from Level 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism, although some Level 4 assays (e.g. TG 234 [FSDT]) may be sufficient for this purpose. If existing data are from an H assay or AFSS or JMASA, then Level 4 mammalian assays or fish screens (OECD TG 229/230/234) will provide data on multiple modalities. Metabolic activation of chemical may occur *in vivo*. Consider possible routes of exposure, implications of metabolism. Check data on chemical analogues. Further mechanistic studies may help determine MOA.
N	–	–	–	No evidence for interaction with AR. No evidence of adverse effects.	Possibly no need for further testing. If there is uncertainty, may perform assay from Levels 3-4, e.g. fish screen (AFSS or JMASA) (Level 3), or male or female PP assay (Level 4).	Lack of binding to mammalian AR indicates binding to AR in other taxa is unlikely. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from Level 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism, although some Level 4 assays (e.g. TG 234 [FSDT]) may be sufficient for this purpose. If existing data are from an H assay or AFSS or JMASA, then Level 4 mammalian assays or fish screens (OECD TG 229/230/234) will provide data on multiple modalities. Check data on chemical analogues.

Scenarios	Result of AR Binding Assay	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
O	–	–	Eq/0	No evidence for interaction with AR. Metabolic differences may explain *in vitro*/*in vivo* differences. Unknown potential for adverse effects via other mechanisms.	Perform AR STTA with added metabolising system or Fish screen (OECD TG 229/230/234) (Level 3) or male or female PP assay (Level 4) if existing data indicate this is needed.	Lack of binding to mammalian AR indicates binding to AR in other taxa is unlikely. Consider possible routes of exposure, implications of metabolism. Check data on chemical analogues.
P	–	Eq/0	+	No evidence for interaction with AR. Metabolic differences may explain *in vitro*/*in vivo* differences. Unknown potential for adverse effects via other mechanisms.	Perform AR STTA with added metabolising system.	Lack of binding to mammalian AR indicates binding to AR in other taxa is unlikely. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude concern for endocrine disruption. If existing data are from Level 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism, although some Level 4 assays (e.g. TG 234 [FSDT]) may be sufficient for this purpose. If existing data are from an H assay or AFSS, then Level 4 mammalian assays or fish screens (OECD TG 229/230/234) will provide data on multiple modalities. Consider possible routes of exposure, implications of metabolism. Check data on chemical analogues. Further mechanistic studies may help determine MOA. Equivocal results may occur if chemical has multiple MOA.
Q	–	Eq/0	–	No evidence for interaction with AR. No evidence of adverse effects.	Perform AR STTA with added metabolising system.	Lack of binding to mammalian AR indicates binding to AR in other taxa is unlikely. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from Level 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism, although some Level 4 assays (e.g. TG 234 [FSDT]) may be sufficient for this purpose. If existing data are from an H assay or AFSS, then Level 4 mammalian assays or fish screens (OECD TG 229/230) will provide data on multiple modalities. Check data on chemical analogues. Further mechanistic studies may help determine MOA.
R	–	Eq/0	Eq/0	No evidence for interaction with AR. Unknown potential for adverse effects via other mechanisms.	For the "0" scenario, perform AR STTA with added metabolising system or Perform assay from Levels 3-4, e.g. H assay or fish screen (OECD TG 229/230/234) (Level 3) or male PP assay (Level 4).	Lack of binding to mammalian AR indicates binding to AR in other taxa is unlikely. Consider possible routes of exposure, implications of metabolism. Check data on chemical analogues. Further mechanistic studies may help determine MOA. Equivocal results may occur if chemical has multiple MOA.

C.1.6. Aromatase Assay (US EPA OPPTS 890.1200)

Status: Assay validated at national level.

Modality detected/endpoints: Inhibition of aromatase (CYP19) enzyme activity.

Background to the assay

259. The Aromatase Assay is an *in vitro* screening assay to detect substances that inhibit aromatase – the cytochrome P450 enzyme complex (CYP 19) responsible for the conversion of androgens to estrogens during steroidogenesis. Inhibition of aromatase enzyme activity alters the levels of circulating estrogens in males and females, which may lead to effects on reproductive organs and other targets such as the mammary gland. Aromatase is found in many vertebrate taxa, including mammals and fish, and therefore the results of this assay are applicable to both human health and mammalian and non-mammalian wildlife populations (US EPA, 2007).

260. The assay determines the conversion of radiolabeled [1-^{3}H]-androstenedione to estrone. The progress of the reaction can be followed by measuring the formation of either of the reaction products: estrone or water. The most common assay in usage determines the formation of tritiated water as the end product of the reaction (US EPA OPPTS 890.1200, published in October 2009; US EPA, 2009). This assay was chosen to be one of the suite of assays comprising United States Environmental Protection Agency's (US EPA) Endocrine Disruptor Screening Program "Tier 1" and has been validated in that context. Aromatase enzyme may be obtained from a number of sources (e.g. human placenta or rat ovary), but human recombinant aromatase has recently become available and this is the preferred source as it is directly relevant to humans, is easily obtained and does not require the use of laboratory animals. Guideline OPPTS 890.1200 utilises the human recombinant enzyme.

261. Inhibition of aromatase may also be determined in the H295R Steroidogenesis Assay. This assay detects substances that affect production of estradiol and testosterone, but the Steroidogenesis Assay contains all the enzymes involved in steroidogenesis, from cholesterol to estradiol and testosterone. Aromatase is the final enzyme in this pathway. Chemicals causing aromatase inhibition will be detected in the Steroidogenesis Assay by causing reduced production of estradiol from the H295R cells, but as the assay is not specific for aromatase it would not be possible to discern which enzyme(s) activity is altered. The H295R Steroidogenesis Assay, as an intact cell system, will also detect chemicals that induce aromatase enzyme activity whilst the aromatase assay itself is not capable of detecting inducers.

262. The aromatase assay may be subject to variability, for example due to degradation of the enzyme, and therefore performance criteria are specified in guideline OPPTS 890.1200 in order to demonstrate that the assay is functioning correctly. An adequate response with the proficiency chemicals econazole, fenarimol, nitrofen (inhibitors) and atrazine (non-inhibitor) should be demonstrated and the inhibitor

4-hydroxyandrostenedione is used as a positive control chemical in each experiment. Compliance with the performance criteria should be checked before evaluating results from this assay. A positive result in guideline OPPTS 890.1200 requires demonstration of inhibition of aromatase activity that fits a four-parameter non-linear regression model and such that the concentration response curve crosses 50% inhibition. The concentration response curve allows the determination of potency, i.e. IC50 (concentration at which the activity of aromatase is reduced to 50% of control values). In some cases, variability may be due to limited solubility of a chemical. The maximum concentration of chemical to be used in the assay is 1mM.

When/why the assay may be used

263. Although the aromatase assay may be used at any stage in the hazard assessment process, the most likely use scenario is during initial assessment of chemicals for their ability to interact with endocrine systems *in vitro* (i.e. estrogen/androgen/thyroid/ steroidogenesis [E,A,T,S] modalities). Assays for interaction with other modalities (e.g. androgen receptor [AR], estrogen receptor [ER] and the Steroidogenesis Assay) are likely to be conducted at the same time, so that all results can be considered together. Thyroid hormone receptor (TR) and other assays concerning mechanisms of thyroid disruption may be available, but they are not in common use. If the *in vitro* assays are not conducted at the same time then positive results in the Steroidogenesis Assay could be followed by an aromatase assay to confirm and clarify a mode of action (MOA). The aromatase assay does not include the use of a xenobiotic metabolising system, but consideration should be given to the inclusion of this (Jacobs et al., 2008, 2013; OECD, 2008) depending on the circumstances (e.g. if the metabolism of a chemical is unknown), although the methods for inclusion of xenobiotic metabolising systems are not yet validated (see Paragraph 50). Alternatively, for a chemical with known metabolites, these could also be tested in the aromatase assay. Another use scenario may be following effects obtained in higher tier tests, for example delayed puberty onset in females, but which are not exclusively indicative of an effect on ER. Selection of the most appropriate tests has to be on a case-by-case basis, but also considering the need to minimise animal testing. A further example could be results obtained in other apical assays, e.g. OECD TG 408 (90-day toxicity test), where effects on reproductive organs may be investigated further by testing in the Aromatase and Steroidogenesis Assays in combination with AR- and ER-based assays.

Introduction to the table of scenarios

264. Table C.1.6 gives guidance on a further step to take in the event of a positive (+) or negative (-) result in the aromatase assay and in the presence of positive (+), negative (-) or equivocal/absent (Eq/0) existing results. “Existing results” are subdivided into “mechanism” and “effects” data (third and fourth columns). The table is divided horizontally into a series of scenarios that represent all the combinations of these events.

265. The results of the aromatase assay are given in the second column. Criteria for positive, negative and equivocal results are given in guideline OPPTS 890.1200. A result is judged positive if the average concentration response curve crosses 50% of control activity (“inhibitor”). A negative result is obtained if the average lowest portion of concentration response curve is greater than 75% of control activity or data do not fit the regression model (“non-inhibitor”). “Equivocal” results lie between these limits. It is

important that quality and proficiency criteria are demonstrated for both positive and negative results.

266. Equivocal results for the guideline are not included in the table because these data require further interrogation about the result itself. This assay is a screen and therefore a clear positive or negative result should be obtained. In the event of an equivocal result, the considerations mentioned above about control quality and proficiency criteria should be taken into account and further investigations made. Equivocal results at high concentrations may indicate solubility issues.

Existing data to be considered

267. Existing "mechanism" *in vitro* data are assumed to be available from ER- and AR-based and Steroidogenesis Assays (Level 2). Assays may also be available for interference with thyroid modalities. In practice, it is possible that data from all of these assays may not be available, so judgement will need to be used to decide which assays to perform.

268. Existing "effects" data refer to *in vivo* effects "of concern" (i.e. data from Level 4 or 5 vertebrate wildlife assays). These may come from varied sources and will depend on the type of substance (e.g. new chemicals, high production volume chemical, pesticide). Thus, available data may range from repeated dose toxicity studies (28-day, 90-day), combined repeat dose/reproductive screening assays or fish screening assays, to chronic toxicity studies and multigeneration reproductive tests in vertebrate wildlife species. Some studies fail to identify endocrine disrupters (EDs) that weakly affect estrogen or androgen receptors as was demonstrated on the basis of data generated in the validation process of the OECD TG 407 assay with endocrine endpoints. In this validation, only moderate EDs such as nonylphenol and DDE, and strong EDs such as ethinylestradiol and flutamide (acting via ER and AR respectively) were detected. The aromatase inhibitor CGS 18320B was detected by the OECD TG 407 assay, although this chemical was developed as a pharmaceutical aromatase inhibitor and therefore is a strong ED, but the ability to detect chemicals that weakly inhibit with aromatase is not known. Thus, OECD TG 407 cannot be regarded as a screening assay for endocrine activity. This means that when a relatively insensitive test is positive for both endocrine-specific and apical endpoints, this should be taken as an indication that the substance is a potential ED. Caution should be exercised, however, because endocrine endpoints may be impacted secondary to non-endocrine toxicity and *in vivo* apical endpoints can be affected by many MOA, including endocrine modalities. Data may also be available on effects in mammalian and non-mammalian wildlife species, although caution should be used when extrapolating between taxa. A chemical causing endocrine effects in non-mammalian vertebrates may also have endocrine effects in mammals, but the physiological consequences of the effects are likely to be different.

269. Data may also be available from Hershberger (H) and Uterotrophic (UT) Assays (Level 3), but as these assays do not generally detect aromatase interference they are only useful in these cases for purposes of elimination.

270. When considering the results of the aromatase assay, all available data should be used in order to reach a conclusion and a weight of evidence approach taken. This may include hgh throughput screening (HTS) data, read-across data from structural analogues and quantitative structure activity relationship (QSAR). Several QSAR models for ER and AR binding/activation are now available (see Sections B.1.1.1 and B.1.1.2).

Scenarios: Positive and negative results combined with existing data

271. The scenarios (A to R) presented in Table C.1.6 represent all the possibilities of positive or negative results in combination with the presence or absence of existing data. The action taken will also depend on the regulatory environment, but the considerations given here are generally science based. Although mammalian aromatase is used, the enzyme is well conserved across taxa and therefore results in this assay are likely to be relevant to other vertebrate species. Wherever possible, the recommended "next step which could be taken" avoids unnecessary animal testing. However, sometimes conducting an animal test will be indicated and then the relevance of species, strain, exposure route and species-specific metabolism should always be considered. Further considerations specific to each scenario are given in the table.

272. Scenarios A to C represent positive results for aromatase inhibition in the presence of positive *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. A positive result is strong evidence for inhibition of aromatase that may or may not be supported by the *in vivo* effects data. In the case of positive *in vivo* effects data, there may be sufficient evidence to conclude concern for endocrine disruption and therefore no need for further screening. *In vivo* assays/tests with negative results should be interpreted with caution as they may either indicate that the tests used do not have sufficient power to detect weak effects or, alternatively, that the effects do not present a concern for endocrine disruption. Generally, a conclusion of lack of concern for endocrine disruption in the presence of positive effects data may only be made given adequate Level 5 assays. Information on some endocrine-related tumours may be detected more comprehensively in carcinogenicity studies (OECD TG 451/453) (Level 4); for example, detection of certain types of thyroid tumors in the absence of reproductive or developmental effects, as well as substances causing tumors in other endocrine-sensitive tissues. MOA data to provide a clear interpretation may be required by some regulatory agencies.

273. Scenarios D to F represent positive results for aromatase inhibition in the presence of negative *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. Unless the metabolic profile of the test substance is known, one option may be to conduct these *in vitro* assays with an added metabolising system. If the metabolic profile is known, then an *in vivo* test may be advisable. The choice of tests will depend on the available *in vivo* effects data. As in Scenarios A to C, *in vivo* assays/tests with negative results should be interpreted with caution as they may either indicate that the tests used do not have sufficient power to detect weak effects or, alternatively, that the effects do not present a concern for endocrine disruption. Generally, a conclusion of lack of concern for endocrine disruption in the presence of positive effects data (Scenario E) may only be made given adequate Level 5 assays. Information on some endocrine-related tumours may be detected more comprehensively in carcinogenicity studies (OECD TG 451/453) (Level 4); for example, detection of certain types of thyroid tumors in the absence of reproductive or developmental effects, as well as substances causing tumors in other endocrine-sensitive tissues. MOA data to provide a clear interpretation may be required by some regulatory agencies.

274. Scenarios G to I represent positive results for aromatase inhibition in the presence of various combinations of missing or equivocal data. The next step to take in these eventualities will depend on the nature of the other available data and the jurisdiction in which it is being used. In some cases, equivocal data may be viewed as positive whilst in others it may or may not contribute to the weight of evidence. The interpretation may also depend on the MOA in question and why the data are considered equivocal, e.g. a study

that is equivocal for thyroid effects may still be of value in evaluating (anti)androgenic effects. In all three scenarios, the recommended first step is to obtain reliable mechanistic (*in vitro*) data rather than proceed directly to *in vivo* testing. Equivocal and missing data are alternative scenarios and two possibilities for the next step are given in most cases, but the nature of equivocal data means that decisions need to be taken on a case-by-case basis. In all cases, the role of metabolism, route of exposure and data from structural analogues should be considered before deciding on the next step. As above, generally a conclusion of lack of concern for endocrine disruption in the presence of some positive effects data (Scenario H) may only be made given adequate Level 5 assays. Information on some endocrine-related tumours may be detected more comprehensively in carcinogenicity studies (OECD TG 451/453) (Level 4); for example, detection of certain types of thyroid tumors in the absence of reproductive or developmental effects, as well as substances causing tumors in other endocrine-sensitive tissues. MOA data to provide a clear interpretation may be required by some regulatory agencies.

275. Scenarios J to L represent negative results for aromatase inhibition in the presence of positive *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. The limitations of the aromatase assay should be considered first (e.g. lack of metabolic activation, possible involvement of factors). The positive *in vitro* mechanistic data indicates possible alternative estrogen/androgen/thyroid (EAT) mechanisms. To confirm lack of aromatase activity in the presence of *in vivo* data, an aromatase assay with added metabolising capability could be performed. Otherwise *in vivo* tests will confirm or refute E,A,T,S activity.

276. Scenarios M to O represent negative results for aromatase inhibition in the presence of negative *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. Negative results for all tests (Scenario N) may be sufficient to enable a conclusion of no concern for endocrine disruption. This will depend on the weight of evidence and may not be possible in some cases. However, in the presence of negative data from robust Level 4 and 5 assays, further animal testing is probably not justified. The limitations of the aromatase assay should also be considered (as described for Scenarios J to L). To confirm lack of aromatase inhibition in the presence of *in vivo* data, an aromatase assay with added metabolising capability could be performed. Otherwise, *in vivo* tests will confirm or refute E,A,T,S activity (Scenarios M and O).

277. The limitations of the aromatase assay should be considered first (as described for Scenarios J to L). As with the positive result scenarios above, the next step to take for Scenarios P to R when negative results in the aromatase assay are obtained in the presence of various combinations of missing or equivocal data will have to be decided on a case-by-case basis. However, the recommended first step is generally to obtain reliable mechanistic (*in vitro*) data rather than proceed directly to *in vivo* testing. In all cases, the role of metabolism, route of exposure and data from structural analogues should be considered before deciding on the next step.

278. In all scenarios (A to R), the next step to take to strengthen weight of evidence will depend on the existing information. Table C.1.6 is meant to provide a succinct guide and may not cover all circumstances or possibilities. The scenarios may also suggest that chemicals have simple or single MOA, when in practice they may have multiple endocrine and non-endocrine MOA. In some cases, for example, two opposite modes of simultaneous action (e.g. estrogenic and anti-estrogenic) could, depending on dose, lead to a minimisation or abolition of effects, while in others two different MOA (e.g. estrogenic and anti-androgenic) could potentially reinforce effects. Endocrine pathways interact and there are

many for which no test guidelines (TGs) yet exist. If multiple MOA are suspected, either from the existing results or based on QSAR/read-across/integrated approaches, this should be investigated further if needed for regulatory decision making.

279. In general, a decision about whether or not to conduct *in vivo* vertebrate wildlife tests will depend on the weight of evidence of new and existing data. If most available data (e.g. the results of the Steroidogenesis Assay, predictions from QSARs, "read-across" from data on similar substances and results from mammalian *in vivo* assays) suggest that the substance has the potential to cause endocrine disruption via inhibition of aromatase (i.e. the level of suspicion about endocrine disrupting action is high – corresponding to Scenario A), then consideration should be given to the conduct of a higher level test.

280. For non-mammalian wildlife species, higher level tests with fish or amphibians (i.e TG 234 [Fish Sexual Development Test], TG 240, TG 241) are recommended. Choice about which of these tests is most appropriate will be driven *inter alia* by mode of action considerations, and by whether multigeneration effects are to be expected. Such tests are unlikely to be needed if exposure of the natural environment is not expected. On the other hand, if available data only raise a low or moderate level of suspicion about endocrine disrupting action (e.g. the data appear to conflict with each other), then consideration should be given to the conduct of a fish or amphibian screen (i.e. OECD TG 229 or TG 230). There are fewer options available for invertebrates, but if ecdysteroid or juvenile hormone activity are suspected in arthropods (e.g. from a screening test with SJHASA), various higher level tests are available, including OECD GD 201, the DMGT and TG 233.

281. For mammals, similar considerations apply but lower level tests (e.g. Level 3 or 4) should be conducted before higher level tests in order to avoid unnecessary animal usage, unless it is apparent that a Level 5 test will be required anyway or will be needed to establish the evidence to conclude on ED properties. At Level 5, the Extended One-Generation Reproductive Toxicity Study (OECD TG 443) is the most sensitive reproduction assay for detecting endocrine disruption because it includes evaluation of a number of endocrine endpoints not included in the two-generation study (OECD TG 416) adopted in 2001. It is recognised, however, that some jurisdictions may require a two-generation study.

References

Jacobs, M. et al. (2013), "*In vitro* metabolism and bioavailability tests for endocrine active substances: What is needed next for regulatory purposes?", *ALTEX – Alternatives to Animal Experimentation*, Vol. 30/3, pp. 331-351.

Jacobs, M.N. et al. (2008), "The use of metabolising systems for *in vitro* testing of endocrine disrupters", *Current Drug Metabolism*, Vol. 9/8, pp. 796-826.

OECD (2008), *Detailed Review Paper on the Use of Metabolising Systems for* In Vitro *Testing of Endocrine Disrupters*, OECD Series on Testing and Assessment, No. 97, OECD Publishing, Paris, https://doi.org/10.1787/9789264085497-en.

US EPA (2009), "Endocrine Disrupter Screening Program Test Guidelines OPPTS 890.1200: Aromatase (human recombinant)", Environmental Protection Agency, Washington, DC, https://www.regulations.gov/document?D=EPA-HQ-OPPT-2009-0576-0004.

US EPA (2007), "Integrated summary report on aromatase", Environmental Protection Agency, Washington, DC.

Table C.1.6. **Aromatase Assay (US EPA OPPTS 890.1200): Guidance for scenarios of combinations of results with existing data**

This table represents possible conclusions to be drawn from assay data, and a next step which could be taken if further evidence is required about possible endocrine disrupting properties and/or effects. The guidance offered is not meant to be prescriptive, but provides science-based considerations. It encourages the use of all available data and expert judgement in a weight of evidence approach. Regional and national interpretation of results and "next steps" may vary.

The conclusions are grouped into a series of scenarios (A-R), each scenario representing a different combination of assay results, existing *in vitro* data and existing *in vivo* data. The symbol "+" indicates that the data in question represent a positive result, "-" indicates a negative result, and "Eq/0" indicates that the data are either equivocal or are not available.

Existing results: * "Mechanism (*in vitro* mechanistic data)" assumes that mechanistic data are available from endocrine receptor (ER-) and androgen receptor (AR-) based assays (Level 2). It is assumed that data from the Steroidogenesis Assay are also available. Thyroid hormone receptor (TR) and other assays concerning mechanisms of thyroid disruption may be available, but they are not in common use. In practice, data from all assays may not be available and therefore this must be taken into account when deciding on the "next step". Quantitative structure activity relationship (QSAR) predictions of estrogen and androgen binding/activation may be made for some substances.

Existing results: ** "Effects (*in vivo* effects of concern)" assumes various information, such as data from repeat dose oral toxicity studies, reproduction/developmental toxicity screen tests, read-across from analogues, will be available.

Scenarios	Result of aromatase assay	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
A	+	+	+	Inhibition of aromatase combined with effects on ER/AR/T/S and potential for adverse effects via multiple mechanisms.	Perform assay from Levels 3-5, e.g. male or female pubertal assay (Level 4) or EOGRTS or two-generation assays or partial/full non-mammalian wildlife life cycle tests, e.g. TG 241 and TG 240 (Level 4/5).	A positive result indicates strong probability of aromatase inhibition in other taxa. If existing data are from adequateLevel 5 assay, there may be sufficient information to conclude evidence of concern for endocrine disruption (the EOGRTS provides the most information; however, for endocrine disrupting chemicals [EDCs] with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from Level 3 or 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism, although some Level 4 assays (e.g. TG 234 [Fish Sexual Development Test]) may be sufficient for this purpose. Compare aromatase assay results with other *in vitro* results to help discern mechanism. Consider route of exposures for existing effects data and possible transformation products and implications of absorption, distribution, metabolism and excretion (ADME) characteristics of the chemical. The chemical itself may give positive *in vitro* results and *in vivo* results but may also be metabolised to a metabolite that also has positive results *in vitro* and *in vivo*. However, note that uptake and metabolism of chemicals can be different between mammalian and non-mammalian wildlife species.
B	+	+	–	Inhibition of aromatase combined with effects on ER/AR/T/S but effects not detected in *in vivo* studies. Weak aromatase inhibition does not result in adverse effects in the selected species under the conditions of the test. Metabolic differences may explain *in vitro*/*in vivo* differences.	Perform aromatase assay with added metabolising system or Assay from Levels 3-4, e.g. fish screen (OECD TG 229/230/234) (Level 3) or male or female pubertal assay (Level 4).	A positive result indicates strong probability of aromatase inhibition in other taxa. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from Level 3 or 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism, although some Level 4 assays (e.g. TG 234 [FSDT]) may be sufficient for this purpose. Compare aromatase assay results with other *in vitro* results to help discern mechanism. Consider route of exposures for existing effects data and possible transformation products and implications of ADME characteristics of the chemical. The chemical itself may give positive *in vitro* results but may not be absorbed or may be metabolised to an inactive metabolite *in vivo*. However, note that uptake and metabolism of chemicals can be different between mammalian and non-mammalian wildlife species.

Scenarios	Result of aromatase assay	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
C	+	+	Eq/0	Inhibition of aromatase combined with effects on ER/AR/T but no or equivocal data from *in vivo* studies. Weak aromatase inhibition may not result in adverse effects in the selected species under the conditions of the test.	Perform assays from Levels 3-4, e.g. fish screen (OECD TG 229/230/234) (Level 3) or male or female pubertal assay (Level 4).	A positive result indicates strong probability of aromatase inhibition in other taxa. Compare aromatase assay results with other *in vitro* results to help discern mechanism. Consider route of exposures for equivocal existing effects data and possible implications of ADME characteristics of the chemical. Equivocal results may occur if chemical has multiple modes of action (MOA). Check data on chemical analogues.
D	+	–	+	Inhibition of aromatase and potential for adverse effects.	Perform assay from Levels 3-4, e.g. fish screen (OECD TG 229/230/234) (Level 3) male or female pubertal assay or (Level 4).	A positive result indicates strong probability of aromatase inhibition in other taxa. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude evidence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from Level 3 or 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism, although some Level 4 assays (e.g. TG 234 [FSDT]) may be sufficient for this purpose. Further mechanistic studies may help determine MOA. A positive result could have arisen from other (E,A,T,S or non-E,A,T,S) mechanisms, e.g. hypothalamic/pituitary/gonadal (HPG) axis.
E	+	–	–	Inhibition of aromatase but effects not detected in *in vivo* studies. Weak aromatase inhibition does not result in adverse effects in the selected species under the conditions of the test. Metabolic differences may explain *in vitro*/*in vivo* differences.	Perform aromatase assay with added metabolising system or Assay from Levels 3-4, e.g. fish screen (OECD TG 229/230/234) (Level 3) or male or female pubertal assay (Level 4).	A positive result indicates strong probability of aromatase inhibition in other taxa. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from Level 3 or 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism, although some Level 4 assays (e.g. TG 234 [FSDT]) may be sufficient for this purpose. Consider route of exposures for existing effects data and possible transformation products and implications of ADME characteristics of the chemical. The chemical itself may give positive *in vitro* results but may not be absorbed or may be metabolised to an inactive metabolite *in vivo*. However, note that uptake and metabolism of chemicals can be different between mammalian and non-mammalian wildlife species. Check data on chemical analogues.

Scenarios	Result of aromatase assay	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
F	+	–	Eq/0	Inhibition of aromatase but no or equivocal data from *in vivo* studies. Weak aromatase inhibition may not result in adverse effects in the selected species under the conditions of the test.	Perform assays from Levels 3-4, e.g. fish screen (OECD TG 229/230/234) (Level 3) or male or female pubertal assay (Level 4).	A positive result indicates strong probability of aromatase inhibition in other taxa. Consider route of exposures for equivocal existing effects data and possible implications of ADME characteristics of the chemical. Equivocal results may indicate chemical has multiple MOA. Check data on chemical analogues.
G	+	Eq/0	+	Inhibition of aromatase and potential for adverse effects via aromatase inhibition or other E,A,T,S mechanisms. May act via non-aromatase inhibition mechanism and may or may not require metabolic activation.	Perform assay from Levels 3-4, e.g. fish screen (OECD TG 229/230/234) (Level 3) or male or female pubertal assay (Level 4).	A positive result indicates strong probability of aromatase inhibition in other taxa. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude evidence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from Level 3 or 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism, although some Level 4 assays (e.g. TG 234 [FSDT]) may be sufficient for this purpose. Check data on chemical analogues. Further mechanistic studies may help determine MOA. Equivocal results may occur if chemical has multiple MOA. A positive result could have arisen from other (E,A,T,S or non-E,A,T,S) mechanisms (e.g. HPG axis).
H	+	Eq/0	–	Inhibition of aromatase but effects not detected in *in vivo* studies. Weak aromatase inhibition does not result in adverse effects in the selected species under the conditions of the test. Metabolic differences may explain *in vitro*/*in vivo* differences.	Perform aromatase assay with added metabolising system.	A positive result indicates strong probability of aromatase inhibition in other taxa. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from Level 3 or 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism, although some Level 4 assays (e.g. TG 234 [FSDT]) may be sufficient for this purpose. Consider route of exposures for existing effects data and possible transformation products and implications of ADME characteristics of the chemical. The chemical itself may give positive *in vitro* results but may not be absorbed or may be metabolised to an inactive metabolite *in vivo*. However, note that uptake and metabolism of chemicals can be different between non-mammalian wildlife species. Check data on chemical analogues. Further mechanistic studies may help determine MOA.

Scenarios	Result of aromatase assay	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
I	+	Eq/0	Eq/0	Inhibition of aromatase with unknown potential for effects in *in vivo* studies. May act via non-aromatase inhibition mechanism and may or may not require metabolic activation. Unknown potential for adverse effects.	Perform aromatase assay with added metabolising system, or assay from Levels 3-4, e.g. fish screen (OECD TG 229/230/234) (Level 3) or male or female pubertal assay (Level 4) if existing data indicate this is needed.	A positive result indicates strong probability of aromatase inhibition in other taxa. Consider route of exposures for equivocal existing effects data and possible implications of ADME characteristics of the chemical. Check data on chemical analogues. Equivocal results may occur if chemical has multiple MOA.
J	–	+	+	No evidence for aromatase inhibition. Effects on ER/AR/T/S and potential for adverse effects via EAT mechanisms.	Perform aromatase assay with added metabolising system or Perform assay from Levels 3-4, e.g. fish screen (OECD TG 229/230/234) (Level 3) or male or female pubertal assay (Level 4).	A negative result indicates that interference with aromatase inhibition in other taxa is unlikely. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude concern for endocrine disruption. If existing data are from Level 3 or 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism, although some Level 4 assays (e.g. TG 234 [FSDT]) may be sufficient for this purpose. Consider route of exposures for existing effects data and possible transformation products and implications of ADME characteristics of the chemical. Check data on chemical analogues. Further mechanistic studies may help determine MOA.
K	–	+	–	No evidence for aromatase inhibition. Effects on ER/AR/T/S but effects not detected in *in vivo* studies. Metabolic differences explain *in vitro*/*in vivo* E,A,T,S differences.	Perform assay from Levels 3-4, e.g. fish screen (OECD TG 229/230/234) (Level 3) or male or female pubertal assay (Level 4).	A negative result indicates that interference with aromatase inhibition in other taxa is unlikely. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from Level 3 or 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism, although some Level 4 assays (e.g. TG 234 [FSDT]) may be sufficient for this purpose. Metabolic deactivation of chemical may occur *in vivo* so that potential *in vitro* E,A,T,S activity is not realised. Consider possible routes of exposure, implications of metabolism.
L	–	+	Eq/0	No evidence for aromatase inhibition. Effects on ER/AR/T/S but effects not detected in *in vivo* studies. Unknown potential for adverse effects. Metabolic differences explain *in vitro*/*in vivo* EAT differences.	Perform assay from Levels 3-4, e.g. fish screen (OECD TG 229/230/234) (Level 3) or male or female pubertal assay (Level 4).	A negative result indicates that interference with aromatase inhibition in other taxa is unlikely. Metabolic deactivation of chemical may occur *in vivo* so that potential *in vitro* E,A,T,S activity is not realised. Consider possible routes of exposure, implications of metabolism. Check data on chemical analogues. Equivocal results may indicate chemical has multiple MOA.

Scenarios	Result of aromatase assay	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
M	–	–	+	No evidence for aromatase inhibition. Metabolic differences or route of exposure explain *in vitro*/*in vivo* differences. Effects seen in existing studies are via non-E,A,T,S or non-endocrine mechanisms.	Perform aromatase assay with added metabolising system or perform assay from Levels 3-4, e.g. fish screen (OECD TG 229/230/234) (Level 3) or male or female pubertal assay (Level 4).	A negative result indicates that interference with aromatase inhibition in other taxa is unlikely. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude concern for endocrine disruption. If existing data are from Level 3 or 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism, although some Level 4 assays (e.g. TG 234 [FSDT]) may be sufficient for this purpose. Metabolic activation of chemical may occur *in vivo*. Consider possible routes of exposure, implications of metabolism. Check data on chemical analogues. Further mechanistic studies may help determine MOA.
N	–	–	–	No evidence for aromatase inhibition. No evidence of adverse effects.	Possibly no need for further testing. If there is uncertainty, may perform assay from Levels 3-4, e.g. fish screen (OECD TG 229/230/234) (Level 3) or male or female pubertal assay (Level 4).	A negative result indicates that interference with aromatase inhibition in other taxa is unlikely. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from Level 3 or 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism, although some Level 4 assays (e.g. TG 234 [FSDT]) may be sufficient for this purpose. Check data on chemical analogues.
O	–	–	Eq/0	No evidence for aromatase inhibition. Unknown potential for adverse effects via other mechanisms.	Perform aromatase assay with added metabolising system or assay from Levels 3-4, e.g. fish screen (OECD TG 229/230/234) (Level 3) or male or female pubertal assay (Level 4) if existing data indicate this is needed.	A negative result indicates that interference with aromatase inhibition in other taxa is unlikely. Consider possible routes of exposure, implications of metabolism. Check data on chemical analogues.
P	–	Eq/0	+	No evidence for aromatase inhibition. Unknown potential for adverse effects via other mechanisms.	Perform aromatase assay with added metabolising system.	A negative result indicates that interference with aromatase inhibition in other taxa is unlikely. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude concern for endocrine disruption, although some Level 4 assays (e.g. TG 234 [FSDT]) may be sufficient for this purpose. Consider possible routes of exposure, implications of metabolism. Check data on chemical analogues. Further mechanistic studies may help determine MOA. Equivocal results may occur if chemical has multiple MOA.

Scenarios	Result of aromatase assay	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
Q	–	Eq/0	–	No evidence for aromatase inhibition. No evidence of adverse effects.	Perform Steroidogenesis Assay with added metabolising system.	A negative result indicates that interference with aromatase inhibition in other taxa is unlikely. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from Level 3 or 4 vertebrate wildlife assays, then Level 5 assays should provide more comprehensive data on adverse effects on endocrine-relevant endpoints over more extensive parts of the life cycle of the organism, although some Level 4 assays (e.g. TG 234 [FSDT]) may be sufficient for this purpose. Check data on chemical analogues. Further mechanistic studies may help determine MOA.
R	–	Eq/0	Eq/0	No evidence for aromatase inhibition. Unknown potential for adverse effects via other mechanisms.	For the "0" scenario, perform Steroidogenesis Assay with added metabolising system or Perform assay from Levels 3-4, e.g. fish screen (OECD TG 229/230/234) (Level 3) or male or female pubertal (Level 4) if existing data indicate this is needed.	A negative result indicates that interference with aromatase inhibition in other taxa is unlikely. Consider possible routes of exposure, implications of metabolism. Check data on chemical analogues. Further mechanistic studies may help determine MOA. Equivocal results may occur if chemical has multiple MOA.

OECD non-mammalian screens and tests (Conceptual Framework Levels 3-5)

C.2.1. Fish Short-Term Reproduction Assay (FSTRA) (OECD TG 229)

Status: Assay validated by the OECD.

282. Modality detected/endpoints: estrogens (♂ VTG ↑; ♂ 2º sex characteristics ↓); anti-estrogens (♀ VTG ↓); androgens (♂ 2º sex characteristics in ♀); anti-androgens (♂ 2º sex characteristics ↓); aromatisable androgens (♂ VTG ↑); aromatase inhibitors (♀ VTG ↓); non-specific effects on hypothalamic/pituitary/gonadal (HPG) axis, plus other reprotox (fecundity ↓); (optional endpoint – gonadal histo-pathology. This may assist with diagnosis of mode of action). Note that this assay may, in some cases, have low statistical power or sensitivity to detect anti-androgenic activity through effects on secondary sexual characteristics. However, if gonad histopathology has been optionally studied, changes in Leydig cells resulting from anti-androgen exposure may have been observed. Finally, diagnostic endpoints (i.e. indicators of hormonal activity) and the apical endpoint (i.e. fecundity) should be considered together to obtain maximum value from this assay.

Background to the assay

283. This assay is primarily designed as a screen for the types of *in vivo* endocrine disruption activity in fish which are listed above, but it also provides information on adverse effects on fecundity which could be used in characterising the hazards of an individual chemical based on a predicted environmental concentration/predicted no-effect concentration approach (although note that only three test concentrations are normally used, so precision of a no-observed-effect-concentration/x% effect concentration (NOEC/ECx) may be relatively low). The fecundity endpoint, which although not necessarily diagnostic of endocrine action, does indicate that apical effects on reproduction are occurring, is sensitive to known endocrine disrupters (EDs). However, the validation studies demonstrated high variability for fecundity (and consequently low power to detect an effect) under certain suboptimal test conditions (e.g. for some fish strains, the recommended degree of replication may provide low power). If the assay gives a positive result, this may be due to a positive indicator of hormonal activity (vitellogenin level, secondary sexual characteristic development), which may or may not be associated with decrease in fecundity. Each of these three possible combinations of positive response should be considered separately (although the distinctions between indicators of hormonal activity and apical effects are not always clear), so they have been listed individually as points 1, 2 and 3 in the possible conclusions column of Table C.2.1. Although this test guideline (TG) is primarily a screening assay where a combination of positive data on hormonal activity and fecundity could lead to a conclusion that higher level testing is desirable (depending on the overall weight of evidence), some regulatory authorities may consider that such a combination is sufficient evidence on its own of endocrine disruption providing an effect on fecundity that is sufficiently large enough to constitute a plausible threat to a fish population. It should be noted, in addition, that due to the relatively short exposure time employed in this screen (three weeks), effects of some chemicals on fecundity might not be as apparent as in longer term exposures, especially for bioaccumulative chemicals. Also, as only three test concentrations are employed, even a reliable short-term NOEC or ECx for fecundity cannot be precisely derived.

When/why the assay may be used

284. Although OECD TG 229 could, in principle, be used at any stage in the hazard assessment process, the most likely use scenario will be when there are relatively few data available about the possible endocrine disrupting properties of a chemical. The results from this assay are most likely to be available after deployment of a battery of *in vitro* and *in vivo* screens (e.g. the United States Environmental Protection Agency's Endocrine Disrupter Screening Program), or as a supplement to existing data which suggest possible endocrine disruption activity. It is also possible that no existing endocrine-relevant data are available (i.e. OECD TG 229 has been used as a primary screen), but in that case a positive result in the screen should ideally be followed up with relevant *in vitro* screening to investigate the suspected mode of action (MOA). Given the high degree of endocrine system conservation across the vertebrates, adverse endocrine-linked effects in the FSTRA may also indicate the possibility of related activity in other organisms such as amphibians, reptiles, birds or mammals.

285. In order to provide information relevant for assessing whether or not a chemical may fulfil the WHO/IPCS (2002) definition of an ED, the study design has to be sufficiently robust to demonstrate the presence or absence of effects. In the dose selection, the investigator should also consider and ensure that data generated are adequate to fulfil the regulatory requirement across OECD countries as appropriate (e.g. hazard and risk assessment and labelling, ED assessment, etc.). The top dose or concentration should be sufficiently high to give clear systemic (i.e. non endocrine-specific) toxicity in order to ensure that a wide range of exposures (high to low) is tested. However, endocrine effects observed solely in the presence of clear systemic toxicity should be interpreted with caution and may be disregarded when sufficiently justified to be caused by secondary effects which are unlikely to be due to endocrine activity. The reason for this advice is a concern that some endocrine active substance (EAS) sensitive assays are being run at doses/concentrations of EASs that are too low to trigger direct impacts on the endocrine system. This guidance document (GD) is not the place to address this issue directly, but it should be considered when EAS-sensitive TGs are revised in the future. In addition, the number and spacing of dose/concentration levels should also be adequate to fulfil the objectives of the study (e.g. to demonstrate dose response relationships if this is required).

Existing data to be considered

286. Given the commonality of endocrine mechanisms in the vertebrates, relevant existing data available before deployment of OECD TG 229 might include *in vivo* results obtained with other vertebrates (e.g. a positive Uterotrophic Assay with rodents, positive findings for endocrine endpoints in mammalian repeat dose toxicity or reproductive studies), or one or more of a range of *in silico* or *in vitro* results which suggest that the modalities indicated above may occur *in vivo*. Such indicators of possible *in vivo* activity might include (quantitative) structure activity relationship QSAR predictions of endocrine activity, high throughput screening (HTS) data, "read-across" from *in vivo* results obtained with structurally related chemicals, or positive results from an *in vitro* screen for estrogen or androgen receptor-mediated activity, or for effects on steroidogenesis (especially aromatase inhibition). Further strong indication of *in vivo* estrogenic activity may also be available from an EASZY Assay with transgenic zebrafish embryos. OECD TG 229 may itself also be used as part of a battery of screening assays. Conduct of OECD TG 229 would be particularly relevant if knowledge is sought about the test chemical's effects on the mature reproductive phase of the fish life cycle (as opposed to effects on the immature sexual development phase), because it provides some apical information on reproductive success and gonad

histopathology. However, this assay is also likely to be responsive to many chemicals which act primarily on sexual development.

Scenarios: Positive and negative results combined with existing data

287. The scenarios (A to R) presented in Table C.2.1 represent all the possibilities of positive or negative results in combination with the presence or absence of existing data. The action taken will also depend on the regulatory environment, but the considerations given here are generally science based. Wherever possible, the recommended "next step which could be taken" avoids unnecessary animal testing. However, sometimes conducting an animal test will be indicated and then the relevance of species, strain and exposure route should always be considered. Further considerations specific to each scenario are given in the table.

288. Positive results obtained with one or more of the indicators of hormonal activity (Scenarios A-I, sub-section 2) result in the conclusion that the test chemical is a potential ED *in vivo*. If both an indicator of hormonal activity and fecundity give a response (Scenarios A-I, sub-section 1), this provides strong evidence for *in vivo* endocrine activity on the HPG axis with potential adverse effects, and some regulatory authorities may consider that this is sufficient evidence of ED. If only fecundity responds (Scenarios A-I, sub-section 3), it suggests that the chemical is a reproductive or general systemic toxicant, with a reduced probability that it is an ED **that acts on one or more of the endocrine modalities covered in the Conceptual Framework** (although existing positive *in vitro* data, or positive *in vivo* data from other species, would have to be weighed against this conclusion).

289. As indicated above, although a combined effect on fecundity and an indicator of hormonal activity in OECD TG 229 suggests that the test chemical is a reproductive toxicant acting through one or more estrogen/androgen/thyroid/steroidogenesis (E,A,T,S) pathways (assuming that the concentration giving this response is not sufficiently high to cause systemic toxicity), a result of this type may need to be followed up with a more comprehensive reproduction test if countries need further evidence (e.g. a Medaka Extended One-Generation Reproduction Test [MEOGRT], OECD TG 240; or Zebrafish Extended One-Generation Reproduction Test [ZEOGRT]) which is able to provide a more reliable and reproducible NOEC or ECx for adverse effects. An exception might be if there are no indications of endocrine activity (either from this or other screens/tests), although in such a case, an NOEC or ECx for reproductive effects would still need to be derived for a non-endocrine hazard identification/characterisation (e.g. using data from OECD TG 210). Equally, if one or more biomarkers for hormonal activity alone respond **without a corresponding response from apical endpoints**, this would also need to be followed up with more comprehensive testing to show whether any adverse apical effects occur at other parts of the life cycle, if countries need further evidence whether the chemical is an ED. In other words, in order to strengthen weight of evidence (WOE) in relation to ED, a positive result of whichever type in OECD TG 229 could be followed by fish partial or full life cycle testing at Level 4 or 5. Existing data suggesting endocrine-specific activity (e.g. positive *in vitro* data, or positive *in vivo* data from other species) will strengthen the case for additional testing still further if the intention is to establish a firmer link between endocrine activity and adverse effects.

290. The situation in which OECD TG 229 gives a negative result (Table C.3.1, Scenarios J-R) needs careful consideration of the WOE based on any existing data. If these data suggest that the chemical is endocrine active both *in vitro* and *in vivo* (Scenario J), then it is possible that OECD TG 229 is simply insufficiently sensitive, perhaps due to rapid metabolism, or because the main mode of action (MOA) acts more potently during sexual development, or because fish in general are simply insensitive to the chemical under

consideration. In some of these circumstances, it might therefore be appropriate to conduct further studies, e.g. of metabolism and characterisation of metabolites for endocrine disruptive properties of the chemical in the tested fish species, or a Fish Sexual Development Test (FSDT) (OECD TG 234), or alternatively, a MEOGRT or ZEOGRT to confirm that there is no endocrine activity in fish.

291. If OECD TG 229 and existing *in vivo* data are all negative, but *in vitro* data reveal some endocrine activity (Scenario K), the probability is that the test chemical is not sufficiently potent to produce endocrine effects *in vivo* in fish (e.g. because it is rapidly hydrolysed or metabolised to ED-inactive metabolites). In such a situation, further testing in fish is probably not necessary. However, if the chemical is known to bioaccumulate slowly, it may be that exposures in the *in vivo* tests have been insufficiently prolonged, in which case longer term testing might be justified. Equally, if the *in vitro* or histopathology data reveal anti-androgenic or thyroid activity, consideration may be given to conducting the Androgenised Female Stickleback Screen or Juvenile Medaka Anti-androgen Screening Assay (JMASA), or the Amphibian Metamorphosis Assay (OECD TG 231) or *Xenopus* Embryonic Thyroid Signalling Assay (XETA), respectively.

292. On the other hand, if OECD TG 229 and the *in vitro* tests are negative (Scenario M), but there are positive existing *in vivo* data, the nature of those existing data should be considered. Unless the existing data are from fish, there is no evidence that the chemical is an ED acting on fish reproduction, but it may act via MOA not covered by the *in vitro* screens, or it may be more potent in species or life stages that have not been tested. Finally, if it does cause endocrine activity-related effects in the test but no effects on fecundity, this may simply be due to the lack of sensitivity of this screening test which, as mentioned above, has limitations due to relatively high variability of the fecundity parameter in combination with the relatively low number of fish per exposure concentration, etc. In this situation, the existing *in vivo* data should be used to guide decisions about whether to conduct any further testing, either for modalities such as anti-androgenicity or thyroid activity, or including life stages represented in OECD TG 234 (FSDT) or in the MEOGRT (OECD TG 240) or ZEOGRT.

293. Finally, a negative OECD TG 229 screen, set against a background of negative *in vitro* and *in vivo* data (Scenario N), suggests that the test chemical is not an ED acting on reproduction in fish, and no further testing for estrogenic, anti-estrogenic, androgenic or steroidogenic MOA will generally be necessary. It remains possible that it has anti-androgenic or thyroid activity, although **this scenario is unlikely if relevant *in vitro* tests for these modalities have shown negative results and if no effects have been detected** by gonadal histopathology. However, it should be noted that a full suite of *in vitro* thyroid assays is not yet available.

294. In each of the above scenarios, it is possible that existing data will be equivocal, or there may be no existing data (Scenarios C, F-I, L and O-R). This will weaken the conclusions which can be drawn about a negative OECD TG 229 test, and this is reflected in Table C.2.1. However, a lack of mechanistic data on endocrine activity should ideally be rectified before any further *in vivo* testing is finally rejected. Indeed, as a general principle, it is desirable to obtain as many relevant ED-related mechanistic non-test and *in vitro* data as possible before doing any *in vivo* testing. On the other hand, if OECD TG 229 is positive, further *in vivo* testing may be needed to establish a more precise NOEC or ECx for any adverse effects, even if all other existing data are equivocal, or if there are no existing data. Again, however, it will always be desirable to obtain some mechanistic information before conducting further *in vivo* testing. There is also the possibility that equivocal mechanistic

data may be the result of multiple modes of endocrine action. Under some circumstances, two opposite modes of simultaneous action (e.g. estrogenic and anti-estrogenic) could, depending on dose, lead to a minimisation or abolition of adverse effects, while in others two different MOA (e.g. estrogenic and anti-androgenic) could potentially reinforce effects on certain apical endpoints. If multiple MOA are suspected, either from the existing results or based on QSAR/read-across/integrated approaches, this situation should be investigated further if needed for regulatory decision making.

295. The scenario in which the results of OECD TG 229 are themselves equivocal has not been dealt with in Table C.2.1, for reasons of brevity. In this context, an equivocal result might be an inconsistent concentration-response (e.g. no effect at a high concentration but effects at a lower concentration), or a result which borders on statistical significance. Without knowing the exact circumstances, reliable advice cannot be given, but the opinions of an experienced ecotoxicologist should be sought. Clearly, however, such equivocal results do not necessarily rule out the existence of *in vivo* endocrine activity. For example, vitellogenin induction in males at a high concentration might be masked by any systemic toxicity, while fecundity depression might just fail to reach a statistically significant level because the sometimes high variability of this endpoint combined with a relatively small sample size might have reduced the power of the test to detect a difference from the controls. If these or other possible reasons for false negatives are suspected with good reason, the screen could be repeated (e.g. conduct it at lower concentrations which avoid systemic toxicity) or a more appropriate version of it (e.g. more fish per replicate) could be designed and conducted. However, note that a repeat test in the event of systemic toxicity would not be needed providing at least one tested concentration was not subject to such toxicity.

296. In summary, positive results in the OECD TG 229 screen indicate that a chemical is either a reproductive toxicant, or a possible endocrine disrupter, or both. In most cases, more comprehensive *in vivo* testing would be needed if the intention is to derive a long-term NOEC/ECx and/or to confirm whether or not the chemical is an actual endocrine disrupter. In this connection, it should also be borne in mind that effects solely on fecundity might be caused by systemic toxicity rather than endocrine disruption or specific reproductive toxicity, if test concentrations were very high. Negative results in OECD TG 229 do not necessarily mean that the chemical is not a potential ED – a judgement about the endocrine disruption potential and the possible need for additional testing will have to be made in the light of existing *in vitro* and *in vivo* data.

Reference

WHO/IPCS (2002), "Global assessment of the state-of-the-science of endocrine disrupters", Damstra, T. et al. (eds.) WHO/PCS/EDC/02.2, World Health Organization, Geneva, www.who.int/ipcs/publications/new_issues/endocrine_disruptors/en.

Table C.2.1. **Fish Short-Term Reproduction Assay (FSTRA) (OECD TG 229): Guidance for scenarios of combinations of results with existing data**

This table represents possible conclusions to be drawn from assay data, and a next step which could be taken if further evidence is required about possible endocrine disrupting properties and/or effects. The guidance offered is not meant to be prescriptive, but provides science-based considerations. It encourages the use of all available data and expert judgement in a weight of evidence approach. Regional and national interpretation of results and "next steps" may vary.

The conclusions are grouped into a series of scenarios (A-R), each scenario representing a different combination of assay results, existing *in vitro* data and existing *in vivo* data. The symbol "+" indicates that the data in question represent a positive result, «-" indicates a negative result, and "Eq/0" indicates that the data are either equivocal or are not available.

The assay under discussion could either be positive for both apical endpoints and indicators of endocrine activity, or positive just for apical endpoints, or positive just for indicators of endocrine activity. For each scenario, each of these three possibilities is addressed separately in the possible conclusions column, **taking into consideration other existing data**.

Existing results: * "Mechanism (*in vitro* mechanistic data)" assumes that mechanistic data are available from estrogen receptor (ER-), androgen receptor (AR-) and steroidogenesis-based assays (Level 2). Thyroid hormone receptor (TR) and other assays concerning mechanisms of thyroid disruption may be available, but they are not in common use. In practice, data from all assays may not be available and therefore this must be taken into account when deciding on the "next step". Quantitative structure activity relationship (QSAR) predictions of estrogen and androgen binding/activation may be made for some substances. There is no evidence at present that equivalent *in vitro* assays with systems derived from fish offer advantages over their mammalian counterparts.

Existing results: ** "Effects (*in vivo* effects of concern)" assumes effects have been observed in other *in vivo* screens/tests which give rise to concern that the test chemical may be an endocrine disrupter.

Scenarios	Result of OECD TG 229 assay	Existing results		Possible conclusions: 1) Indicators of endocrine activity and apical endpoints positive 2) Indicators of endocrine activity positive and apical endpoints negative 3) Indicators of endocrine activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
A	+	+	+	1) Strong evidence for *in vivo* endocrine activity with potential adverse effects (reproductive toxicity) in fish. 2) Strong evidence for *in vivo* endocrine activity in fish. 3) Evidence for *in vivo* endocrine activity in other species, and strong evidence for reproductive toxicity in fish.	Consider performing a fish life cycle test (MEOGRT – OECD TG 240, or ZEOGRT), especially if the intention is to obtain precise data on a reproductive or developmental no-observed-effect-concentration/x% effect concentration (NOEC/ECx).	An alternative approach would be to deploy the Fish Sexual Development Test (FSDT – OECD TG 234), especially if sexual development is expected to give a response at lower concentrations than reproduction.
B	+	+	–	1) Strong-moderate evidence for *in vivo* endocrine activity with potential adverse effects (reproductive toxicity) in fish. 2) Strong-moderate evidence for *in vivo* endocrine activity in fish. 3) Moderate-weak evidence for *in vivo* endocrine activity, and strong evidence for reproductive toxicity in fish.	Consider performing a fish life cycle test (MEOGRT – OECD TG 240, or ZEOGRT), especially if the intention is to obtain precise data on a reproductive or developmental NOEC/ECx.	An alternative approach would be to deploy the FSDT, especially if sexual development is expected to give a response at lower concentrations than reproduction. If the negative *in vivo* data are from another fish endocrine assay, consider possible reasons for the disparity (e.g. differences in species sensitivity) before conducting a life cycle test.
C	+	+	Eq/0	1) Moderate evidence for *in vivo* endocrine activity with potential adverse effects in fish. 2) Moderate evidence for *in vivo* endocrine activity in fish. 3) Weak evidence for *in vivo* endocrine activity, and strong evidence for reproductive toxicity in fish.	Consider performing a fish life cycle test (MEOGRT – OECD TG 240, or ZEOGRT), especially if the intention is to obtain precise data on a reproductive or developmental NOEC/ECx.	An alternative approach would be to deploy the FSDT, especially if sexual development is expected to give a response at lower concentrations than reproduction. If no existing fish data are available, it may be worth performing an FSDTbefore a life cycle test in order to obtain information on whether sexual development is the most sensitive part of the life cycle. Such information could influence the design of the life cycle test. It should be borne in mind that equivocal data may be due to a variety of causes including experimental error, very weak endocrine activity or multiple modes of action (MOA). If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
D	+	–	+	1) Moderate evidence for *in vivo* endocrine activity with potential adverse effects in fish. 2) Moderate evidence for *in vivo* endocrine activity in fish. 3) Evidence for *in vivo* endocrine activity in other species, and strong evidence for reproductive toxicity in fish.	Consider performing a fish life cycle test (MEOGRT – OECD TG 240, or ZEOGRT), especially if the intention is to obtain precise data on a reproductive or developmental NOEC/ECx.	The negative *in vitro* data suggest that the test chemical may be metabolically activated *in vivo* or may not act via the screened receptor. An alternative approach would be to deploy the FSDT, especially if sexual development is expected to give a response at lower concentrations than reproduction.

Scenarios	Result of OECD TG 229 assay	Existing results		Possible conclusions: 1) Indicators of endocrine activity and apical endpoints positive 2) Indicators of endocrine activity positive and apical endpoints negative 3) Indicators of endocrine activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
E	+	–	–	1) Moderate-strong evidence for *in vivo* endocrine activity with potential adverse effects in fish. 2) Moderate-strong evidence for *in vivo* endocrine activity in fish. 3) Weak-moderate evidence for *in vivo* endocrine activity, and strong evidence for reproductive toxicity in fish.	Consider performing a fish life cycle test (MEOGRT – OECD TG 240, or ZEOGRT), especially if the intention is to obtain precise data on a reproductive or developmental NOEC/ECx.	The negative *in vitro* data suggest that the test chemical may be metabolically activated *in vivo*. An alternative approach would be to deploy the FSDT, especially if sexual development is expected to give a response at lower concentrations than reproduction. If the negative *in vivo* data are from another fish endocrine assay, consider possible reasons for the disparity (e.g. differences in species sensitivity) before conducting a life cycle test.
F	+	–	Eq/0	1) Moderate-strong evidence for *in vivo* endocrine activity with potential adverse effects in fish. 2) Moderate-strong evidence for *in vivo* endocrine activity in fish. 3) Weak evidence for *in vivo* endocrine activity, and strong evidence for reproductive toxicity in fish.	Consider performing a fish life cycle test (MEOGRT – OECD TG 240, or ZEOGRT), especially if the intention is to obtain precise data on a reproductive or developmental NOEC/ECx.	The negative *in vitro* data suggest that the test chemical may be metabolically activated *in vivo*. If no existing fish data are available, it may be worth performing an FSDT before a life cycle test in order to obtain information on whether sexual development is the most sensitive part of the life cycle. Such information could influence the design of the life cycle test. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
G	+	Eq/0	+	1) Strong evidence for *in vivo* endocrine activity with potential adverse effects in fish, but mechanism unconfirmed. 2) Strong evidence for *in vivo* endocrine activity in fish, but mechanism unconfirmed. 3) Moderate evidence for *in vivo* endocrine activity (mechanism unconfirmed), and strong evidence for reproductive toxicity in fish.	Obtain more predictive mechanistic data and then consider performing a fish life cycle test (MEOGRT – OECD TG 240, or ZEOGRT).	An alternative approach would be to deploy the FSDT, especially if sexual development is expected to give a response at lower concentrations than reproduction. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

Scenarios	Result of OECD TG 229 assay	Existing results		Possible conclusions: 1) Indicators of endocrine activity and apical endpoints positive 2) Indicators of endocrine activity positive and apical endpoints negative 3) Indicators of endocrine activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
H	+	Eq/0	–	1) Moderate-strong evidence for *in vivo* endocrine activity with potential adverse effects in fish, but mechanism unconfirmed. 2) Moderate-strong evidence for *in vivo* endocrine activity in fish, but mechanism unconfirmed. 3) Weak-moderate evidence for *in vivo* endocrine activity (mechanism unconfirmed), and strong evidence for reproductive toxicity in fish.	Obtain more predictive mechanistic data and then consider performing a fish life cycle test (MEOGRT – OECD TG 240, or ZEOGRT).	An alternative approach would be to deploy the FSDT, especially if sexual development is expected to give a response at lower concentrations than reproduction. If the negative *in vivo* data are from a fish test (e.g. the 21-day fish assay), consider possible reasons for the disparity (e.g. differences in species sensitivity) before conducting a life cycle test. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
I	+	Eq/0	Eq/0	1) Moderate evidence for *in vivo* endocrine activity with potential adverse effects in fish, but mechanism unconfirmed. 2) Moderate evidence for *in vivo* endocrine activity in fish, but mechanism unconfirmed. 3) Weak evidence for *in vivo* endocrine activity (mechanism unconfirmed), but strong evidence for reproductive toxicity in fish.	Obtain more predictive mechanistic data and then consider performing a fish life cycle test (MEOGRT – OECD TG 240, or ZEOGRT).	If no existing fish data are available, it may be worth performing a FSDTbefore a life cycle test in order to obtain information on whether sexual development is the most sensitive part of the life cycle. Such information could influence the design of the life cycle test. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
J	–	+	+	Based on the existing data, the chemical has endocrine activity *in vivo*. The lack of response in OECD TG 229 suggests that sexually mature fish are not responsive, unless the existing data are from fish.	If existing *in vivo* data are from fish, consider performing an FSDT (OECD TG 234) (unless reproduction is known to be the most sensitive life stage).	

Scenarios	Result of OECD TG 229 assay	Existing results		Possible conclusions: 1) Indicators of endocrine activity and apical endpoints positive 2) Indicators of endocrine activity positive and apical endpoints negative 3) Indicators of endocrine activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
K	–	+	–	There is no evidence that the chemical is an endocrine disrupter (ED) *in vivo*, probably because it is very weakly acting, rapidly metabolised or simply does not reach the target site.	Probably no further action, but see comments in right-hand column.	It is possible that EDs which bioaccumulate slowly may only cause effects *in vivo* after exposure times longer than three weeks. If this is suspected, and depending on which part of the life cycle is suspected of being the most sensitive, consider performing an FSDT, a MEOGRT or a ZEOGRT. It is also possible that the chemical may be an anti-androgen *in vivo* (consider performing the Androgenised Female Stickleback Screen, or Juvenile Medaka Anti-Androgen Screening Assay [JMASA]), or a thyroid-active chemical *in vivo* (consider performing the Amphibian Metamorphosis Assay ([AMA] OECD TG 231, or *Xenopus* Embryo Thyroid Signalling Assay [XETA]).
L	–	+	Eq/0	The chemical may not be an ED *in vivo*, but the confidence in this conclusion is relatively low as there is only one unequivocal *in vivo* test result (a negative).	If the existing *in vivo* data are equivocal and from a fish, consider performing a fish assay (OECD TG 229 or TG 230) with a different species, or consider a longer term test (TG 234 [FSDT] or life cycle (EOGRT or ZEOGRT).	It is also possible that the chemical may be an anti-androgen *in vivo* (consider performing the Androgenised Female Stickleback Screen, or JMASA), or a thyroid-active chemical *in vivo* (consider performing the AMA – OECD TG 231, or XETA). It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
M	–	–	+	The chemical is probably not an ED acting on reproduction in fish. However, it has endocrine activity in another species and may act through MOA not covered by the available *in vitro* assays, or it may be more potent in a species other than that tested, or over a longer exposure period.	If further evidence is required, consider using the existing *in vivo* data to help choose a longer term test with an appropriate species.	It is also possible that the chemical may be an anti-androgen *in vivo* (consider performing the Androgenised Female Stickleback Screen, or JMASA), or a thyroid-active chemical *in vivo* (consider performing the AMA – OECD TG 231, or XETA), although lack of *in vitro* binding affinity with receptors suggests this is unlikely. Use the existing *in vivo* data to guide any further testing.
N	–	–	–	The chemical is probably not an ED acting on reproduction in fish. There is a possibility that the chemical is able to affect sexual development in fish, but the probability of this is low given the apparent absence of estrogenic, androgenic or steroidogenic properties.	No further action with respect to estrogenic, anti-estrogenic, androgenic or steroidogenic MOA.	It is still possible that the chemical may be an anti-androgen *in vivo* (consider performing the Androgenised Female Stickleback Screen, or JMASA), or a thyroid-active chemical *in vivo* (consider performing the AMA – OECD TG 231, or XETA), although lack of *in vitro* binding affinity with receptors suggests this is unlikely.

Scenarios	Result of OECD TG 229 assay	Existing results		Possible conclusions: 1) Indicators of endocrine activity and apical endpoints positive 2) Indicators of endocrine activity positive and apical endpoints negative 3) Indicators of endocrine activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
O	–	–	Eq/0	The chemical is probably not an ED acting on reproduction in fish. There is a possibility that the chemical is able to affect sexual development in fish, but the probability of this is low given the apparent absence of estrogenic, androgenic or steroidogenic properties.	Probably no further action. However, see comments in right-hand column.	If the paucity of *in vivo* data are a concern, performance of a screening test (OECD TG 229 or TG 230) with a different species, or a longer term test (i.e. TG 234 [FSDT] or life cycle – MEOGRT or ZEOGRT) could be considered. It is also possible that the chemical may be an anti-androgen *in vivo* (consider performing the Androgenised Female Stickleback Screen, or JMASA), or a thyroid-active chemical *in vivo* (consider performing the AMA – OECD TG 231, or XETA), although lack of *in vitro* binding affinity with receptors suggests this is unlikely. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
P	–	Eq/0	+	The chemical may not be an ED acting on reproduction in fish, but confidence in this conclusion is low given the lack of mechanistic *in vitro* data and the availability of positive existing *in vivo* data.	Obtain more predictive mechanistic data, then consider further testing.	If the mechanistic data confirm that the chemical has potential endocrine action, consider conducting a fish assay (OECD TG 229 or TG 230) with another species, or a longer term test (TG 234 [FSDT] or life cycle – MEOGRT or ZEOGRT). Use the existing *in vivo* data as a guide to test design. If the mechanistic data reveal anti-androgenic or thyroid activity, perform the Androgenised Female Stickleback Screen or JMASA, or the AMA (OECD TG 231) or XETA, respectively. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

Scenarios	Result of OECD TG 229 assay	Existing results		Possible conclusions: 1) Indicators of endocrine activity and apical endpoints positive 2) Indicators of endocrine activity positive and apical endpoints negative 3) Indicators of endocrine activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
Q	–	Eq/0	–	The chemical is probably not an ED acting on reproduction in fish, but the lack of more predictive mechanistic data are a concern, even though the existing *in vivo* data are negative.	Obtain more predictive mechanistic data, then consider further testing.	If the mechanistic data confirm that the chemical has potential endocrine action, consider conducting a fish assay (OECD TG 229 or TG 230) with another species, or a longer term test (TG 234 [FSDT] or life cycle – MEOGRT or ZEOGRT). Use the existing *in vivo* data as a guide to test design. If the mechanistic data reveal anti-androgenic or thyroid activity, perform the Androgenised Female Stickleback Screen or JMASA, or the AMA (OECD TG 231) or XETA, respectively. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
R	–	Eq/0	Eq/0	The chemical is probably not an ED acting on reproduction in fish, but confidence in this conclusion is low given the lack of mechanistic *in vitro* and existing *in vivo* data.	Obtain more predictive mechanistic data, then consider further testing.	If the mechanistic data confirm that the chemical has potential endocrine action, consider conducting a fish assay (OECD TG 229 or TG 230) with another species, or a longer term test (TG 234 [FSDT] or life cycle – MEOGRT or ZEOGRT). Use the existing *in vivo* data as a guide to test design. If the mechanistic data reveal anti-androgenic or thyroid activity, perform the Androgenised Female Stickleback Screen or JMASA, or the AMA (OECD TG 231) or XETA, respectively. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

C.2.2. 21-Day Fish Assay (OECD TG 230)

Status: Assay validated by the OECD.

297. Modality detected/endpoints: estrogens (♂ VTG ↑; ♂ 2º sex characteristics ↓); anti-estrogens (♀ VTG ↓); androgens (♂ 2º sex characteristics in ♀); anti-androgens (♂ 2º sex characteristics ↓); aromatisable androgens (♂ VTG ↑); aromatase inhibitors (♀ VTG ↓). Note that this assay has low statistical power to identify anti-androgenic activity.

Background to the assay

298. This assay is designed as a screen for the types of *in vivo* endocrine disruption activity in fish which are listed above. The endpoints are indicators of hormonal activity and there are no apical measures of adverse effects **that can be attributed to a single estrogen/androgen/thyroid/steroidogenesis (E,A,T,S) modality** (although it is possible that some substances could cause cessation of spawning). A variation of this assay specifically designed for the detection of androgens and anti-androgens, the Androgenised Female Stickleback Screen (AFSS), is described elsewhere in this document. Anti-androgens may also be detected by the Juvenile Medaka Anti-Androgen Screening Assay (JMASA).

When/why the assay may be used

299. Although data from OECD TG 230 could, in principle, be available at any stage in the hazard assessment process, the most likely scenario will be when there are relatively few data available about the possible endocrine disrupting properties of a chemical. The assay is most likely to be used either as part of a battery of *in vitro* and *in vivo* screens, or to follow up on existing data which suggest possible endocrine disruption activity. It is also possible that no existing endocrine-relevant data are available (i.e. OECD TG 230 has been used as a primary screen), but in that case a positive result in the screen should ideally be followed up with relevant *in vitro* screening in an attempt to confirm the suspected mode of action (MOA). Given the high degree of endocrine system conservation across the vertebrates, endocrine activity in this assay may also indicate the possibility of related activity in other organisms such as amphibians, reptiles, birds or mammals. Possible conclusions to be derived from the results of OECD TG 230, and guidance about potential additional studies to strengthen weight of evidence, are summarised in Table C.2.2.

300. In order to provide information relevant for assessing whether or not a chemical may fulfil the WHO/IPCS (2002) definition of an endocrine disrupter (ED), the study design has to be sufficiently robust to demonstrate the presence or absence of effects. In the dose selection, the investigator should also consider and ensure that data generated are adequate to fulfil the regulatory requirement across OECD countries as appropriate (e.g. hazard and risk assessment and labelling, ED assessment, etc.). The top dose or concentration should be sufficiently high to give clear systemic (i.e. non endocrine-specific) toxicity in order to ensure that a wide range of exposures (high to low) is tested. However, endocrine effects observed solely in the presence of clear systemic toxicity should be interpreted with caution

and may be disregarded when sufficiently justified to be caused by secondary effects which are unlikely to be due to endocrine activity. The reason for this advice is a concern that some endocrine active substance (EAS) sensitive assays are being run at doses/concentrations of EASs that are too low to trigger direct impacts on the endocrine system. This guidance document is not the place to address this issue directly, but it should be considered when EAS-sensitive test guidelines (TGs) are revised in the future. In addition, the number and spacing of dose/concentration levels should also be adequate to fulfil the objectives of the study (e.g. to demonstrate dose response relationships if this is required).

Existing data to be considered

301. Given the commonality of endocrine mechanisms in the vertebrates, relevant existing data available before deployment of OECD TG 230 might include *in vivo* results obtained with other vertebrates (e.g. a Uterotrophic Assay with rodents, positive findings for endocrine endpoints in mammalian repeat dose toxicity or reproductive studies), or one or more of a range of *in silico* or *in vitro* results which suggest that the modalities indicated above may occur *in vivo*. Such indicators of possible *in vivo* activity might include quantitative structure activity relationship (QSAR) predictions of endocrine activity, high throughput screening (HTS) data, "read-across" from *in vivo* results obtained with structurally related chemicals, or positive results from an *in vitro* screen for estrogen or androgen receptor-mediated activity, or for effects on steroidogenesis (especially aromatase inhibition). Further strong indication of *in vivo* estrogenic activity may also be available from an EASZY Assay with transgenic zebrafish embryos.

Scenarios: Positive and negative results combined with existing data

302. The scenarios (A to R) presented in Table C.2.2 represent all the possibilities of positive or negative results in combination with the presence or absence of existing data. The action taken will also depend on the regulatory environment, but the considerations given here are generally science based. Wherever possible, the recommended "next step which could be taken" avoids unnecessary animal testing. However, sometimes conducting an animal test will be indicated and then the relevance of species, strain and exposure route should always be considered. Further considerations specific to each scenario are given in the table.

303. Positive results obtained with one or more of the endpoints (Table C.2.2, Scenarios A-I) result in the conclusion that the test chemical is a potential ED *in vivo*. This would ideally need to be followed up with more comprehensive testing to show whether adverse apical effects related to endocrine impacts occur at any part of the life cycle (and hence to discover whether the chemical is **an ED acting through E,A,T,S pathways**). In other words, a positive result in OECD TG 230 may trigger TG 234 (Fish Sexual Development Test [FSDT]) at Level 4 or fish life cycle testing at Level 5. Existing data suggesting endocrine activity will strengthen the case for additional testing.

304. The situation in which OECD TG 230 gives a negative result (Table C.2.2, Scenarios J-R) needs careful consideration of any existing data. If the weight of evidence of these data suggests that the chemical is endocrine active both *in vitro* and *in vivo* in other species (Scenario J), then the probability is that OECD TG 230 may simply be insufficiently responsive in that case, or fish in general may be unresponsive. In some of these circumstances, it might be appropriate to conduct an FSDT (OECD TG 234), or

alternatively, a fish life cycle test (either MEOGRT – OECD TG 240, or ZEOGRT) to confirm that there is no endocrine activity in fish.

305. If OECD TG 230 and existing *in vivo* data are all negative, but *in vitro* data reveal some endocrine activity (Scenario K), the probability is that the test chemical is not sufficiently potent to produce endocrine effects *in vivo* in adult fish, or it may be rapidly metabolised. However, TG 230 does not include some endpoints which are included in TG 229 (fecundity and histopathology), which is able to detect certain endocrine-active substances not detected by TG 230 alone. In such a situation, further testing may or may not be necessary. A lack of effects in adult fish does not preclude the possibility that endocrine-mediated effects may manifest in fish exposed during a more sensitive life stage (e.g. as embryos or larvae). If the chemical is known to bioaccumulate slowly, it may be that exposures in the *in vivo* tests are not of sufficient duration, in which case longer term testing might be justified. If the *in vitro* data reveal anti-androgenic or thyroid activity, consideration should be given to conducting the Androgenised Female Stickleback Screen (AFSS – OECD GD 148) or Juvenile Medaka Anti-Androgen Screening Assay (JMASA) or the Amphibian Metamorphosis Assay (OECD TG 231), respectively.

306. On the other hand, if OECD TG 230 and the *in vitro* tests are negative but there are positive existing *in vivo* data (Scenario M), the chemical is probably not a potential ED with the modalities listed above, but it may act via estrogen- or androgen-related modes of action (MOA) not covered by the *in vitro* screens, or it may be more potent in species or life stages that have not been tested. In this situation, the relevant existing *in vitro* and *in vivo* data should be used to guide decisions about whether to conduct any further testing, either for modalities such as anti-androgenicity or including life stages represented in OECD TG 234 (FSDT) or in the MEOGRT or ZEOGRT.

307. Finally, a negative OECD TG 230 screen, set against a background of negative *in vitro* and *in vivo* data (Scenario N) **that includes relevant *in vivo* data for fish**, suggests that the test chemical is not a potential ED in fish or other vertebrates via estrogenic, anti-estrogenic, androgenic or steroidogenic MOA, and no further testing will generally be necessary for these modalities. It remains possible that it has anti-androgenic or thyroid activity, although negative *in vitro* tests for these modalities would suggest that this scenario is unlikely.

308. In each of the above scenarios, it is possible that existing data will be equivocal, or there may be no existing data (Scenarios C, F-I, L and O-R). This will weaken the conclusions which can be drawn about a negative OECD TG 230 test, and this is reflected in Table C.2.2. However, a lack of mechanistic data on endocrine activity should usually be rectified before any further *in vivo* testing is finally rejected. Indeed, as a general principle, it is desirable to obtain mechanistic data before any *in vivo* testing. On the other hand, if OECD TG 230 is positive, further *in vivo* testing is generally indicated, particularly when existing data are equivocal, or if there are no existing data. There is also the possibility that equivocal mechanistic data may be the result of multiple modes of endocrine action. Under some circumstances, two opposite modes of simultaneous action (e.g. estrogenic and anti-estrogenic) could, depending on dose, lead to a minimisation or abolition of effects, while in others two different MOA (e.g. estrogenic and anti-androgenic) could potentially reinforce effects. If multiple MOA are suspected, either from the existing results or based on QSAR/read-across/integrated approaches, this situation should be investigated further if needed for regulatory decision making and, if necessary, the weight given to the apparently equivocal mechanistic data should be increased.

309. The scenario in which the results of OECD TG 230 are themselves equivocal has not been dealt with in Table C.2.2, for reasons of brevity. In this context, an equivocal result might be an inconsistent concentration-response (e.g. no effect at a high concentration but effects at a lower concentration), or a result which borders on statistical significance. Without knowing the exact circumstances, reliable advice cannot be given, but the opinions of an experienced ecotoxicologist should be sought. Clearly, however, such equivocal results do not necessarily rule out the existence of *in vivo* endocrine activity. For example, vitellogenin (VTG) induction in males at a high concentration might be masked by any systemic toxicity, while VTG depression in females might just fail to reach a statistically significant level because VTG levels were relatively low to begin with. If these or other possible reasons for false negatives are suspected with good reason, the screen could be repeated if none of the test concentrations have given reliable data (e.g. conduct it at lower concentrations which avoid systemic toxicity), or a more appropriate version of it (e.g. ensure females have high VTG levels at the start of the test) could be conducted. However, note that a repeat test in the event of systemic toxicity would not be needed providing at least one tested concentration was not subject to such effects.

310. In summary, positive results in the OECD TG 230 screen indicate that a chemical is a possible endocrine disrupter. More predictive *in vivo* testing would then be necessary to produce a long-term no-observed-effect-concentration/x% effect concentration (NOEC/ECx) and/or to confirm whether or not the chemical is an actual endocrine disrupter with adverse effects *in vivo*. Negative results in OECD TG 230 do not necessarily mean that the chemical is not a potential ED – a judgement about its endocrine disruption potential and the possible need for additional testing will have to be made based on a weight of evidence evaluation of existing *in vitro* and *in vivo* data.

Reference

WHO/IPCS (2002), "Global assessment of the state-of-the-science of endocrine disrupters", Damstra, T. et al. (eds.) WHO/PCS/EDC/02.2, World Health Organization, Geneva, www.who.int/ipcs/publications/new_issues/endocrine_disruptors/en.

Table C.2.2. **21-Day Fish Assay (OECD TG 230):**
Guidance for scenarios of combinations of results with existing data

This table represents possible conclusions to be drawn from assay data, and a next step which could be taken if further evidence is required about possible endocrine disrupting properties and/or effects. The guidance offered is not meant to be prescriptive, but provides science-based considerations. It encourages the use of all available data and expert judgement in a weight of evidence approach. Regional and national interpretation of results and "next steps" may vary.

The conclusions are grouped into a series of scenarios (A-R), each scenario representing a different combination of assay results, existing *in vitro* data and existing *in vivo* data. The symbol "+" indicates that the data in question represent a positive result, "-" indicates a negative result, and "Eq/0" indicates that the data are either equivocal or are not available. Note that there are no apical endpoints in this assay considered to be diagnostic of an E,A,T,S modality.

Existing results: * "Mechanism (*in vitro* mechanistic data)" assumes that mechanistic data are available from estrogen receptor (ER-), androgen receptor (AR-) and steroidogenesis-based assays (Level 2). Thyroid hormone receptor (TR) and other assays concerning mechanisms of thyroid disruption may be available, but they are not in common use. In practice, data from all assays may not be available and therefore this must be taken into account when deciding on the "next step". Quantitative structure activity relationship (QSAR) predictions of etrogen and androgen binding/activation may be made for some substances. There is no evidence at present that equivalent *in vitro* assays with systems derived from fish offer advantages over their mammalian counterparts.

Existing results: ** "Effects (*in vivo* effects of concern)" assumes effects have been observed in other *in vivo* screens/tests which give rise to concern that the test chemical may be an endocrine disrupter.

Scenarios	Result of OECD TG 230 assay	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
A	+	+	+	Strong evidence for *in vivo* endocrine activity in fish and other organisms.	Consider performing a fish life cycle test (MEOGRT – OECD TG 240, or ZEOGRT), especially if the intention is to obtain precise data on a reproductive or developmental no-observed-effect-concentration/x% effect concentration (NOEC/ECx).	An alternative approach would be to deploy TG 234 (Fish Sexual Development Test [FSDT]), especially if sexual development is expected to give a response at lower concentrations than reproduction.
B	+	+	–	Strong evidence for *in vivo* endocrine activity in fish, despite lack of *in vivo* effects in existing tests.	Consider performing a fish life cycle test (MEOGRT – OECD TG 240, or ZEOGRT), especially if the intention is to obtain precise data on a reproductive or developmental NOEC/ECx.	An alternative approach would be to deploy OECD TG 234 (FSDT), especially if sexual development is expected to give a response at lower concentrations than reproduction. If the negative *in vivo* data are from a fish test (e.g. OECD TG 229), consider possible reasons for the disparity (e.g. differences in species sensitivity) before conducting a life cycle test.
C	+	+	Eq/0**	Strong evidence for *in vivo* endocrine activity in fish, despite equivocal or absent *in vivo* data in other species.	Consider performing a fish life cycle test (MEOGRT – OECD TG 240, or ZEOGRT), especially if the intention is to obtain precise data on a reproductive or developmental NOEC/ECx.	If no existing fish data are available, it may be worth performing OECD TG 234 (FSDT) before a possible life cycle test in order to obtain information on whether sexual development is a sensitive part of the life cycle. Such information could influence the design of the life cycle test. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple modes of action (MOA). If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
D	+	–	+	Moderate evidence for *in vivo* endocrine activity in fish and other species, but confidence about MOA is reduced by negative mechanistic data.	Consider performing a fish life cycle test (MEOGRT – OECD TG 240, or ZEOGRT), especially if the intention is to obtain precise data on a reproductive or developmental NOEC/ECx.	The negative *in vitro* data suggest that the test chemical may be metabolically activated *in vivo,* or it may operate via mechanisms not covered by the *in vitro* screens. An alternative approach to a life cycle test would be to deploy OECD TG 234 (FSDT), especially if sexual development is expected to give a response at lower concentrations than reproduction.
E	+	–	–	Moderate-strong evidence for *in vivo* endocrine activity in fish, but confidence is reduced by negative *in vitro* data and negative *in vivo* activity in other species.	Consider performing a fish life cycle test (MEOGRT – OECD TG 240, or ZEOGRT), especially if the intention is to obtain precise data on a reproductive or developmental NOEC/ECx.	The negative *in vitro* data suggest that the test chemical may be metabolically activated *in vivo*, or it may operate via mechanisms not covered by the *in vitro* screens. An alternative approach to a life cycle test would be to deploy OECD TG 234 (FSDT), especially if sexual development is expected to give a response at lower concentrations than reproduction. If the negative *in vivo* data are from a fish test (e.g. OECD TG 229), consider possible reasons for the disparity (e.g. differences in species sensitivity) before conducting a life cycle test.

Scenarios	Result of OECD TG 230 assay	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
F	+	–	Eq/0	Moderate-strong evidence for *in vivo* endocrine activity in fish, but confidence is reduced by negative *in vitro* data and equivocal or absent *in vivo* activity in other species.	Consider performing a fish life cycle test (MEOGRT – OECD TG 240, or ZEOGRT), especially if the intention is to obtain precise data on a reproductive or developmental NOEC/ECx.	The negative *in vitro* data suggest that the test chemical may be metabolically activated *in vivo*, or it may operate via mechanisms not covered by the *in vitro* screens. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information. If no existing fish data are available, it may be worth performing OECD TG 234 (FSDT) before a possible life cycle test in order to obtain information on whether sexual development is a sensitive part of the life cycle. Such information could influence the design of the life cycle test.
G	+	Eq/0	+	Strong evidence for *in vivo* endocrine activity in fish, but mechanism unconfirmed.	Obtain mechanistic data, then consider performing a fish life cycle test (MEOGRT – OECD TG 240, or ZEOGRT).	An alternative approach to a life cycle test would be to deploy OECD TG 234 (FSDT), especially if sexual development is expected to give a response at lower concentrations than reproduction. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
H	+	Eq/0	–	Strong-moderate evidence for *in vivo* endocrine activity in fish, but mechanism unconfirmed.	Obtain mechanistic data, then consider performing a fish life cycle test (MEOGRT – OECD TG 240, or ZEOGRT).	An alternative approach to a life cycle test would be to deploy OECD TG 234 (FSDT), especially if sexual development is expected to give a response at lower concentrations than reproduction. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information. If the negative *in vivo* data are from a fish test (e.g. OECD TG 229), consider possible reasons for the disparity (e.g. differences in species sensitivity) before possibly conducting a life cycle test.
I	+	Eq/0	Eq/0	Moderate evidence for *in vivo* endocrine activity in fish, but mechanism unconfirmed.	Obtain mechanistic data, then consider performing a fish life cycle test (MEOGRT – OECD TG 240, or ZEOGRT).	It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information. If no existing fish data are available, it may be worth performing OECD TG 234 (FSDT) before a possible life cycle test (MEOGRT – OECD TG 240, or ZEOGRT) in order to obtain information on whether sexual development is a sensitive part of the life cycle. Such information could influence the design of the life cycle test.
J	–	+	+	Based on the existing data, the chemical has endocrine activity *in vivo*. The lack of response in OECD TG 230 suggests that fish are not responsive, unless the existing data are from fish.	Consider performing OECD TG 234 (FSDT).	It is possible that the failure to give a positive result in OECD TG 230 was caused by the relatively short exposure time (three weeks). If this is suspected (e.g. the chemical only bioaccumulates slowly), or if the existing *in vivo* data are from a fish, OECD TG 234 (FSDT) or potentially a life cycle test (MEOGRT – OECD TG 240, or ZEOGRT) would be able to study the effects of longer exposure and confirm whether there is a hazard to fish. Choice of test should be guided by the existing *in vivo* data.

Scenarios	Result of OECD TG 230 assay	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
K	–	+	–	There is no evidence that the chemical is a potential ED *in vivo*, probably because it is very weakly acting or rapidly metabolised.	Probably no further action, but see comments in right-hand column.	It is possible that EDs which bioaccumulate slowly may only cause effects *in vivo* after exposure times longer than three weeks. If this is suspected, and depending on which part of the life cycle is suspected of being the most sensitive, consider performing OECD TG 234 (FSDT), or a fish life cycle test (MEOGRT – OECD TG 240, or ZEOGRT). It is also possible that the chemical may be an anti-androgen *in vivo* (consider performing the Androgenised Female Stickleback Screen [AFSS] or Juvenile Medaka Anti-Androgen Screening Assay [JMASA]), or a thyroid-active chemical *in vivo* (consider performing the Amphibian Metamorphosis Assay [AMA] – OECD TG 231, or *Xenopus* Embryo Thyroid Signalling Assay [XETA]).
L	–	+	Eq/0	The chemical may not be an ED *in vivo*, but the confidence in this conclusion is relatively low as there is only one unequivocal *in vivo* test result (a negative).	Probably no further action, but see comments in right-hand column.	If the equivocal existing data are from a fish assay, consider performing a fish assay (OECD TG 229 or TG 230) with a different species, or a longer term test (TG 234 [FSDT] or life cycle [MEOGRT – OECD TG 240, or ZEOGRT]) if the chemical is a slow bioaccumulator. It is also possible that the chemical may be an anti-androgen *in vivo* (consider performing the AFSS or JMASA), or a thyroid-active chemical *in vivo* (consider performing the AMA – OECD TG 231, or XETA). It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
M	–	–	+	The chemical is apparently not a potential ED in fish but it does have activity in other species.	Use the existing *in vivo* data to help decide whether a longer term test with an appropriate fish species is indicated.	It is also possible that the chemical may be an anti-androgen *in vivo* (consider performing the AFSS or JMASA), or a thyroid-active chemical *in vivo* (consider performing the AMA – OECD TG 231, or XETA), although lack of *in vitro* binding affinity with the androgen receptor suggests the former is unlikely. Use the existing *in vivo* data to guide any further testing.
N	–	–	–	The chemical is probably not a potential ED *in vivo*.	No further action with respect to estrogenic, anti-estrogenic, androgenic or steroidogenic MOA.	It is still possible that the chemical may be an anti-androgen *in vivo* (consider performing the AFSS or JMASA), or a thyroid-active chemical *in vivo* (consider performing the AMA – OECD TG 231, or XETA), although lack of *in vitro* binding affinity with the androgen receptor suggests the former is unlikely.

Scenarios	Result of OECD TG 230 assay	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
O	–	–	Eq/0	The chemical is probably not a potential ED in fish.	Probably no further action. However, see comments in right-hand column.	If the paucity of *in vivo* data is a concern, performance of a screening test (OECD TG 229 or TG 230) with a different species, or a longer term test (i.e. TG 234 [FSDT] or life cycle) could be considered. It is still possible that the chemical may be an anti-androgen *in vivo* (consider performing the AFSS or JMASA), or a thyroid-active chemical *in vivo* (consider performing the AMA – OECD TG 231, or XETA), although lack of *in vitro* binding affinity with the androgen receptor suggests the former is unlikely. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
P	–	Eq/0	+	The chemical is probably not a potential ED in fish, but confidence in this conclusion is low given the lack of mechanistic *in vitro* data and the availability of positive existing *in vivo* data.	Obtain mechanistic data, then consider whether further testing is desirable.	If the mechanistic data confirm that the chemical has potential endocrine action, consider conducting a fish assay (OECD TG 229 or TG 230) with another species, or a longer term test (OECD TG 234 [FSDT] or life cycle). Use the existing *in vivo* data as a guide to test choice. If the mechanistic data reveal anti-androgenic or thyroid activity, consider performing the AFSS or JMASA), or a thyroid-active chemical *in vivo* (consider performing the AMA – OECD TG 231, or XETA), respectively. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
Q	–	Eq/0	–	The chemical is probably not a potential ED in fish, but the lack of mechanistic *in vitro* data are a concern, even though the existing *in vivo* data are negative.	Obtain mechanistic data, then consider whether further testing is desirable.	If the mechanistic data confirm that the chemical has potential endocrine action, consider conducting a fish assay (OECD TG 229 or TTG 230) with another species, or a longer term test (OECD TG 234 [FSDT]) or life cycle). Use the existing *in vivo* data as a guide to test choice. If the mechanistic data reveal anti-androgenic or thyroid activity, consider performing the AFSS or JMASA) or a thyroid-active chemical *in vivo* (consider performing the AMA – OECD TG 231, or XETA), respectively. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

Scenarios	Result of OECD TG 230 assay	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
R	–	Eq/0	Eq/0	The chemical is probably not a potential ED in fish, but confidence in this conclusion is low given the lack of mechanistic *in vitro* and existing *in vivo* data.	Obtain mechanistic data, then consider whether further testing is desirable.	If the mechanistic data confirm that the chemical has potential endocrine action, consider conducting a fish assay (OECD TG 229 or TG 230) with another species, or a longer term test (OECD TG 234 [FSDT] or life cycle). Use the existing *in vivo* data as a guide to test choice. If the mechanistic data reveal anti-androgenic or thyroid activity, consider performing the AFSS or JMASA), or a thyroid-active chemical *in vivo* (consider performing the AMA – OECD TG 231, or XETA), respectively. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

C.2.3. Amphibian Metamorphosis Assay (AMA) (OECD TG 231)

Status: Assay validated by the OECD.

311. Modality detected/endpoints: thyroid activity (advanced development; asynchronous development; delayed development in absence of non-specific systemic toxicity; thyroid histopathology), but note that this covers several different modes of action (MOA), including thyroid agonists and antagonists, as well as substances interfering with thyroid hormone synthesis and transport. According to OECD TG 231, there is disagreement about the implications of the different endpoints in this larval development screen. Some experts accept that changes in one of the thyroid-relevant apical endpoints (advanced development; asynchronous development; delayed development in absence of non-specific systemic toxicity) may on their own provide information on thyroid activity, while others will only reach this conclusion if one of the apical endpoints is accompanied by significant thyroid histopathology, such as moderate or severe follicular hypertrophy and/or hyperplasia (OECD, 2007). Note that the AMA is subject to indirect thyroid effects such as those that result from cytochrome P450 induction (e.g. phenobarbital, the model compound for the latter effect, tests positive in the AMA). Therefore, interpretation of the AMA may be complicated.

Background to the assay

312. This assay is designed as a screen for thyroid activity in amphibians, and not to provide information on endocrine activity for use in assessing the environmental risks of an individual chemical based on a predicted environmental concentration/predicted no-effect concentration (PEC/PNEC) approach. Delay in metamorphosis could be considered an apical endpoint, but the significance of a short delay in metamorphosis for amphibian populations is poorly understood except for amphibian species living in temporary pools that dry out at, or shortly after, the normal time for metamorphosis to be completed. Furthermore, the use of only three concentrations of test chemical precludes the reliable establishment of a no-observed-effect-concentration/x% effect concentration (NOEC/LOEC). It is important to note that there are several types of thyroid disruption, not all of which involve interactions with the thyroid receptor, and they have differential effects on the various endpoints in this screen. OECD TG 231 does not, however, allow unequivocal diagnosis of which type of thyroid disruption is occurring. It includes a specific endpoint (thyroid gland histopathology) for some types of thyroid activity, but also includes apical measurements (hind limb length, snout-vent length, developmental stage and wet weight), which are used to determine other thyroid-responsive endpoints: advanced development, asynchronous development or delayed development. The first two of these are considered by some authorities to be diagnostic of thyroid activity, while the latter is only diagnostic if non-specific systemic toxicity is absent. It should also be noted that a review (Pickford, 2010) concluded that for thyroid agonists, the response of amphibian thyroid histopathology is not as predictable or as sensitive as developmental stage or hind limb development. However, it is probable that a diagnosis of thyroid activity on the basis of the apical endpoints will be more robust if accompanied by thyroid histopathology, and vice versa.

313. Consequently, if the assay gives a positive result, this may be due to a combination of a positive indicator of hormonal activity (thyroid histopathology) and a positive apical endpoint (advanced development, asynchronous development or delayed development), or a positive indicator of hormonal activity alone (possibly accompanied by a negative apical endpoint), or for an apical endpoint alone (possibly accompanied by a negative indicator of hormonal activity). Each of these possible combinations of positive response should be considered separately (although the distinctions between indicators of hormonal activity and apical effects are not always clear), so they have been listed individually as points 1, 2 and 3 in the possible conclusions column of Table C.2.3. It should be noted, however, that due to the relatively short exposure time employed in this screen (three weeks), one cannot be sure if the effects of some chemicals on apical endpoints would result in adverse effects on development, growth or reproduction in the longer term. This is primarily relevant for hazard identification/characterisation. Given the high degree of endocrine system conservation across the vertebrates, endocrine-linked effects in the AMA may also indicate the possibility of related activity in other organisms such as fish, reptiles, birds or mammals.

When/why the assay may be used

314. Although OECD TG 231 could, in principle, be used at any stage in the hazard assessment process, the most likely use scenario will be when there are relatively few data available about the possible thyroid disrupting properties of a chemical. The results from this assay are most likely to be available after deployment of a battery of *in vitro* and *in vivo* screens (e.g. the United States Environmental Protection Agency's Endocrine Disrupter Screening Program), or as a supplement to existing data which suggest potential endocrine disrupter (ED) activity (e.g. a positive result in the *Xenopus* embryonic thyroid signalling assay [XETA]). A number of mammalian (rat) assays are sensitive to thyroid disruption, particularly thyroid antagonists, including the pubertal assay (male or female), the enhanced repeat dose assay (OECD TG 407) and the intact male screening assay. Note that these assays utilise different routes of exposure than OECD TG 231 and therefore, depending on the properties of the chemical, have differing potentials for the test substance to be metabolised. It should also be noted that only the AMA and the XETA appear to be sensitive to thyroid agonists. It has been argued by Pickford (2010) that only one thyroid-disrupting chemical (methoxychlor) shows activity in the AMA but not in any rodent screens, but the number of chemicals tested in the former is less than in the latter.

315. It is possible that no endocrine-relevant data are available before the AMA is deployed (i.e. if OECD TG 231 has been used as a primary screen), but in that case a positive result in the screen could be followed up with relevant *in vitro* screening to investigate the suspected MOA. However, it should be noted that *in vitro* screens essentially only exist for thyroid agonists and antagonists (e.g. GH_3 rat pituitary somatotroph cell proliferation; solid state thyroid receptor binding assays; transfected reporter gene assays in yeast or mammalian cell lines), while thyroid disruption can occur at other points in the endocrine system for which *in vitro* screens do not exist, or are still at the research stage (e.g. FRTL-5 rat cell lines sensitive to iodide uptake inhibitors) (see Paragraph 18). Furthermore, none of these screens have yet been validated and standardised at the international level.

316. In order to provide information relevant for assessing whether or not a chemical may fulfil the WHO/IPCS (2002) definition of an ED, the study design has to be sufficiently robust to demonstrate the presence or absence of effects. In the dose selection, the investigator should also consider and ensure that data generated are adequate to fulfil the regulatory

requirement across OECD countries as appropriate (e.g. hazard and risk assessment and labelling, ED assessment, etc.). The top dose or concentration should be sufficiently high to give clear systemic (i.e. non endocrine-specific) toxicity in order to ensure that a wide range of exposures (high to low) is tested. However, endocrine effects observed solely in the presence of clear systemic toxicity should be interpreted with caution and may be disregarded when sufficiently justified to be caused by secondary effects which are unlikely to be due to endocrine activity. The reason for this advice is a concern that some endocrine active substance (EAS) sensitive assays are being run at doses/concentrations of EASs that are too low to trigger direct impacts on the endocrine system. This guidance document is not the place to address this issue directly, but it should be considered when EAS-sensitive test guidelines (TGs) are revised in the future. In addition, the number and spacing of dose/concentration levels should also be adequate to fulfil the objectives of the study (e.g. to demonstrate dose response relationships if this is required).

Existing data to be considered

317. Given the commonality of endocrine mechanisms in the vertebrates, relevant existing data available before deployment of OECD TG 231 might include *in vivo* results obtained with other vertebrates (e.g. a positive *in vivo* assay with rats – see above, positive findings for thyroid endpoints in mammalian repeat dose toxicity or reproductive studies), or one or more of a range of *in silico* or *in vitro* results which suggest that thyroid disruption may occur *in vivo* (but note the limitations of this approach, as indicated above). Such indicators of possible thyroid activity might include quantitative structure activity relationship (QSAR) predictions of thyroid activity, "read-across" from *in vivo* results obtained with chemically related chemicals or positive results from an *in vitro* screen for thyroid agonist/antagonist activity. Data from the XETA, if available, should also be considered.

Scenarios: Positive and negative results combined with existing data

318. The scenarios (A to R) presented in Table C.2.3 represent all the possibilities of positive or negative results in combination with the presence or absence of existing data. The action taken will also depend on the regulatory environment, but the considerations given here are generally science based. Wherever possible, the recommended "next step which could be taken" avoids unnecessary animal testing. However, sometimes conducting an animal test will be indicated and then the relevance of species, strain and exposure route should always be considered. Further considerations specific to each scenario are given in the table.

319. Positive results obtained with the thyroid histopathology endpoint (Table C.2.3, Scenarios A-I, sub-section 2) result in the conclusion that the test chemical is a potential ED *in vivo*. If both thyroid histopathology and an apical endpoint give a response (Table C.2.3, Scenarios A-I, sub-section 1), this may provide even stronger evidence that one is dealing with a potential ED, especially if its action is not receptor-mediated. If only an apical endpoint responds (Table C.2.3, Scenarios A-I, sub-section 3), it suggests that the chemical is a possible thyroid disrupter, but with somewhat reduced confidence in some cases compared to sub-section 2 (although existing positive *in vitro* data, or positive *in vivo* data from other species, would have to be weighed against this conclusion). Note, however, that apical endpoints alone are probably sufficiently responsive to thyroid receptor agonists (i.e. in these cases thyroid histopathology is unlikely to make the assay more robust) (Daniel Pickford, pers. comm., 2010).

320. As indicated above, although a positive response of OECD TG 231 indicates that the chemical is a possible thyroid disrupter, a result of this type would generally need to be followed up with a more comprehensive growth, development and/or reproduction test if countries need further evidence (i.e. a Larval Amphibian Growth and Development Assay [LAGDA] – OECD TG 241) which is able to provide a precise NOEC/ECx for adverse effects. In other words, in order to strengthen weight of evidence, a positive result of whichever type in OECD TG 231 could be followed by a LAGDA at Level 4. Existing data suggesting endocrine-specific activity (e.g. positive *in vitro* or XETA data, or positive *in vivo* data from other species) will strengthen the case for additional testing still further. Note, however, that the LAGDA is not a true life cycle test and does not include all aspects of reproduction. For that reason, it is worth considering whether a positive result in OECD TG 231 could be more usefully followed up under some circumstances by a MEOGRT/ZEOGRT with thyroid-specific endpoints such as thyroid hormone induction or depression, although at present the responsiveness of apical endpoints in these tests (e.g. growth) to thyroid-active substances is not well understood.

321. The situation in which OECD TG 231 gives a negative result (Table C.2.3, Scenarios J-R) needs careful consideration of any existing data. If these data suggest that the chemical is endocrine active both *in vitro* and *in vivo* (Scenario J), then it is possible that OECD TG 231 is simply insufficiently sensitive, although most known thyroid disrupters have been shown to give a response in the AMA. Depending on the robustness of the existing data, it might therefore be appropriate to conduct a LAGDA.

322. If OECD TG 231 and existing *in vivo* data are all negative, but *in vitro* data reveal some endocrine activity (Scenario K), the probability is that the test chemical is not sufficiently potent to produce thyroid effects *in vivo* in amphibians or other organisms, or it may be rapidly metabolised. In such a situation, further testing is probably not necessary. However, if the chemical is known to bioaccumulate slowly, it may be that exposures in the *in vivo* tests have been insufficiently prolonged, in which case longer term testing with the LAGDA might be justified.

323. On the other hand, if OECD TG 231 and the *in vitro* tests are negative (Scenario M), but there are positive existing *in vivo* data, the nature of those existing data should be considered. Unless the existing data are from another amphibian, the chemical is probably not an ED acting on amphibian growth or development, but it may act via MOA not covered by the *in vitro* screens, or it may be more potent in species or life stages that have not been tested. In this situation, the existing *in vivo* data should be used to guide decisions about whether to conduct any further testing.

324. Finally, a negative OECD TG 231 screen, set against a background of negative *in vitro* and *in vivo* data (Scenario N), suggests that the test chemical is not a possible thyroid-active ED, and further action is unnecessary.

325. In each of the above scenarios, it is possible that existing data will be equivocal, or there may be no existing data. This will weaken the conclusions which can be drawn about a negative OECD TG 231 test, and this is reflected in Table C.2.3. However, a lack of mechanistic data on thyroid activity should ideally be rectified before any further *in vivo* testing is finally conducted, although as indicated above, many thyroid modalities are not detectable in *in vitro* screens. On the other hand, if OECD TG 231 is positive, further *in vivo* testing would generally be needed to quantify any adverse effects and/or to establish a NOEC or ECx for such effects, even if all existing data are equivocal, or if there are no existing data. Again, however, it may be useful to obtain some mechanistic information before conducting further *in vivo* testing. There is also the possibility that equivocal

mechanistic data may be the result of multiple modes of endocrine action. Under some circumstances, two opposite modes of simultaneous action (e.g. thyroid and anti-thyroid) could, depending on dose, lead to a minimisation or abolition of adverse effects, while in others two different MOA could potentially reinforce effects on certain apical endpoints. If multiple MOA are suspected, either from the existing results or based on QSAR/read-across/integrated approaches, this situation should be investigated further if needed for regulatory decision making and if necessary, the weight given to the apparently equivocal mechanistic data should be increased.

326. The scenario in which the results of OECD TG 231 are themselves equivocal has not been dealt with in Table C.2.3, for reasons of brevity. In this context, an equivocal result might be an inconsistent concentration-response (e.g. no effect at a high concentration but effects at a lower concentration) or a result which borders on statistical significance. Without knowing the exact circumstances, reliable advice cannot be given, but the opinions of an experienced ecotoxicologist should be sought. Clearly, however, such equivocal results do not necessarily rule out the existence of *in vivo* endocrine activity. For example, thyroid histopathology at a high concentration might be masked by any systemic toxicity, while growth measurements might just fail to reach a statistically significant level due to unexpectedly high variability. If these or other possible reasons for false negatives are suspected with good reason, the screen could be repeated (e.g. conduct it at lower concentrations which avoid systemic toxicity), or a more appropriate version of it (e.g. more larvae per replicate) could be designed and conducted. However, note that a repeat screen in the event of systemic toxicity would not be needed providing at least one tested concentration was not subject to such effects.

327. In summary, certain positive results in the OECD TG 231 screen may indicate that a chemical is a possible endocrine disrupter via one of several types of thyroid activity. This suggests that more comprehensive *in vivo* testing would be needed if the intention is to derive a long-term NOEC/ECx and/or to confirm whether or not the chemical is an actual endocrine disrupter due to the occurrence of adverse effects. Negative results in OECD TG 231 do not necessarily mean that the chemical is not a potential ED – a judgement about the endocrine disruption potential and the possible need for additional testing will have to be made based on a weight of evidence evaluation of existing *in vitro* and *in vivo* data.

References

OECD (2007), "Guidance document on amphibian thyroid histology", OECD Series on Testing and Assessment, No. 82, OECD, Paris, www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2007)31&doclanguage=en.

Pickford, D.B. (2010), "Screening chemicals for thyroid-disrupting activity: A critical comparison of mammalian and amphibian models", *Critical Reviews in Toxicology*, Vol. 40/10, pp. 845-892, https://doi.org/10.3109/10408444.2010.494250.

WHO/IPCS (2002), "Global assessment of the state-of-the-science of endocrine disrupters", Damstra, T. et al. (eds.) WHO/PCS/EDC/02.2, World Health Organization, Geneva, www.who.int/ipcs/publications/new_issues/endocrine_disruptors/en.

Table C.2.3. **Amphibian Metamorphosis Assay (AMA) (OECD TG 231): Guidance for scenarios of combinations of results with existing data**

This table represents possible conclusions to be drawn from assay data, and a next step which could be taken if further evidence is required about possible endocrine disrupting properties and/or effects. The guidance offered is not meant to be prescriptive, but provides science-based considerations. It encourages the use of all available data and expert judgement in a weight of evidence approach. Regional and national interpretation of results and "next steps" may vary.

The conclusions are grouped into a series of scenarios (A-R), each scenario representing a different combination of assay results, existing *in vitro* data and existing *in vivo* data. The symbol "+" indicates that the data in question represent a positive result, "-" indicates a negative result, and "Eq/0" indicates that the data are either equivocal or are not available.

Existing results: * "Mechanism (*in vitro* mechanistic data)" assumes that mechanistic data are available from thyroid hormone receptor (TR) and other assays concerning mechanisms of thyroid disruption although these are not yet in common use. In practice, data from all assays may not be available and therefore this must be taken into account when deciding on the "next step". Quantitative structure activity relationship (QSAR) predictions of TR binding/activation may be made for some substances.

Existing results: ** "Effects (*in vivo* effects of concern)" assumes effects have been observed in other *in vivo* screens/tests which give rise to concern that the test chemical may be an thyroid disrupter.

The assay under discussion could either be positive for both apical endpoints and indicators of endocrine activity, or positive just for an apical endpoint or the indicator of endocrine activity. For each scenario, each of these three possibilities is addressed separately in the possible conclusions column.

Scenarios	Result of OECD TG 231 assay (AMA)	Existing results		Possible conclusions: 1) Indicators of endocrine activity and apical endpoints positive 2) Indicators of endocrine activity positive 3) Apical endpoint positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
A	+	+	+	1) Strong evidence for *in vivo* thyroid activity with potential adverse effects (developmental/growth toxicity) in amphibians, plus thyroid effects in other species. 2) Strong evidence for *in vivo* thyroid activity in amphibians, plus thyroid effects in other species. 3) Some evidence for *in vivo* thyroid activity with potential adverse effects (developmental/growth toxicity) in amphibians, plus thyroid effects in other species.	Consider performing a Larval Amphibian Growth and Development Assay (LAGDA – OECD TG 241).	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is a thyroid (ant)agonist.
B	+	+	–	1) Strong evidence for *in vivo* thyroid activity with potential adverse effects (developmental/growth toxicity) in amphibians. 2) Strong evidence for *in vivo* thyroid activity in amphibians. 3) Some evidence for *in vivo* thyroid activity with potential adverse effects (developmental/growth toxicity) in amphibians.	Consider performing a LAGDA (OECD TG 241).	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is a thyroid (ant)agonist. Cases where chemicals are active in the AMA but not in thyroid-responsive rodent assays are rare. In this scenario, it is therefore particularly important to discover if adverse effects appear in a longer term amphibian test.
C	+	+	Eq/0	1) Strong evidence for *in vivo* thyroid activity with potential adverse effects (developmental/growth toxicity) in amphibians. 2) Strong evidence for *in vivo* thyroid activity in amphibians. 3) Some evidence for *in vivo* thyroid activity with potential adverse effects (developmental/growth toxicity) in amphibians.	Consider performing a LAGDA (OECD TG 241).	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is a thyroid (ant)agonist. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple modes of action (MOA). If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
D	+	–	+	1) Moderate evidence for *in vivo* thyroid activity with potential adverse effects (developmental/growth toxicity) in amphibians, plus thyroid effects in other species. 2) Moderate evidence for *in vivo* thyroid activity in amphibians, plus thyroid effects in other species. 3) Some evidence for *in vivo* thyroid activity with potential adverse effects (developmental/growth toxicity) in amphibians, plus thyroid effects in other species.	Consider performing a LAGDA (OECD TG 241).	The lack of *in vitro* thyroid activity is not evidence against any thyroid activity, due to the limited nature of current *in vitro* thyroid screens.

Scenarios	Result of OECD TG 231 assay (AMA)	Existing results		Possible conclusions: 1) Indicators of endocrine activity and apical endpoints positive 2) Indicators of endocrine activity positive 3) Apical endpoint positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
E	+	–	–	1) Strong evidence for *in vivo* thyroid activity with potential adverse effects (developmental/growth toxicity) in amphibians. 2) Strong evidence for *in vivo* thyroid activity in amphibians. 3) Some evidence for *in vivo* thyroid activity with potential adverse effects (developmental/growth toxicity) in amphibians.	Consider performing a LAGDA (OECD TG 241).	The lack of *in vitro* thyroid activity is not evidence against any thyroid activity, due to the limited nature of current *in vitro* thyroid screens. Cases where chemicals are active in the AMA but not in thyroid-responsive rodent assays are rare. In this scenario, it is therefore particularly important to discover if adverse effects appear in a longer term amphibian test.
F	+	–	Eq/0	1) Strong evidence for *in vivo* thyroid activity with potential adverse effects (developmental/growth toxicity) in amphibians. 2) Strong evidence for *in vivo* thyroid activity in amphibians. 3) Some evidence for *in vivo* thyroid activity with potential adverse effects (developmental/growth toxicity) in amphibians.	Consider performing a LAGDA (OECD TG 241). Given the absence or equivocal nature of existing *in vivo* data, it might also be sensible to conduct a thyroid-responsive mammalian assay (e.g. rat pubertal).	The lack of *in vitro* thyroid activity is not evidence against any thyroid activity, due to the limited nature of current *in vitro* thyroid screens. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
G	+	Eq/0	+	1) Strong evidence for *in vivo* thyroid activity with potential adverse effects (developmental/growth toxicity) in amphibians, plus thyroid effects in other species. 2) Strong evidence for *in vivo* thyroid activity in amphibians, plus thyroid effects in other species. 3) Some evidence for *in vivo* thyroid activity with potential adverse effects (developmental/growth toxicity) in amphibians, plus thyroid effects in other species.	Consider performing a LAGDA (OECD TG 241). Given the absence or equivocal nature of the *in vitro* mechanistic data, it might also be helpful to conduct an *in vitro* screen for thyroid (ant)agonistic activity.	If a new *in vitro* mechanistic assay is conducted, note that a negative resultdoes not mean that the test material has no thyroid activity. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
H	+	Eq/0	–	1) Strong evidence for *in vivo* thyroid activity with potential adverse effects (developmental/growth toxicity) in amphibians. 2) Strong evidence for *in vivo* thyroid activity in amphibians. 3) Some evidence for *in vivo* thyroid activity with potential adverse effects (developmental/growth toxicity) in amphibians.	Consider performing a LAGDA (OECD TG 241). Given the absence or equivocal nature of the *in vitro* mechanistic data, it might also be helpful to conduct an *in vitro* screen for thyroid (ant)agonistic activity.	Cases where chemicals are active in the AMA but not in thyroid-responsive rodent assays are rare. In this scenario, it is therefore particularly important to discover if adverse effects appear in a longer term amphibian test. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

Scenarios	Result of OECD TG 231 assay (AMA)	Existing results		Possible conclusions: 1) Indicators of endocrine activity and apical endpoints positive 2) Indicators of endocrine activity positive 3) Apical endpoint positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
I	+	Eq/0	Eq/0	1) Strong evidence for *in vivo* thyroid activity with potential adverse effects (developmental/growth toxicity) in amphibians. 2) Strong evidence for *in vivo* thyroid activity in amphibians. 3) Some evidence for *in vivo* thyroid activity with potential adverse effects (developmental/growth toxicity) in amphibians.	Consider performing a LAGDA (OECD TG 241). Given the absence or equivocal nature of the *in vitro* mechanistic data, it might also be helpful to conduct an *in vitro* screen for thyroid (ant)agonistic activity. Given the absence or equivocal nature of existing *in vivo* data, it might also be sensible to conduct a thyroid-responsive mammalian assay (e.g. rat pubertal).	If a new *in vitro* mechanistic assay is conducted, note that a negative result does not mean that the test material has no thyroid activity. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
J	–	+	+	The test chemical is probably a thyroid (ant)agonist without activity in amphibians, although it is possible that *Xenopus laevis* responds atypically in this case.	Some regulatory authorities may conclude that no further evidence is required, but it might be desirable to conduct a LAGDA with a species other than *X. laevis* (none have been validated at present) if the existing data are sufficiently persuasive.	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is a thyroid (ant)agonist.
K	–	+	–	The test chemical is probably a thyroid (ant)agonist without activity in amphibians or other taxa, although it is possible that *Xenopus laevis* responds atypically in this case.	If there is no activity in amphibian or mammals, further evidence is probably not needed.	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is a thyroid (ant)agonist.
L	–	+	Eq/0	The test chemical is probably a thyroid (ant)agonist without activity in amphibians, although it is possible that *Xenopus laevis* responds atypically in this case.	Some regulatory authorities may conclude that no further evidence is required, but if mammalian data are absent, it might be desirable to conduct a thyroid-responsive rodent screen (e.g. rat pubertal).	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is a thyroid (ant)agonist. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

Scenarios	Result of OECD TG 231 assay (AMA)	Existing results		Possible conclusions: 1) Indicators of endocrine activity and apical endpoints positive 2) Indicators of endocrine activity positive 3) Apical endpoint positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
M	–	–	+	The test chemical is probably without thyroid activity in amphibians, although it is possible that *Xenopus laevis* responds atypically in this case.	Some regulatory authorities may conclude that no further evidence is required, but the positive existing *in vivo* data suggest that it might be helpful to perform a LAGDA with a species other than *X. laevis*.	The lack of *in vitro* thyroid activity is not evidence against any thyroid activity, due to the limited nature of current *in vitro* thyroid screens.
N	–	–	–	The test chemical is probably without thyroid activity in amphibians or other taxa.	No further action is necessary.	–
O	–	–	Eq/0	The test chemical is probably without thyroid activity in amphibians.	Some regulatory authorities may conclude that no further evidence is required, but if mammalian data are absent, it might be desirable to conduct a thyroid-responsive rodent screen (e.g. rat pubertal).	The lack of *in vitro* thyroid activity is not evidence against any thyroid activity, due to the limited nature of current *in vitro* thyroid screens. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
P	–	Eq/0	+	The test chemical is probably without thyroid activity in amphibians, although it is possible that *Xenopus laevis* responds atypically in this case.	Some regulatory authorities may conclude that no further evidence is required, but the positive existing *in vivo* data suggest that it might be helpful to perform a LAGDA with a species other than *X. laevis* (none have been validated at present). Also, if clear *in vitro* mechanistic data are missing, it might be desirable to obtain some.	If a new *in vitro* mechanistic assay is conducted, note that a negative response does not mean that the test material has no thyroid activity. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
Q	–	Eq/0	–	The test chemical is probably without thyroid activity in amphibians or other taxa.	No further action is necessary.	It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
R	–	Eq/0	Eq/0	The test chemical is probably without thyroid activity in amphibians.	Some regulatory authorities may conclude that no further evidence is required, but if mammalian data are absent, it might be desirable to conduct a thyroid-responsive rodent screen (e.g. rat pubertal).	It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

C.2.4. *Potamopyrgus antipodarum* Reproduction Test (OECD TG 242)

Status: Assay validated by the OECD.

328. Modality detected/endpoints: This medium-term reproduction *in vivo* assay with the parthenogenetic female mudsnail *Potamopyrgus antipodarum*, a prosobranch mollusc, is expected to be responsive *inter alia* to retinoid X receptor (RXR) (ant)agonists which can interfere with reproduction in molluscs. It may also respond to certain vertebrate steroid agonists (e.g. estradiol) (Duft et al., 2007), but is not recommended for use in this regard as it was not validated with such chemicals except for the androgen agonist trenbolone (to which it did not respond). It exposes the test organisms for less than a whole generation. It is important to note, however, that none of the endpoints in this apical test are specifically responsive to endocrine-active chemicals, and the assay will give positive results with many other substances. The lack of mechanistic assays for endocrine activity in molluscs will prevent firm conclusions about whether test chemicals are endocrine disrupters (EDs) in this taxon, but the data from the test may nevertheless be of value for classification and hazard identification/characterisation.

Background to the assay

329. This assay is run with the parthenogenetic mudsnail *Potamopyrgus antipodarum*. Adult females are exposed to a range of dilutions of the test chemical for 28 days, after which the numbers of ovoviviparous embryos in the brood pouch are measured. Mortality of the adults is also recorded but does not constitute an endpoint likely to be sensitive to endocrine-active chemicals.

When/why the assay may be used

330. Although OECD TG 242 could, in principle, be used at any stage in the hazard assessment process, the most likely use scenario will be when there are already some *in vitro* data available about the possible endocrine activity of a chemical. However, many chemicals with non-endocrine action will also give positive responses in OECD TG 242.

331. The precise modes of action (MOA) of endocrine-active chemicals in molluscs are unknown, and even the MOA of the well-known mollusc ED tributyltin (TBT) is not fully understood, although it appears to act at least partly via the RXR receptor. OECD TG 242 should therefore not be deployed as a primary screen for endocrine activity, because of its lack of specificity. Furthermore, it should be noted that there are no standardised *in vitro* screens for RXR agonists, although some are described in the scientific literature (e.g. Li, Ma and Wang, 2008). As *P. antipodarum* reproduces parthenogenetically, endocrine-active chemicals may produce different responses in sexually reproducing species.

332. In order to provide information relevant for assessing whether or not a chemical may fulfil the WHO/IPCS (2002) definition of an ED, the study design has to be sufficiently robust to demonstrate the presence or absence of effects. In the dose selection, the

investigator should also consider and ensure that data generated are adequate to fulfil the regulatory requirement across OECD countries as appropriate (e.g. hazard and risk assessment and labelling, ED assessment, etc.). The top dose or concentration should be sufficiently high to give clear systemic (i.e. non endocrine-specific) toxicity in order to ensure that a wide range of exposures (high to low) is tested. However, endocrine effects observed solely in the presence of clear systemic toxicity should be interpreted with caution and may be disregarded when sufficiently justified to be caused by secondary effects which are unlikely to be due to endocrine activity. The reason for this advice is a concern that some endocrine active substance (EAS) sensitive assays are being run at doses/concentrations of EASs that are too low to trigger direct impacts on the endocrine system. This guidance document is not the place to address this issue directly, but it should be considered when EAS-sensitive test guidelines (TGs) are revised in the future. In addition, the number and spacing of dose/concentration levels should also be adequate to fulfil the objectives of the study (e.g. to demonstrate dose response relationships if this is required).

Existing data to be considered

333. Existing information on endocrine-related effects from other molluscs should also be considered before deployment of OECD TG 242, but no short-term screens with this phylum have been internationally standardised. Existing data available might also include one or more of a range of *in silico* or *in vitro* results which suggest that endocrine disruption may occur *in vivo* (but note the limitations of this approach, as indicated above). Such indicators of possible activity might include quantitative structure activity relationship (QSAR) predictions, "read-across" from *in vivo* results obtained with structurally related chemicals or positive results from an *in vitro* screen.

Scenarios: Positive and negative results combined with existing data

334. The scenarios (A to R) presented in Table C.2.4 represent all the possibilities of positive or negative results in combination with the presence or absence of existing data. The action taken will also depend on the regulatory environment, but the considerations given here are generally science based. Wherever possible, the recommended "next step which could be taken" avoids unnecessary animal testing. However, sometimes conducting an animal test will be indicated and then the relevance of species, strain and exposure route should always be considered. Further considerations specific to each scenario are given in the table.

335. Positive results obtained with OECD TG 242 (Table C.2.4, Scenarios A-I) result in the conclusion that the test chemical has adverse apical effects, at least in parthenogenetic molluscs, but these are not necessarily caused by endocrine activity. However, although a positive response of OECD TG 242 indicates that the chemical has adverse effects in parthenogenetic molluscs, it should be noted that many mollusc species such as *Lymnaea stagnalis* have a sexual reproductive strategy and so may respond differently to *P. antipodarum*. Therefore, if countries need further evidence concerning growth and sexual development, etc. in this phylum, a *Lymnaea stagnalis* Reproduction Test (OECD TG 243) would be able to provide information on adverse effects in such mollusc groups. In other words, in order to strengthen weight of evidence, a positive result in OECD TG 242 could be followed by TG 243 (Level 4). Existing data suggesting endocrine-specific activity (e.g. positive *in vitro* data, or positive *in vivo* data from other species) will strengthen the case for additional testing still further.

336. The situation in which OECD TG 242 gives a negative result (Table C.2.4, Scenarios J-R) needs careful consideration of any existing data. If these data suggest that the chemical is endocrine active both *in vitro* and *in vivo* (Scenario J), then it is possible that OECD TG 242 is simply insufficiently sensitive.

337. If OECD TG 242 and existing *in vivo* data are all negative, but *in vitro* data reveal some endocrine activity (Scenario K), the probability is that the test chemical is not sufficiently potent to produce endocrine disruption *in vivo* in molluscs, or it may be rapidly metabolised. In such a situation, further testing is probably not necessary.

338. On the other hand, if OECD TG 242 and the *in vitro* tests are negative (Scenario M), but there are positive existing *in vivo* data, the nature of those existing data should be considered. Unless the existing data are from another mollusc, the chemical is possibly not endocrine active in molluscs, but it may be more potent in sexually reproducing species (e.g. *L. stagnalis*) or life stages that have not been tested. In this situation, the existing *in vivo* data should be used to guide decisions about whether to conduct any further testing (e.g. with OECD TG 243).

339. Finally, a negative OECD TG 242, set against a background of negative *in vitro* and *in vivo* data (Scenario N), suggests that the test chemical is probably not endocrine active *in vitro* or *in vivo*, and further action is unnecessary.

340. In each of the above scenarios, it is possible that existing data will be equivocal, or there may be no existing data. This will weaken the conclusions which can be drawn about a negative OECD TG 242, and this is reflected in Table C.2.4. However, a lack of mechanistic data on endocrine activity should ideally be rectified before any further *in vivo* testing is finally conducted, although as indicated above, *in vitro* endocrine screens for molluscs have not yet been internationally standardised. On the other hand, if OECD TG 242 is positive, further *in vivo* testing would generally be needed to quantify any adverse effects in other molluscs, even if all existing data are equivocal, or if there are no existing data. Again, however, it may be useful to obtain some mechanistic information before conducting further *in vivo* testing. Under some circumstances, two opposite modes of simultaneous action could, depending on dose, lead to a minimisation or abolition of adverse effects, while in others two different MOA could potentially reinforce effects on the OECD TG 242 endpoint. If multiple MOA are suspected, either from the existing results or based on QSAR/read-across/integrated approaches, this situation should be investigated further if needed for regulatory decision making.

341. The scenario in which the results of OECD TG 242 are themselves equivocal has not been dealt with in Table C.2.4, for reasons of brevity. In this context, an equivocal result might be an inconsistent concentration-response (e.g. no effect at a high concentration but effects at a lower concentration, although note that such responses to ED are sometimes repeatably observed *in vivo*), or a result which borders on statistical significance. Without knowing the exact circumstances, reliable advice cannot be given, but the opinions of an experienced ecotoxicologist should be sought. Clearly, however, such equivocal results do not necessarily rule out the existence of *in vivo* endocrine activity. If possible reasons for false negatives are suspected, OECD TG 242 could be repeated (e.g. conduct it at lower concentrations which avoid systemic toxicity).

342. In summary, positive results in OECD TG 242 indicate that a chemical has adverse effects in molluscs which may or may not be via endocrine activity. This may need to be followed up with further apical testing with sexually reproducing molluscs such as *L. stagnalis*. Negative results in OECD TG 242 do not necessarily mean that the chemical is

not a potential ED – a judgement about the endocrine disruption potential in other arthropods will have to be made based on a weight of evidence evaluation of existing *in vitro* and *in vivo* data.

References

Duft, M. et al. (2007), "Prosobranch snails as test organisms for the assessment of endocrine-active chemicals: An overview and a guideline proposal for a reproduction test with the freshwater mudsnail *Potamopyrgus antipodarum*", *Ecotoxicology*, Vol. 16/1, pp. 169-182, https://doi.org/10.1007/s10646-006-0106-0.

Li, J., M. Ma and Z. Wang (2008), "A two-hybrid yeast assay to quantify the effects of xenobiotics on retinoid X receptor-mediated gene expression", *Toxicology Letters*, Vol. 176/3, pp. 198-206, https://doi.org/10.1016/j.toxlet.2007.11.006.

WHO/IPCS (2002), "Global assessment of the state-of-the-science of endocrine disrupters", Damstra, T. et al. (eds.) WHO/PCS/EDC/02.2, World Health Organization, Geneva, www.who.int/ipcs/publications/new_issues/endocrine_disruptors/en.

Table C.2.4. ***Potamopyrgus antipodarum* Reproduction Test (OECD TG 242): Guidance for scenarios of combinations of results with existing data**

This table represents possible conclusions to be drawn from assay data, and a next step which could be taken if further evidence is required about possible endocrine disrupting properties and/or effects. The guidance offered is not meant to be prescriptive, but provides science-based considerations. It encourages the use of all available data and expert judgement in a weight of evidence approach. Regional and national interpretation of results and "next steps" may vary.

The conclusions are grouped into a series of scenarios (A-R), each scenario representing a different combination of assay results, existing *in vitro* data and existing *in vivo* data. The symbol "+" indicates that the data in question represent a positive result, «-" indicates a negative result, and "Eq/0" indicates that the data are either equivocal or are not available.

Existing results: * "Mechanism (*in vitro* mechanistic data)" assumes that mechanistic data are available from available from suitable assays. Some assays concerning mechanisms of disruption in molluscs (e.g. retinoid X receptor [RXR] agonism) may be available, but they have not yet been internationally standardised. In practice, data from few if any assays may be available and therefore this must be taken into account when deciding on the "next step".

Existing results: ** "Effects (*in vivo* effects of concern)" assumes effects have been observed in other *in vivo* screens/tests which give rise to concern that the test chemical may be an endocrine disrupter.

Scenarios	Result of OECD TG 242	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
A	+	+	+	Strong evidence for adverse *in vivo* effects in parthenogenetic molluscs, possibly but not necessarily caused by endocrine disruption, plus possible endocrine effects in other molluscs.	It would be desirable (if not already conducted) to perform an apical test with a sexually reproducing mollusc (e.g. the *Lymnaea stagnalis* Reproduction Test – OECD TG 243).	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is endocrine active.
B	+	+	–	Strong evidence for adverse *in vivo* effects in parthenogenetic molluscs, possibly but not necessarily caused by endocrine disruption.	It would be desirable (if not already conducted) to perform an apical test with a sexually reproducing mollusc (e.g. the *Lymnaea stagnalis* Reproduction Test – OECD TG 243).	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is endocrine active.
C	+	+	Eq/0	Strong evidence for adverse *in vivo* effects in parthenogenetic molluscs, possibly but not necessarily caused by endocrine disruption.	It would be desirable (if not already conducted) to perform an apical test with a sexually reproducing mollusc (e.g. the *Lymnaea stagnalis* Reproduction Test – OECD TG 243).	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is endocrine active. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple modes of action (MOA). If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
D	+	–	+	Strong evidence for adverse *in vivo* effects in parthenogenetic molluscs, possibly but not necessarily caused by endocrine disruption, plus possible endocrine effects in other molluscs.	It would be desirable (if not already conducted) to perform an apical test with a sexually reproducing mollusc (e.g. the *Lymnaea stagnalis* Reproduction Test – OECD TG 243).	The lack of *in vitro* evidence of endocrine activity is not evidence against any such activity, due to the limited nature of current *in vitro* screens.
E	+	–	–	Strong evidence for adverse *in vivo* effects in parthenogenetic molluscs, possibly but not necessarily caused by endocrine disruption.	It would be desirable (if not already conducted) to perform an apical test with a sexually reproducing mollusc (e.g. the *Lymnaea stagnalis* Reproduction Test – OECD TG 243).	The lack of *in vitro* evidence of endocrine activity is not evidence against any such activity, due to the limited nature of current *in vitro* screens.
F	+	–	Eq/0	Strong evidence for adverse *in vivo* effects in parthenogenetic molluscs, possibly but not necessarily caused by endocrine disruption.	It would be desirable (if not already conducted) to perform an apical test with a sexually reproducing mollusc (e.g. the *Lymnaea stagnalis* Reproduction Test – OECD TG 243).	The lack of *in vitro* evidence of endocrine activity is not evidence against any such activity, due to the limited nature of current *in vitro* screens. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

Scenarios	Result of OECD TG 242	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
G	+	Eq/0	+	Strong evidence for adverse *in vivo* effects in parthenogenetic molluscs, possibly but not necessarily caused by endocrine disruption, plus possible endocrine effects in other molluscs.	Given the absence or equivocal nature of existing *in vitro* data, it would be desirable to obtain further *in vitro* data on endocrine activity if possible. It might also be sensible to conduct a *Lymnaea stagnalis* Reproduction Test (OECD TG 243).	If a new *in vitro* mechanistic assay is conducted, note that a negative result does not mean that the test material has no endocrine activity. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
H	+	Eq/0	–	Strong evidence for adverse *in vivo* effects in parthenogenetic molluscs, possibly but not necessarily caused by endocrine disruption.	Given the absence or equivocal nature of the *in vitro* mechanistic data, it might also be helpful to conduct an *in vitro* screen for endocrine activity. It might also be sensible to conduct a *Lymnaea stagnalis* Reproduction Test (OECD TG 243).	If a new *in vitro* mechanistic assay is conducted, note that a negative result does not mean that the test material has no endocrine activity. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
I	+	Eq/0	Eq/0	Strong evidence for adverse *in vivo* effects in parthenogenetic molluscs, possibly but not necessarily caused by endocrine disruption.	Given the absence or equivocal nature of the *in vitro* mechanistic data, it might also be helpful to conduct an *in vitro* screen for endocrine activity. It might also be sensible to conduct a *Lymnaea stagnalis* Reproduction Test (OECD TG 243).	If a new *in vitro* mechanistic assay is conducted, note that a negative result does not mean that the test material has no endocrine activity. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
J	–	+	+	The test chemical is probably endocrine active without adverse effects in parthenogenetic molluscs, although it is possible that *P. antipodarum* responds atypically in this case.	Some regulatory authorities may conclude that no further evidence is required. However, it might be desirable to obtain data from a *Lymnaea stagnalis* Reproduction Test (OECD TG 243) if these are not already available.	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is endocrine active.
K	–	+	–	The test chemical is probably endocrine active without adverse effects in molluscs, although it is possible that *P. antipodarum* responds atypically in this case.	If there is no activity in parthenogenetic or sexually reproducing molluscs, further evidence is probably not needed.	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is endocrine active.

Scenarios	Result of OECD TG 242	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
L	–	+	Eq/0	The test chemical is probably endocrine active without adverse effects in molluscs, although it is possible that *P. antipodarum* responds atypically in this case.	Some regulatory authorities may conclude that no further evidence is required, but if data from a sexually reproducing mollusc species are absent, it might be desirable to conduct a *Lymnaea stagnalis* Reproduction Test (OECD TG 243).	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is endocrine active. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
M	–	–	+	The test chemical is probably endocrine active without adverse effects in parthenogenetic molluscs, although it is possible that *P. antipodarum* responds atypically in this case.	Some regulatory authorities may conclude that no further evidence is required. However, it might be desirable to obtain data from a *Lymnaea stagnalis* Reproduction Test (OECD TG 243).	The lack of *in vitro* endocrine activity is not evidence against any such activity, due to the limited nature of current *in vitro* screens.
N	–	–	–	The test chemical is probably without activity in molluscs.	No further action is necessary.	–
O	–	–	Eq/0	The test chemical is probably without activity in molluscs.	Some regulatory authorities may conclude that no further evidence is required. However, it might be desirable to obtain data from a sexually reproducing mollusc (e.g. the *Lymnaea stagnalis* Reproduction Test – OECD TG 243).	The lack of *in vitro* endocrine activity is not evidence against any such activity, due to the limited nature of current *in vitro* endocrine screens. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
P	–	Eq/0	+	The test chemical is probably without activity in parthenogenetic molluscs, although it is possible that *P. antipodarum* responds atypically in this case.	Some regulatory authorities may conclude that no further evidence is required. Also, if clear *in vitro* mechanistic data are missing, it might be desirable to obtain some.	If a new *in vitro* mechanistic assay is conducted, note that a negative result does not mean that the test material has no endocrine activity. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
Q	–	Eq/0	–	The test chemical is probably without endocrine activity in molluscs.	No further action is necessary.	It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

Scenarios	Result of OECD TG 242	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
R	–	Eq/0	Eq/0	The test chemical is probably without endocrine activity in molluscs.	Some regulatory authorities may conclude that no further evidence is required. However, it might be desirable to obtain data from a *Lymnaea stagnalis* Reproduction Test (OECD TG 243) if these are not already available.	It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

C.2.5. *Lymnaea stagnalis* Reproduction Test (OECD TG 243)

Status: Assay validated by the OECD.

343. Modality detected/endpoints: This medium-term reproduction *in vivo* assay with the sexually reproducing pulmonate pond snail *Lymnaea stagnalis* is expected to be responsive to endocrine disrupters of reproduction in molluscs including *inter alia* to retinoid X receptor (RXR) (ant)agonists. It exposes the test organisms for less than a whole generation. It is important to note, however, that none of the endpoints in this apical test are specifically responsive to endocrine-active chemicals, and the assay will give positive results with many other substances. The lack of mechanistic assays for endocrine activity in molluscs will prevent firm conclusions about whether test chemicals are endocrine disrupters (EDs) in this taxon, but the data from the test may nevertheless be of value for classification and hazard identification/characterisation.

Background to the assay

344. This assay is run with the sexually reproducing pulmonate pond snail *Lymnaea stagnalis*. Adults are exposed to a range of dilutions of the test chemical for 28 days, during which the cumulative numbers of egg clutches per snail per day are recorded. Numbers of eggs per clutch can also be measured. The assay therefore primarily measures fecundity. Mortality of the adults is also recorded, but does not constitute an endpoint likely to be sensitive to endocrine-active chemicals.

When/why the assay may be used

345. Although OECD TG 243 could, in principle, be used at any stage in the hazard assessment process, the most likely use scenario will be when there are already some *in vitro* data available about the possible endocrine activity of a chemical. However, many chemicals with non-endocrine action will also give positive responses in OECD TG 243.

346. The precise modes of action (MOA) of endocrine-active chemicals in molluscs are unknown, and even the MOA of the well-known mollusc ED tributyltin (TBT) is not fully understood, although it appears to act at least partly via the RXR receptor. OECD TG 243 should therefore not be deployed as a primary screen for endocrine activity, because of its lack of specificity. Furthermore, it should be noted that there are no standardised *in vitro* screens for RXR agonists, although some are described in the scientific literature (e.g. Li, Ma and Wang, 2008). As *L. stagnalis* reproduces sexually, endocrine-active chemicals may produce different responses in parthenogenetic species.

347. In order to provide information relevant for assessing whether or not a chemical may fulfil the WHO/IPCS (2002) definition of an ED, the study design has to be sufficiently robust to demonstrate the presence or absence of effects. In the dose selection, the investigator should also consider and ensure that data generated are adequate to fulfil the regulatory requirement across OECD countries as appropriate (e.g. hazard and risk assessment and labelling, ED assessment, etc.). The top dose or concentration should be

sufficiently high to give clear systemic (i.e. non endocrine-specific) toxicity in order to ensure that a wide range of exposures (high to low) is tested. However, endocrine effects observed solely in the presence of clear systemic toxicity should be interpreted with caution and may be disregarded when sufficiently justified to be caused by secondary effects which are unlikely to be due to endocrine activity. The reason for this advice is a concern that some endocrine active substance (EAS) sensitive assays are being run at doses/concentrations of EASs that are too low to trigger direct impacts on the endocrine system. This guidance document is not the place to address this issue directly, but it should be considered when EAS-sensitive test guidelines (TGs) are revised in the future. In addition, the number and spacing of dose/concentration levels should also be adequate to fulfil the objectives of the study (e.g. to demonstrate dose response relationships if this is required).

Existing data to be considered

348. Existing information on endocrine-related effects from other molluscs should also be considered before deployment of OECD TG 243, but no short-term screens with this phylum have been internationally standardised. Existing data available might also include one or more of a range of *in silico* or *in vitro* results which suggest that endocrine disruption may occur *in vivo* (but note the limitations of this approach, as indicated above). Such indicators of possible activity might include quantitative structure activity relationship (QSAR) predictions, "read-across" from *in vivo* results obtained with structurally related chemicals or positive results from an *in vitro* screen.

Scenarios: Positive and negative results combined with existing data

349. The scenarios (A to R) presented in Table C.2.5 represent all the possibilities of positive or negative results in combination with the presence or absence of existing data. The action taken will also depend on the regulatory environment, but the considerations given here are generally science based. Wherever possible, the recommended "next step which could be taken" avoids unnecessary animal testing. However, sometimes conducting an animal test will be indicated and then the relevance of species, strain and exposure route should always be considered. Further considerations specific to each scenario are given in the table.

350. Positive results obtained with OECD TG 243 (Table C.2.5, Scenarios A-I) result in the conclusion that the test chemical has adverse apical effects, at least in sexually reproducing molluscs, but these are not necessarily caused by endocrine activity. However, although a positive response of the OECD TG 243 indicates that the chemical has adverse effects in sexually reproducing molluscs, it should be noted that many mollusc species such *Pomatopyrgus antipodarum* have a parthenogenetic reproductive strategy and so may respond differently to *L. stagnalis*. Therefore, if countries need further evidence concerning growth and sexual development etc. in this phylum, a *Pomatopyrgus antipodarum* Reproduction Test – OECD TG 242 would be able to provide information on adverse effects in such mollusc groups. In other words, in order to strengthen weight of evidence, a positive result in OECD TG 243 could be followed by TG 242 (Level 4). Existing data suggesting endocrine-specific activity (e.g. positive *in vitro* data, or positive *in vivo* data from other species) will strengthen the case for additional testing still further.

351. The situation in which OECD TG 243 gives a negative result (Table C.2.5, Scenarios J-R) needs careful consideration of any existing data. If these data suggest that

the chemical is endocrine active both *in vitro* and *in vivo* (Scenario J), then it is possible that OECD TG 243 is simply insufficiently sensitive.

352. If OECD TG 243 and existing *in vivo* data are all negative, but *in vitro* data reveal some endocrine activity (Scenario K), the probability is that the test chemical is not sufficiently potent to produce endocrine disruption *in vivo* in molluscs, or it may be rapidly metabolised. In such a situation, further testing is probably not necessary.

353. On the other hand, if OECD TG 243 and the *in vitro* tests are negative (Scenario M), but there are positive existing *in vivo* data, the nature of those existing data should be considered. Unless the existing data are from another mollusc, the chemical is possibly not endocrine active in molluscs, but it may be more potent in parthenogenetic species (e.g. *P. antipodarum*) or life stages that have not been tested. In this situation, the existing *in vivo* data should be used to guide decisions about whether to conduct any further testing (e.g. with OECD TG 242).

354. Finally, a negative OECD TG 243, set against a background of negative *in vitro* and *in vivo* data (Scenario N), suggests that the test chemical is probably not endocrine active *in vitro* or *in vivo*, and further action is unnecessary.

355. In each of the above scenarios, it is possible that existing data will be equivocal, or there may be no existing data. This will weaken the conclusions which can be drawn about a negative OECD TG 243, and this is reflected in Table C.2.5. However, a lack of mechanistic data on endocrine activity should ideally be rectified before any further *in vivo* testing is finally conducted, although as indicated above, *in vitro* endocrine screens for molluscs have not yet been internationally standardised. On the other hand, if OECD TG 243 is positive, further *in vivo* testing would generally be needed to quantify any adverse effects in other molluscs, even if all existing data are equivocal, or if there are no existing data. Again, however, it may be useful to obtain some mechanistic information before conducting further *in vivo* testing. Under some circumstances, two opposite modes of simultaneous action could, depending on dose, lead to a minimisation or abolition of adverse effects, while in others two different MOA could potentially reinforce effects on the OECD TG 243 endpoint. If multiple MOA are suspected, either from the existing results or based on QSAR/read-across/integrated approaches, this situation should be investigated further if needed for regulatory decision making.

356. The scenario in which the results of OECD TG 243 are themselves equivocal has not been dealt with in Table C.2.5, for reasons of brevity. In this context, an equivocal result might be an inconsistent concentration-response (e.g. no effect at a high concentration but effects at a lower concentration), or a result which borders on statistical significance. Without knowing the exact circumstances, reliable advice cannot be given, but the opinions of an experienced ecotoxicologist should be sought. Clearly, however, such equivocal results do not necessarily rule out the existence of *in vivo* endocrine activity. If possible reasons for false negatives are suspected, OECD TG 243 could be repeated (e.g. conduct it at lower concentrations which avoid systemic toxicity).

357. In summary, positive results in OECD TG 243 indicate that a chemical has adverse effects in molluscs which may or may not be via endocrine activity. This may need to be followed up with further apical testing with parthenogenetic molluscs such as *P. antipodarum*. Negative results in OECD TG 243 do not necessarily mean that the chemical is not a potential ED – a judgement about the endocrine disruption potential in other arthropods will have to be made based on a weight of evidence evaluation of existing *in vitro* and *in vivo* data.

References

Li, J., M. Ma and Z. Wang (2008), "A two-hybrid yeast assay to quantify the effects of xenobiotics on retinoid X receptor-mediated gene expression", *Toxicology Letters*, Vol. 176/3, pp. 198-206, https://doi.org/10.1016/j.toxlet.2007.11.006.

WHO/IPCS (2002), "Global assessment of the state-of-the-science of endocrine disrupters", Damstra, T. et al. (eds.) WHO/PCS/EDC/02.2, World Health Organization, Geneva, www.who.int/ipcs/publications/new_issues/endocrine_disruptors/en.

Table C.2.5. ***Lymnaea stagnalis* Reproduction Test (OECD TG 243): Guidance for scenarios of combinations of results with existing data**

This table represents possible conclusions to be drawn from assay data, and a next step which could be taken if further evidence is required about possible endocrine disrupting properties and/or effects. The guidance offered is not meant to be prescriptive, but provides science-based considerations. It encourages the use of all available data and expert judgement in a weight of evidence approach. Regional and national interpretation of results and "next steps" may vary.

The conclusions are grouped into a series of scenarios (A-R), each scenario representing a different combination of assay results, existing *in vitro* data and existing *in vivo* data. The symbol "+" indicates that the data in question represent a positive result, «-" indicates a negative result, and "Eq/0" indicates that the data are either equivocal or are not available.

Existing results: * "Mechanism (*in vitro* mechanistic data)" assumes that mechanistic data are available from available from suitable assays. Some assays concerning mechanisms of disruption in molluscs (e.g. RXR agonism) may be available, but have not yet been internationally standardised. In practice, data from few if any assays may be available and therefore this must be taken into account when deciding on the "next step".

Existing results: ** "Effects (*in vivo* effects of concern)" assumes effects have been observed in other *in vivo* screens/tests which give rise to concern that the test chemical may be an endocrine disrupter.

Scenarios	Result of OECD TG 243	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
A	+	+	+	Strong evidence for adverse *in vivo* effects in sexually reproducing molluscs, possibly but not necessarily caused by endocrine disruption, plus possible endocrine effects in other molluscs.	It would be desirable (if not already conducted) to perform an apical test with a parthenogenetic mollusc (e.g. the *Pomatopyrgus antipodarum* Reproduction Test – OECD TG 242).	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is endocrine active.
B	+	+	–	Strong evidence for adverse *in vivo* effects in sexually reproducing molluscs, possibly but not necessarily caused by endocrine disruption.	It would be desirable (if not already conducted) to perform an apical test with a parthenogenetic mollusc (e.g. the *Pomatopyrgus antipodarum* Reproduction Test – OECD TG 242).	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is endocrine active.
C	+	+	Eq/0	Strong evidence for adverse *in vivo* effects in sexually reproducing molluscs, possibly but not necessarily caused by endocrine disruption.	It would be desirable (if not already conducted) to perform an apical test with a parthenogenetic mollusc (e.g. the *Pomatopyrgus antipodarum* Reproduction Test – OECD TG 242).	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is endocrine active. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple modes of action (MOA). If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
D	+	–	+	Strong evidence for adverse *in vivo* effects in sexually reproducing molluscs, possibly but not necessarily caused by endocrine disruption, plus possible endocrine effects in other molluscs.	It would be desirable (if not already conducted) to perform an apical test with a parthenogenetic mollusc (e.g. the *Pomatopyrgus antipodarum* Reproduction Test – OECD TG 242).	The lack of *in vitro* evidence of endocrine activity is not evidence against any such activity, due to the limited nature of current *in vitro* screens.
E	+	–	–	Strong evidence for adverse *in vivo* effects in sexually reproducing molluscs, possibly but not necessarily caused by endocrine disruption.	It would be desirable (if not already conducted) to perform an apical test with a parthenogenetic mollusc (e.g. the *Pomatopyrgus antipodarum* Reproduction Test – OECD TG 242).	The lack of *in vitro* evidence of endocrine activity is not evidence against any such activity, due to the limited nature of current *in vitro* screens.
F	+	–	Eq/0	Strong evidence for adverse *in vivo* effects in sexually reproducing molluscs, possibly but not necessarily caused by endocrine disruption.	It would be desirable (if not already conducted) to perform an apical test with a parthenogenetic mollusc (e.g. the *Pomatopyrgus antipodarum* Reproduction Test – OECD TG 242).	The lack of *in vitro* evidence of endocrine activity is not evidence against any such activity, due to the limited nature of current *in vitro* screens. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

Scenarios	Result of OECD TG 243	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
G	+	Eq/0	+	Strong evidence for adverse *in vivo* effects in sexually reproducing molluscs, possibly but not necessarily caused by endocrine disruption, plus possible endocrine effects in other molluscs.	Given the absence or equivocal nature of existing *in vitro* data, it would be desirable to obtain further *in vitro* data on endocrine activity if possible. It might also be sensible to conduct the *Pomatopyrgus antipodarum* Reproduction Test (OECD TG 242).	If a new *in vitro* mechanistic assay is conducted, note that a negative result does not mean that the test material has no endocrine activity. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
H	+	Eq/0	–	Strong evidence for adverse *in vivo* effects in sexually reproducing molluscs, possibly but not necessarily caused by endocrine disruption.	Given the absence or equivocal nature of the *in vitro* mechanistic data, it might also be helpful to conduct an *in vitro* screen for endocrine activity. It might also be sensible to conduct the *Pomatopyrgus antipodarum* Reproduction Test (OECD TG 242).	If a new *in vitro* mechanistic assay is conducted, note that a negative result does not mean that the test material has no endocrine activity. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
I	+	Eq/0	Eq/0	Strong evidence for adverse *in vivo* effects in sexually reproducing molluscs, possibly but not necessarily caused by endocrine disruption.	Given the absence or equivocal nature of the *in vitro* mechanistic data, it might also be helpful to conduct an *in vitro* screen for endocrine activity. It might also be sensible to conduct the *Pomatopyrgus antipodarum* Reproduction Test (OECD TG 242).	If a new *in vitro* mechanistic assay is conducted, note that a negative result does not mean that the test material has no endocrine activity. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
J	–	+	+	The test chemical is probably endocrine active without adverse effects in sexually reproducing molluscs, although it is possible that *L. stagnalis* responds atypically in this case.	Some regulatory authorities may conclude that no further evidence is required. However, it might be desirable to obtain data from the *Pomatopyrgus antipodarum* Reproduction Test (OECD TG 242) if these are not already available.	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is endocrine active.

Scenarios	Result of OECD TG 243	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
K	–	+	–	The test chemical is probably endocrine active without adverse effects in molluscs, although it is possible that *Lymnaea stagnalis* responds atypically in this case.	If there is no activity in parthenogenetic or sexually reproducing molluscs, further evidence is probably not needed.	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is endocrine active.
L	–	+	Eq/0	The test chemical is probably endocrine active without adverse effects in molluscs, although it is possible that *L. stagnalis* responds atypically in this case.	Some regulatory authorities may conclude that no further evidence is required, but if data from a parthenogenetic mollusc species are absent, it might be desirable to conduct the *Pomatopyrgus antipodarum* Reproduction Test (OECD TG 242).	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is endocrine active. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
M	–	–	+	The test chemical is probably endocrine active without adverse effects in sexually reproducing molluscs, although it is possible that *L. stagnalis* responds atypically in this case.	Some regulatory authorities may conclude that no further evidence is required. However, it might be desirable to obtain data from the *Pomatopyrgus antipodarum* Reproduction Test (OECD TG 242).	The lack of *in vitro* endocrine activity is not evidence against any such activity, due to the limited nature of current *in vitro* screens.
N	–	–	–	The test chemical is probably without activity in molluscs.	No further action is necessary.	–
O	–	–	Eq/0	The test chemical is probably without activity in molluscs.	Some regulatory authorities may conclude that no further evidence is required. However, it might be desirable to obtain data from the *Pomatopyrgus antipodarum* Reproduction Test (OECD TG 242).	The lack of *in vitro* endocrine activity is not evidence against any such activity, due to the limited nature of current *in vitro* endocrine screens. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

Scenarios	Result of OECD TG 243	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
P	–	Eq/0	+	The test chemical is probably without activity in sexually reproducing molluscs, although it is possible that *L. stagnalis* responds atypically in this case.	Some regulatory authorities may conclude that no further evidence is required. Also, if clear *in vitro* mechanistic data are missing, it might be desirable to obtain some.	If a new *in vitro* mechanistic assay is conducted, note that a negative result does not mean that the test material has no endocrine activity. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
Q	–	Eq/0	–	The test chemical is probably without endocrine activity in molluscs.	No further action is necessary.	It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
R	–	Eq/0	Eq/0	The test chemical is probably without endocrine activity in molluscs.	Some regulatory authorities may conclude that no further evidence is required. However, it might be desirable to obtain data from the *Pomatopyrgus antipodarum* Reproduction Test (OECD TG 242) if these are not already available.	It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

C.2.6. Chironomid Toxicity Test Using Spiked Sediment (OECD TG 218) or Spiked Water (OECD TG 219)

Status: Assay validated by the OECD.

358. Modality detected/endpoints: This medium-term *in vivo* assay with the dipteran insect *Chironomus* spp. is responsive to juvenile hormone (JH) (ant)agonists and ecdysteroid (Ec) (ant)agonists which can interfere with such processes as metamorphosis, moulting and growth (e.g. Hahn, Liess and Schulz [2001]; Taenzler et al. [2007]; Jungmann et al. [2009]; Tassou and Schulz [2009, 2013]). It exposes the test organisms over a single generation. It is important to note, however, that none of the endpoints in this apical test are specifically responsive to JH- or Ec-active chemicals, and the assay will give positive results with many other substances. The lack of internationally validated mechanistic assays for endocrine activity in insects may prevent firm conclusions about whether test chemicals are endocrine disrupters (EDs) in this taxon, although *in vitro* assays for JH and estrogen (E) activity are available in the literature. However, the data from the test may nevertheless be of value for classification and hazard identification/characterisation.

Background to the assay

359. This assay can be run with one of several chironomid species, including *Chironomus riparius*, *C. dilutus* and *C. yoshimatsui*. It can also be operated in one of two formats, with the test chemical spiked either into the ambient water (OECD TG 219) or into the sediment (OECD TG 218), thus allowing sparingly soluble or hydrophobic chemicals to be tested. The test with *C. riparius* and *C. yoshimatsui* takes 20-28 days, while with *C. dilutus* it continues for 28-65 days. The exposure to a range of test concentrations begins with first instar larvae and continues to their fully emerged adulthood. The main endpoints are time to emergence, and emergence itself, but larval survival and growth may also be measured.

360. Available data from a two-generation test with *C. riparius* (Tassou and Schulz, 2009) show that an agonist of the JH pathway will impact emergence rate in a single generation, though the no-observed-effect-concentration (NOEC) becomes lower when considering the second generation. It is expected that chemicals affecting the Ec pathway will also have an effect on emergence.

When/why the assay may be used

361. Although OECD TG 218/219 could, in principle, be used at any stage in the hazard assessment process, the most likely use scenario will be when there are already some *in vitro* data available about the possible JH or E activity of a chemical. Given the significant degree of endocrine system conservation across the arthropods, effects in OECD TG 218/219 may also indicate the possibility of related activity in other arthropods such as crustaceans (cladocera, copepods and decapods). However, many chemicals with non-endocrine action will also give positive responses in OECD TG 218/219.

362. It is not recommended that OECD TG 218/219 be deployed as a primary screen for JH or Ec activity and effects because of its lack of specificity, but it should be noted that there are no standardised *in vitro* screens for JH- or Ec- (ant)agonists, although some are described in the scientific literature (e.g. Dinan et al. [2001]; Smagghe et al. [2003]; Swevers et al. [2003]).

363. In order to provide information relevant for assessing whether or not a chemical may fulfil the WHO/IPCS (2002) definition of an ED, the study design has to be sufficiently robust to demonstrate the presence or absence of effects. In the dose selection, the investigator should also consider and ensure that data generated are adequate to fulfil the regulatory requirement across OECD countries as appropriate (e.g. hazard and risk assessment and labelling, ED assessment, etc.). The top dose or concentration should be sufficiently high to give clear systemic (i.e. non endocrine-specific) toxicity in order to ensure that a wide range of exposures (high to low) is tested. However, endocrine effects observed solely in the presence of clear systemic toxicity should be interpreted with caution and may be disregarded when sufficiently justified to be caused by secondary effects which are unlikely to be due to endocrine activity. The reason for this advice is a concern that some endocrine active substance (EAS) sensitive assays are being run at doses/concentrations of EASs that are too low to trigger direct impacts on the endocrine system. This guidance document is not the place to address this issue directly, but it should be considered when EAS-sensitive test guidelines (TGs) are revised in the future. In addition, the number and spacing of dose/concentration levels should also be adequate to fulfil the objectives of the study (e.g. to demonstrate dose response relationships if this is required).

Existing data to be considered

364. Existing information on endocrine-related effects from other arthropods should also be considered before deployment of OECD TG 218/219, given the commonality of endocrine mechanisms in these taxa. Existing data available might also include one or more of a range of *in silico* or *in vitro* results which suggest that JH or Ec disruption may occur *in vivo* (but note the limitations of this approach, as indicated above). Such indicators of possible JH or Ec activity might include QSAR predictions of JH/Ec activity, "read-across" from *in vivo* results obtained with structurally related chemicals or positive results from an *in vitro* screen for JH/Ec (ant)agonist activity. In addition, *in vivo* data may also be available from one or more short-medium term assays, including the Short-Term Juvenile Hormone Activity Screening Assay (SJHASA), or the *Daphnia magna* Reproduction Test with male neonate option (OECD TG 211).

Scenarios: Positive and negative results combined with existing data

365. The scenarios (A to R) presented in Table C.2.6 represent all the possibilities of positive or negative results in combination with the presence or absence of existing data. The action taken will also depend on the regulatory environment, but the considerations given here are generally science based. Wherever possible, the recommended "next step which could be taken" avoids unnecessary animal testing. However, sometimes conducting an animal test will be indicated and then the relevance of species, strain and exposure route should always be considered. Further considerations specific to each scenario are given in the table.

366. Positive results obtained with OECD TG 218/219 (Table C.2.6, Scenarios A-I) result in the conclusion that the test chemical has adverse apical effects, at least in insects, but these are not necessarily caused by JH or Ec activity. However, although a positive response of OECD TG 218/219 indicates that the chemical has adverse effects in insects, it should be noted that crustacean species such as *Daphnia* have a parthenogenetic reproductive strategy and so may respond differently to *Chironomus*. Therefore, if countries need further evidence concerning growth and sexual development, etc. in this phylum, a *Daphnia* Multigeneration Test (DMGT) and/or a Harpacticoid Copepod Development and Reproduction Test (OECD GD 201) would be able to provide information on adverse effects in other arthropod groups. In other words, in order to strengthen weight of evidence, a positive result in OECD TG 218/219 could be followed by OECD GD 201 (Level 4), and/or the DMGT (Level 5). Existing data suggesting endocrine-specific activity (e.g. positive *in vitro* data, or positive *in vivo* data from other species) will strengthen the case for additional testing still further.

367. The situation in which OECD TG 218/219 gives a negative result (Table C.2.6, Scenarios J-R) needs careful consideration of any existing data. If these data suggest that the chemical is endocrine active both *in vitro* and *in vivo* (Scenario J), then it is possible that OECD TG 218/219 is simply insufficiently sensitive.

368. If OECD TG 218/219 and existing *in vivo* data are all negative, but *in vitro* data reveal some JH or Ec activity (Scenario K), the probability is that the test chemical is not sufficiently potent to produce JH/Ec (ant)agonism *in vivo* in arthropods, or it may be rapidly metabolised. In such a situation, further testing is probably not necessary.

369. On the other hand, if OECD TG 218/219 and the *in vitro* tests are negative (Scenario M), but there are positive existing *in vivo* data, the nature of those existing data should be considered. Unless the existing data are from another insect, the chemical is possibly not a JH or Ec (ant)agonist acting in insects, but it may be more potent in species (e.g. crustaceans) or life stages that have not been tested. In this situation, the existing *in vivo* data should be used to guide decisions about whether to conduct any further testing (e.g. with OECD GD 201).

370. Finally, a negative OECD TG 218/219, set against a background of negative *in vitro* and *in vivo* data (Scenario N), suggests that the test chemical is probably not a JH or Ec (ant)agonist *in vitro* or *in vivo*, and further action is unnecessary.

371. In each of the above scenarios, it is possible that existing data will be equivocal, or there may be no existing data. This will weaken the conclusions which can be drawn about a negative OECD TG 218/219, and this is reflected in Table C.2.6. However, a lack of mechanistic data on JH or Ec activity should ideally be addressed before any further *in vivo* testing is finally conducted, although as indicated above, *in vitro* JH/Ec screens have not yet been internationally standardised. On the other hand, if OECD TG 218/219 is positive, further *in vivo* testing would generally be needed to quantify any adverse effects in crustaceans, even if all existing data are equivocal, or if there are no existing data. Again, however, it may be useful to obtain some mechanistic information before conducting further *in vivo* testing. There is also the possibility that equivocal mechanistic data may be the result of multiple modes of endocrine action. Under some circumstances, two opposite modes of simultaneous action (e.g. JH or Ec agonistic and antagonistic) could, depending on dose, lead to a minimisation or abolition of adverse effects, while in others two different modes of action (MOA) could potentially reinforce effects on the OECD TG 218/219 endpoint. If multiple MOA are suspected, either from the existing results or based on

quantitative structure activity relationship (QSAR)/read-across/integrated approaches, this situation should be investigated further if needed for regulatory decision making.

372. The scenario in which the results of OECD TG 218/219 are themselves equivocal has not been dealt with in Table C.2.6, for reasons of brevity. In this context, an equivocal result might be an inconsistent concentration-response (e.g. no effect at a high concentration but effects at a lower concentration), or a result which borders on statistical significance. Without knowing the exact circumstances, reliable advice cannot be given, but the opinions of an experienced ecotoxicologist should be sought. Clearly, however, such equivocal results do not necessarily rule out the existence of *in vivo* endocrine activity. If possible reasons for false negatives are suspected, OECD TG 218/219 could be repeated (e.g. conduct it at lower concentrations which avoid systemic toxicity).

373. In summary, positive results in OECD TG 218/219 indicate that a chemical has adverse effects in insects which may or may not be via JH or Ec (ant)agonism. This may need to be followed up with further apical testing with crustaceans. Negative results in OECD TG 218/219 do not necessarily mean that the chemical is not a potential ED – a judgement about the endocrine disruption potential in other arthropods will have to be made based on a weight of evidence evaluation of existing *in vitro* and *in vivo* data.

References

Dinan, L. et al. (2001), "Screening of environmental contaminants for ecdysteroid agonist and antagonist activity using the *Drosophila melanogaster* B-II cell *in vitro* assay", *Environmental Toxicology and Chemistry*, Vol. 20/9, pp. 2038-2046, https://doi.org/10.1002/etc.5620200924.

Hahn, T., M. Liess and R. Schulz (2001), "Effects of the hormone mimetic insecticide tebufenozide in *Chironomus riparius* larvae in two different exposure setups", *Ecotoxicology and Environmental Safety*, Vol. 49/2, pp. 171-178, https://doi.org/10.1006/eesa.2001.2055.

Jungmann, D. et al. (2009), "Chronic toxicity of fenoxycarb to the midge *Chironomus riparius* after exposure in sediments of different composition", *Journal of Soils and Sediments*, Vol. 9/2, pp. 94-102, http://dx.doi.org/10.1007/s11368-009-0056-2.

Smagghe, G. et al. (2003), "Cultured mosquito cells *Aedes albopictus* C6/36 (Dip, Culicidae) responsive to 20-hydroxyecdysone and non-steroidal ecdysone antagonists", *Journal of Applied Entomology*, Vol. 127/3, pp. 167-173, https://doi.org/10.1046/j.1439-0418.2003.00727.x.

Swevers, L. et al. (2003), "A high-throughput screening system for fast detection of ecdysteroid mimetic and antagonistic substances using transformed *Bombyx mori*-derived cell lines", *The FASEB Journal*, Vol. 18/1, pp. 134-136, http://dx.doi.org/10.1096/fj.03-0627fje.

Taenzler, V. et al. (2007), "Chironomids: Suitable test organisms for risk assessment investigations on the potential endocrine disrupting properties of pesticides", *Ecotoxicology*, Vol. 16/1, pp. 221-230, https://doi.org/10.1007/s10646-006-0117-x.

Tassou, K.T. and R. Schulz (2013), "Low field-relevant tebufenozide concentrations affect reproduction in *Chironomus riparius* (Diptera: Chironomidae) in a long-term toxicity test", *Environmental Science and Pollution Research*, Vol. 20/6, pp. 3735-3742, http://dx.doi.org/10.1007/s11356-012-1311-4.

Tassou, K.T. and R. Schulz (2009), "Effects of the insect growth regulator pyriproxyfen in a two-generation test with *Chironomus riparius*", *Ecotoxicology and Environmental Safety*, Vol. 72, pp. 1058-1062, https://doi.org/10.1016/j.ecoenv.2009.02.001.

WHO/IPCS (2002), "Global assessment of the state-of-the-science of endocrine disrupters", Damstra, T. et al. (eds.) WHO/PCS/EDC/02.2, World Health Organization, Geneva, www.who.int/ipcs/publications/new_issues/endocrine_disruptors/en.

Table C.2.6. **Chironomid Toxicity Test Using Spiked Sediment (OECD TG 218) or Spiked Water (OECD TG 219): Guidance for scenarios of combinations of results with existing data**

This table represents possible conclusions to be drawn from assay data, and a next step which could be taken if further evidence is required about possible endocrine disrupting properties and/or effects. The guidance offered is not meant to be prescriptive, but provides science-based considerations. It encourages the use of all available data and expert judgement in a weight of evidence approach. Regional and national interpretation of results and "next steps" may vary.

The conclusions are grouped into a series of scenarios (A-R), each scenario representing a different combination of assay results, existing *in vitro* data and existing *in vivo* data. The symbol "+" indicates that the data in question represent a positive result, «-" indicates a negative result, and "Eq/0" indicates that the data are either equivocal or are not available.

Existing results: * "Mechanism (*in vitro* mechanistic data)" assumes that mechanistic data are available from juvenile hormone-(JH) or ecdysteroid (Ec)-based assays. JH or Ec assays concerning mechanisms of disruption may be available, but they have not yet been internationally standardised. In practice, data from all assays may not be available and therefore this must be taken into account when deciding on the "next step".

Existing results: ** "Effects (*in vivo* effects of concern)" assumes effects have been observed in other *in vivo* screens/tests which give rise to concern that the test chemical may be a JH or Ec disrupter.

Scenarios	Result of OECD TG 218/219	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
A	+	+	+	Strong evidence for adverse *in vivo* effects in insects, possibly but not necessarily caused by juvenile hormone (JH) or ecdysteroid (Ec) (ant)agonists, plus possible JH or Ec effects in other arthropods.	It would be desirable (if not already conducted) to perform an apical test with crustaceans (e.g. the Harpacticoid Copepod Development and Reproduction Test – OECD GD 201; and/or the *Daphnia* Multigeneration Test [DMGT]).	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is a JH or Ec (ant)agonist. It should be noted that chironomids reproduce sexually and therefore may respond differently to a parthenogenetic species such as *Daphnia*. It should also be considered whether the full extent of effects in chironomids needs to be followed up with a Chironomid Life-Cycle Toxicity Test (OECD TG 233).
B	+	+	–	Strong evidence for adverse *in vivo* effects in insects, possibly but not necessarily caused by JH or Ec (ant)agonists.	It would be desirable (if not already conducted) to perform an apical test with crustaceans (e.g. the Harpacticoid Copepod Development and Reproduction Test – OECD GD 201; and/or the DMGT).	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is a JH or Ec (ant)agonist. It should be noted that chironomids reproduce sexually and therefore may respond differently to a parthenogenetic species such as *Daphnia*. It should also be considered whether the full extent of effects in chironomids needs to be followed up with a Chironomid Life-Cycle Toxicity Test (OECD TG 233).
C	+	+	Eq/0	Strong evidence for adverse *in vivo* effects in insects, possibly but not necessarily caused by JH or Ec (ant)agonists.	It would be desirable (if not already conducted) to perform an apical test with crustaceans (e.g. the Harpacticoid Copepod Development and Reproduction Test – OECD GD 201; and/or the DMGT).	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is a JH or Ec (ant)agonist. It should be noted that chironomids reproduce sexually and therefore may respond differently to a parthenogenetic species such as *Daphnia*. It should also be considered whether the full extent of effects in chironomids needs to be followed up with a Chironomid Life-Cycle Toxicity Test (OECD TG 233). It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity, or multiple modes of action (MOA). If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
D	+	–	+	Strong evidence for adverse *in vivo* effects in insects, possibly but not necessarily caused by JH or Ec (ant)agonists, plus possible JH or Ec effects in other arthropods.	It would be desirable (if not already conducted) to perform an apical test with crustaceans (e.g. the Harpacticoid Copepod Development and Reproduction Test – OECD GD 201; and/or the DMGT).	The lack of *in vitro* JH or Ec activity is not necessarily evidence against any JH/Ec activity, due to the limited nature of current *in vitro* JH and Ec screens. It should be noted that chironomids reproduce sexually and therefore may respond differently to a parthenogenetic species such as *Daphnia*.

						It should also be considered whether the full extent of effects in chironomids needs to be followed up with a Chironomid Life-Cycle Toxicity Test (OECD TG 233).

Scenarios	Result of OECD TG 218/219	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
E	+	–	–	Some evidence for adverse *in vivo* effects in insects, possibly but not necessarily caused by JH or Ec (ant)agonists.	It would be desirable (if not already conducted) to perform an apical test with crustaceans (e.g. the Harpacticoid Copepod Development and Reproduction Test – OECD GD 201; and/or the DMGT).	The lack of *in vitro* JH or Ec activity is not necessarily evidence against any JH/Ec activity, due to the limited nature of current *in vitro* JH and Ec screens. It should be noted that chironomids reproduce sexually and therefore may respond differently to a parthenogenetic species such as *Daphnia*. It should also be considered whether the full extent of effects in chironomids needs to be followed up with a Chironomid Life-Cycle Toxicity Test (OECD TG 233).
F	+	–	Eq/0	Some evidence for adverse *in vivo* effects in insects, possibly but not necessarily caused by JH or Ec (ant)agonists.	Given the absence or equivocal nature of existing *in vivo* data, it might also be sensible to conduct a JH/Ec-responsive crustacean assay (e.g. the Harpacticoid Copepod Development and Reproduction Test – OECD GD 201) or a JH-responsive DMGT.	The lack of *in vitro* JH/Ec activity is not evidence against any JH/Ec activity, due to the limited nature of current *in vitro* JH/Ec screens. It should be noted that chironomids reproduce sexually and therefore may respond differently to a parthenogenetic species such as *Daphnia*. It should also be considered whether the full extent of effects in chironomids needs to be followed up with a Chironomid Life-Cycle Toxicity Test (OECD TG 233). It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
G	+	Eq/0	+	Strong evidence for adverse *in vivo* effects in insects, possibly but not necessarily caused by JH or Ec (ant)agonists, plus possible JH or E effects in other arthropods.	Given the absence or equivocal nature of existing *in vitro* data, it would be desirable to obtain further *in vitro* data on JH/Ec activity if possible. It might also be sensible to conduct a JH/Ec-responsive crustacean assay (e.g. the Harpacticoid Copepod Development and Reproduction Test – OECD GD 201) or a JH-responsive DMGT.	If a new *in vitro* mechanistic assay is conducted, note that a negative result does not mean that the test material has no JH/Ec activity. It should be noted that chironomids reproduce sexually and therefore may respond differently to a parthenogenetic species such as *Daphnia*. It should also be considered whether the full extent of effects in chironomids needs to be followed up with a Chironomid Life-Cycle Toxicity Test (OECD TG 233). It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

Scenarios	Result of OECD TG 218/219	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
H	+	Eq/0	–	Some evidence for adverse *in vivo* effects in insects, possibly but not necessarily caused by JH or Ec (ant)agonists.	Given the absence or equivocal nature of the *in vitro* mechanistic data, it might also be helpful to conduct an *in vitro* screen for JH/Ec activity. It might also be sensible to conduct a JH/Ec-responsive crustacean assay (e.g. the Harpacticoid Copepod Development and Reproduction Test – OECD GD 201) or a JH-responsive DMGT.	If a new *in vitro* mechanistic assay is conducted, note that a negative result does not mean that the test material has no JH/Ec activity. It should be noted that chironomids reproduce sexually and therefore may respond differently to a parthenogenetic species such as *Daphnia*. It should also be considered whether the full extent of effects in chironomids needs to be followed up with a Chironomid Life-Cycle Toxicity Test (OECD TG 233). It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
I	+	Eq/0	Eq/0	Some evidence for adverse *in vivo* effects in insects, possibly but not necessarily caused by JH or Ec (ant)agonists.	Given the absence or equivocal nature of the *in vitro* mechanistic data, it might also be helpful to conduct an *in vitro* screen for JH/Ec activity. It might also be sensible to conduct a JH/Ec-responsive crustacean assay (e.g. the Harpacticoid Copepod Development and Reproduction Test – OECD GD 201) or a JH-responsive DMGT.	If a new *in vitro* mechanistic assay is conducted, note that a negative result does not mean that the test material has no JH/Ec activity. It should be noted that chironomids reproduce sexually and therefore may respond differently to a parthenogenetic species such as *Daphnia*. It should also be considered whether the full extent of effects in chironomids needs to be followed up with a Chironomid Life-Cycle Toxicity Test (OECD TG 233). It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
J	–	+	+	The test chemical is probably a JH or Ec (ant)agonist without adverse effects in insects, although it is possible that *Chironomus* spp. responds atypically in this case.	Some regulatory authorities may conclude that no further evidence is required. However, it might be desirable to obtain data from crustaceans (e.g. the Harpacticoid Copepod Development and Reproduction Test – OECD GD 201; or the DMGT) if these are not already available.	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is a JH/Ec (ant)agonist. It should be noted that chironomids reproduce sexually and therefore may respond differently to a parthenogenetic species such as *Daphnia*.
K	–	+	–	The test chemical is probably a JH or Ec (ant)agonist without adverse effects in insects or other taxa, although it is possible that *Chironomus* responds atypically in this case.	If there is no activity in crustaceans or insects, further evidence is probably not needed.	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is a JH/Ec (ant)agonist.

Scenarios	Result of OECD TG 218/219	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
L	–	+	Eq/0	The test chemical is probably a JH or Ec (ant)agonist without adverse effects in insects, although it is possible that *Chironomus* responds atypically in this case.	Some regulatory authorities may conclude that no further evidence is required, but if crustacean data are absent, it might be desirable to conduct a Harpacticoid Copepod Development and Reproduction Test (OECD GD 201); or a DMGT.	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is a JH/Ec agonist. It should be noted that chironomids reproduce sexually and therefore may respond differently to a parthenogenetic species such as *Daphnia*. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
M	–	–	+	The test chemical is probably without JH or Ec activity in insects, although it is possible that *Chironomus* responds atypically in this case.	Some regulatory authorities may conclude that no further evidence is required. However, it might be desirable to obtain data from crustaceans (e.g. Harpacticoid Copepod Development and Reproduction Test – OECD GD 201; or a DMGT) if these are not already available.	The lack of *in vitro* JH/Ec activity is not evidence against any JH activity, due to the limited nature of current *in vitro* JH screens. It should be noted that chironomids reproduce sexually and therefore may respond differently to a parthenogenetic species such as *Daphnia*.
N	–	–	–	The test chemical is probably without JH or Ec activity in arthropods.	No further action is necessary.	–
O	–	–	Eq/0	The test chemical is probably without JH or Ec activity in arthropods.	Some regulatory authorities may conclude that no further evidence is required. However, it might be desirable to obtain data from crustaceans (e.g. the Harpacticoid Copepod Development and Reproduction Test – OECD GD 201; or a DMGT) if these are not already available.	The lack of *in vitro* JH/Ec activity is not evidence against any JH activity, due to the limited nature of current *in vitro* JH screens. It should be noted that chironomids reproduce sexually and therefore may respond differently to a parthenogenetic species such as *Daphnia*. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
P	–	Eq/0	+	The test chemical is probably without JH or Ec activity in insects, although it is possible that *Chironomus* responds atypically in this case.	Some regulatory authorities may conclude that no further evidence is required. Also, if clear *in vitro* mechanistic data are missing, it might be desirable to obtain some.	If a new *in vitro* mechanistic assay is conducted, note that a negative result does not mean that the test material has no JH/Ec activity. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

Scenarios	Result of OECD TG 218/219	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
Q	–	Eq/0	–	The test chemical is probably without JH or Ec activity in arthropods.	No further action is necessary.	It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
R	–	Eq/0	Eq/0	The test chemical is probably without JH or Ec activity in insects and possibly crustaceans.	Some regulatory authorities may conclude that no further evidence is required. However, it might be desirable to obtain data from crustaceans (e.g. the Harpacticoid Copepod Development and Reproduction Test – OECD GD 201; or a DMGT) if these are not already available.	It should be noted that chironomids reproduce sexually and therefore may respond differently to a parthenogenetic species such as *Daphnia*. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

C.2.7. *Daphnia magna* Reproduction Test (OECD TG 211)

Status: Assay validated by the OECD.

374. Modality detected/endpoints: This *in vivo* assay with *Daphnia magna* was not primarily designed to detect endocrine-active substances, but can be responsive to juvenile hormone (JH) agonists which lead to the production of male offspring (an optional endpoint described in Annex 7 of OECD TG 211). However, it is important to note that *D. magna* can also produce male offspring in response to such natural factors as short photoperiod, temperature fluctuations, decreased food density and increased F0 population density. The lack of internationally validated mechanistic assays for endocrine activity in crustaceans may prevent firm conclusions about whether test chemicals are endocrine disrupters (EDs) in this taxon, although *in vitro* assays for JH and ecdysteroid (Ec) activity are available in the literature. However, the data from the test may nevertheless be of value for classification and hazard identification/characterisation.

Background to the assay

375. This *in vivo* assay with parthenogenetic female *Daphnia magna* is widely used to evaluate the chronic effects of non-endocrine active chemicals, but if it is used to test a JH agonist, it can lead to the production of male neonates. An adverse outcome pathway for this process is under development[1] – significant male production in a population could potentially lead to its decline. However, due to the relatively short-term nature of OECD TG 211, the endpoint of male production should not be considered as an adverse apical endpoint without further investigation in longer term tests.

376. Used in this mode, OECD TG 211 should only be considered as a non-specific screen for *in vivo* JH activity, and a positive result should ideally be followed up with longer term tests such as the *Daphnia* Multi Generation Test (DMGT). However, OECD TG 211 is resource-intensive, and a cheaper option for screening *in vivo* JH activity would be to use the Short-Term Juvenile Hormone Activity Screening Assay (SJHASA). At present, however, OECD TG 211 is the only validated *in vivo* assay which is able to identify potential JH activity.

When/why the assay may be used

377. Although OECD TG 211 could, in principle, be used at any stage in the hazard assessment process, the most likely use scenario will be when there are relatively few data available about the possible JH-disrupting properties of a chemical. The results from this assay are most likely to be available after deployment of a battery of *in vitro* screens, or as a supplement to existing data which suggest possible JH-related activity. Given the significant degree of endocrine system conservation across the arthropods, endocrine-linked effects in OECD TG 211 may also indicate the possibility of related activity in other arthropods such as copepods, decapods and insects.

378. It is possible that no endocrine-relevant data are available before OECD TG 211 is deployed (e.g. if OECD TG 211 has been used as a primary screen, even though the SJHASA may be more appropriate), but in that case a positive result in the screen should probably be followed up with relevant *in vitro* screening, if available, to investigate the suspected mode of action (MOA) in more detail. However, it should be noted that there are no standardised *in vitro* screens for JH agonists, although some are described in the scientific literature (e.g. Cherbas, Koehler and Cherbas [1989]).

Existing data to be considered

379. Existing information on endocrine-related effects from other arthropods should also be considered before deployment of OECD TG 211, given the commonality of endocrine mechanisms in these taxa. Existing data available might also include one or more of a range of *in silico* or *in vitro* results which suggest that JH disruption may occur *in vivo* (but note the limitations of this approach, as indicated above). Such indicators of possible JH activity might include quantitative structure activity relationship (QSAR) predictions of JH activity, "read-across" from *in vivo* results obtained with structurally related chemicals or positive results from an *in vitro* screen for JH agonist activity.

380. In order to provide information relevant for assessing whether or not a chemical may fulfil the WHO/IPCS (2002) definition of an ED, the study design has to be sufficiently robust to demonstrate the presence or absence of effects. In the dose selection, the investigator should also consider and ensure that data generated are adequate to fulfil the regulatory requirement across OECD countries as appropriate (e.g. hazard and risk assessment and labelling, ED assessment, etc.). The top dose or concentration should be sufficiently high to give clear systemic (i.e. non endocrine-specific) toxicity in order to ensure that a wide range of exposures (high to low) is tested. However, endocrine effects observed solely in the presence of clear systemic toxicity should be interpreted with caution and may be disregarded when sufficiently justified to be caused by secondary effects which are unlikely to be due to endocrine activity. The reason for this advice is a concern that some endocrine active substance (EAS) sensitive assays are being run at doses/concentrations of EASs that are too low to trigger direct impacts on the endocrine system. This guidance document is not the place to address this issue directly, but it should be considered when EAS-sensitive test guidelines (TGs) are revised in the future. In addition, the number and spacing of dose/concentration levels should also be adequate to fulfil the objectives of the study (e.g. to demonstrate dose response relationships if this is required).

Scenarios: Positive and negative results combined with existing data

381. In the context of this section, the terms positive and negative refer solely to the production or otherwise of male neonates. The scenarios (A to R) presented in Table C.2.7 represent all the possibilities of positive or negative results in combination with the presence or absence of existing data. The action taken will also depend on the regulatory environment, but the considerations given here are generally science based. Wherever possible, the recommended "next step which could be taken" avoids unnecessary animal testing. However, sometimes conducting an animal test will be indicated and then the relevance of species, strain and exposure route should always be considered. Further considerations specific to each scenario are given in the table.

382. Positive results obtained with the OECD TG 211 (Table C.2.7, Scenarios A-I) result in the conclusion that the test chemical is a possible JH disrupter *in vivo*, at least in crustaceans. However, as indicated above, although a positive response of OECD TG 211 indicates that the chemical is a possible JH agonist, a result of this type would generally need to be followed up with a more comprehensive screen. The most appropriate choice for this is the DMGT (a draft OECD TG). However, if countries need further evidence concerning growth and sexual development etc., a Harpacticoid Copepod Development and Reproduction Test (OECD GD 201) and/or the Sediment-Water Chironomid Life Cycle Toxicity Test (OECD TG 233) would be able to provide a precise no-observed-effect-concentration/x% effect concentration (NOEC/ECx) for adverse effects. This may be particularly important because *Daphnia* are parthenogenic, while *Amphiascus* and *Chironomus* reproduce sexually. In other words, in order to strengthen weight of evidence, a positive result in OECD TG 211 could be followed by a DMGT at Level 5. Existing data suggesting endocrine-specific activity (e.g. positive *in vitro* data, or positive *in vivo* data from other species) will strengthen the case for additional testing still further.

383. The situation in which OECD TG 211 gives a negative result (Table C.2.7, Scenarios J-R) needs careful consideration of any existing data. If these data suggest that the chemical is endocrine active both *in vitro* and *in vivo* (Scenario J), then it is possible that OECD TG 211 is simply insufficiently sensitive.

384. If OECD TG 211 and existing *in vivo* data are all negative, but *in vitro* data reveal some JH activity (Scenario K), the probability is that the test chemical is not sufficiently potent to produce JH agonism *in vivo* in arthropods, or it may be rapidly metabolised. In such a situation, further testing is probably not necessary. However, if the chemical is known to bioaccumulate slowly, it may be that exposures in the *in vivo* tests have been insufficiently prolonged, in which case longer term testing with OECD TG 233 might be justified.

385. On the other hand, if OECD TG 211 and the *in vitro* tests are negative (Scenario M), but there are positive existing *in vivo* data, the nature of those existing data should be considered. Unless the existing data are from another crustacean, the chemical is possibly not a JH agonist acting in crustaceans, but it may be more potent in species (e.g. insects) or life stages that have not been tested. In this situation, the existing *in vivo* data should be used to guide decisions about whether to conduct any further testing.

386. Finally, a negative OECD TG 211, set against a background of negative *in vitro* and *in vivo* data (Scenario N), suggests that the test chemical is probably not a JH agonist *in vitro* or *in vivo*, and further action is unnecessary.

387. In each of the above scenarios, it is possible that existing data will be equivocal, or there may be no existing data. This will weaken the conclusions which can be drawn about a negative OECD TG 211, and this is reflected in Table C.2.7. However, a lack of mechanistic data on JH activity should ideally be rectified before any further *in vivo* testing is finally conducted, although as indicated above, *in vitro* JH screens have not yet been internationally standardised. On the other hand, if OECD TG 211 is positive, further *in vivo* testing would generally be needed to quantify any adverse effects and/or to establish a NOEC or ECx for such effects, even if all existing data are equivocal, or if there are no existing data. Again, however, it may be useful to obtain some mechanistic information before conducting further *in vivo* testing. There is also the possibility that equivocal mechanistic data may be the result of multiple modes of endocrine action. Under rare circumstances, two opposite modes of simultaneous action (e.g. JH agonistic and antagonistic) could, depending on dose, lead to a minimisation or abolition of adverse

effects, while in others two different MOA could potentially reinforce effects on the OECD TG 211 endpoint. If multiple MOA are suspected, either from the existing results or based on QSAR/read-across/integrated approaches, this situation should be investigated further if needed for regulatory decision making.

388. The scenario in which the results of OECD TG 211 are themselves equivocal has not been dealt with in Table C.2.7, for reasons of brevity. In this context, an equivocal result might be an inconsistent concentration-response (e.g. no effect at a high concentration but effects at a lower concentration), or a result which borders on statistical significance. Without knowing the exact circumstances, reliable advice cannot be given, but the opinions of an experienced ecotoxicologist should be sought. Clearly, however, such equivocal results do not necessarily rule out the existence of *in vivo* endocrine activity. If possible reasons for false negatives are suspected, OECD TG 211 could be repeated (e.g. conduct it at lower concentrations which avoid systemic toxicity). However, note that a repeat screen in the event of systemic toxicity would not be needed providing at least one tested concentration was not subject to such effects. It should also be borne in mind that changing environmental conditions such as shortening photoperiod, temperature and food shortages can also cause the production of male neonates in *D. magna*, so if these have accidentally occurred during the test, the results may constitute false positives and should be treated as suspect.

389. In summary, positive results in OECD TG 211 may indicate that a chemical is endocrine active *in vivo* via JH agonism. This suggests that more comprehensive *in vivo* testing would be needed if the intention is to derive a long-term NOEC/ECx and/or to confirm whether or not the chemical is an actual endocrine disrupter in arthropods due to the occurrence of adverse effects. Negative results in OECD TG 211 do not necessarily mean that the chemical is not a potential ED – a judgement about the endocrine disruption potential in other arthropods (especially sexually reproducing species) and the possible need for additional testing will have to be made based on a weight of evidence evaluation of existing *in vitro* and *in vivo* data.

Note

1. See: https://aopwiki.org/wiki/index.php/Aop:201.

References

Cherbas, L., M.M.D. Koehler and P. Cherbas (1989), "Effects of juvenile hormone on the ecdysone response of *Drosophila* Kc cells", *Developmental Genetics*, Vol. 10/3, pp. 177-188, https://doi.org/10.1002/dvg.1020100307.

WHO/IPCS (2002), "Global assessment of the state-of-the-science of endocrine disrupters", Damstra, T. et al. (eds.) WHO/PCS/EDC/02.2, World Health Organization, Geneva, www.who.int/ipcs/publications/new_issues/endocrine_disruptors/en.

Table C.2.7. ***Daphnia magna*** **Reproduction Test (OECD TG 211): Guidance for scenarios of combinations of results with existing data**

This table represents possible conclusions to be drawn from assay data, and a next step which could be taken if further evidence is required about possible endocrine disrupting properties and/or effects. The guidance offered is not meant to be prescriptive, but provides science-based considerations. It encourages the use of all available data and expert judgement in a weight of evidence approach. Regional and national interpretation of results and "next steps" may vary.

The conclusions are grouped into a series of scenarios (A-R), each scenario representing a different combination of assay results, existing *in vitro* data and existing *in vivo* data. The symbol "+" indicates that the data in question represent a positive result, "-" indicates a negative result, and "Eq/0" indicates that the data are either equivocal or are not available. Note that the terms positive and negative refer solely to the optional male-production endpoint of TG 211.

Existing results: * "Mechanism (*in vitro* mechanistic data)" assumes that mechanistic data are available from available from juvenile hormone (JH-) based assays. JH assays concerning mechanisms of JH disruption may be available, but have not yet been internationally standardised. In practice, data from all assays may not be available and therefore this must be taken into account when deciding on the "next step".

Existing results: ** "Effects (*in vivo* effects of concern)" assumes effects have been observed in other *in vivo* screens/tests which give rise to concern that the test chemical may be a JH disrupter.

Scenarios	Result of OECD TG 211 (male endpoint only)	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
A	+	+	+	Strong evidence for *in vivo* juvenile hormone (JH) activity in crustaceans, plus possible JH effects in other arthropods.	Consider performing a *Daphnia* Multigeneration Test (DMGT – draft OECD TG).	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is a JH agonist. The DMGT will show if sex ratio bias towards males carries over into the F2 generation and some regulatory authorities may consider that this provides sufficient information on adverse apical effects in crustaceans. However, as *Daphnia* are parthenogenetic, it would be desirable to perform an additional apical test with sexually reproducing crustaceans and/or insects (e.g. the Harpacticoid Copepod Development and Reproduction Test – OECD GD 201; and/or the Sediment-Water Chironomid Life Cycle Toxicity Test – OECD TG 233).
B	+	+	–	Strong evidence for *in vivo* JH activity in crustaceans.	Consider performing a DMGT (draft OECD TG).	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is a JH agonist. The DMGT will show if sex ratio bias towards males carries over into the F2 generation and some regulatory authorities may consider that this provides sufficient information on adverse apical effects in crustaceans. However, as *Daphnia* are parthenogenetic, it would be desirable to perform an additional apical test with sexually reproducing crustaceans and/or insects (e.g. the Harpacticoid Copepod Development and Reproduction Test – OECD GD 201; and/or the Sediment-Water Chironomid Life Cycle Toxicity Test – OECD TG 233).
C	+	+	Eq/0	Strong evidence for *in vivo* JH activity in crustaceans.	Consider performing a DMGT (draft OECD TG).	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is a JH agonist. The DMGT will show if sex ratio bias towards males carries over into the F2 generation and some regulatory authorities may consider that this provides sufficient information on adverse apical effects in crustaceans. However, as *Daphnia* are parthenogenetic, it would be desirable to perform an additional apical test with sexually reproducing crustaceans and/or insects (e.g. the Harpacticoid Copepod Development and Reproduction Test – OECD GD 201; and/or the Sediment-Water Chironomid Life Cycle Toxicity Test – OECD TG 233). It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple modes of action (MOA). If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

Scenarios	Result of OECD TG 211 (male endpoint only)	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
D	+	–	+	Moderate evidence for *in vivo* JH activity in crustaceans, plus possible JH effects in other arthropods.	Consider performing a DMGT (draft OECD TG).	The lack of *in vitro* JH activity is not evidence against any JH activity, due to the limited nature of current *in vitro* JH screens. The DMGT will show if sex ratio bias towards males carries over into the F2 generation and some regulatory authorities may consider that this provides sufficient information on adverse apical effects in crustaceans. However, as *Daphnia* are parthenogenetic, it would be desirable to perform an additional apical test with sexually reproducing crustaceans and/or insects (e.g. the Harpacticoid Copepod Development and Reproduction Test – OECD GD 201; and/or the Sediment-Water Chironomid Life Cycle Toxicity Test – OECD TG 233).
E	+	–	–	Some evidence for *in vivo* JH activity in crustaceans.	Consider performing a DMGT (draft OECD TG).	The lack of *in vitro* JH activity is not evidence against any JH activity, due to the limited nature of current *in vitro* JH screens. The DMGT will show if sex ratio bias towards males carries over into the F2 generation and some regulatory authorities may consider that this provides sufficient information on adverse apical effects in crustaceans. However, as *Daphnia* are parthenogenetic, it would be desirable to perform an additional apical test with sexually reproducing crustaceans and/or insects (e.g. the Harpacticoid Copepod Development and Reproduction Test – OECD GD 201; and/or the Sediment-Water Chironomid Life Cycle Toxicity Test – OECD TG 233).
F	+	–	Eq/0	Some evidence for *in vivo* JH activity in crustaceans.	Consider performing a DMGT (draft OECD TG). Given the absence or equivocal nature of existing *in vivo* data, it might also be sensible to conduct a JH-responsive insect assay (e.g. the Sediment-Water Chironomid Life Cycle Toxicity Test – OECD TG 233).	The lack of *in vitro* JH activity is not evidence against any JH activity, due to the limited nature of current *in vitro* JH screens. The DMGT will show if sex ratio bias towards males carries over into the F2 generation and some regulatory authorities may consider that this provides sufficient information on adverse apical effects in crustaceans. However, as *Daphnia* are parthenogenetic, it would be desirable to perform an additional apical test with sexually reproducing crustaceans (e.g. the Harpacticoid Copepod Development and Reproduction Test – OECD GD 201) It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

Scenarios	Result of OECD TG 211 (male endpoint only)	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
G	+	Eq/0	+	Moderate evidence for *in vivo* JH activity in crustaceans, plus possible JH effects in other arthropods.	Consider performing a DMGT (draft OECD TG). Given the absence or equivocal nature of the *in vitro* mechanistic data, it might also be helpful to conduct an *in vitro* screen for JH activity.	If a new *in vitro* mechanistic assay is conducted, note that a negative result does not mean that the test material has no JH activity. The DMGT will show if sex ratio bias towards males carries over into the F2 generation and some regulatory authorities may consider that this provides sufficient information on adverse apical effects in crustaceans. However, as *Daphnia* are parthenogenetic, it would be desirable to perform an additional apical test with sexually reproducing crustaceans (e.g. the Harpacticoid Copepod Development and Reproduction Test – OECD GD 201 and/or the Sediment-Water Chironomid Life Cycle Toxicity Test – OECD TG 233). It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
H	+	Eq/0	–	Some evidence for *in vivo* JH activity in crustaceans.	Consider performing a DMGT (draft OECD TG). Given the absence or equivocal nature of the *in vitro* mechanistic data, it might also be helpful to conduct an *in vitro* screen for JH activity.	The DMGT will show if sex ratio bias towards males carries over into the F2 generation and some regulatory authorities may consider that this provides sufficient information on adverse apical effects in crustaceans. However, as *Daphnia* are parthenogenetic, it would be desirable to perform an additional apical test with sexually reproducing crustaceans (e.g. the Harpacticoid Copepod Development and Reproduction Test – OECD GD 201 and/or the Sediment-Water Chironomid Life Cycle Toxicity Test – OECD TG 233). It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
I	+	Eq/0	Eq/0	Some evidence for *in vivo* JH activity in crustaceans.	Consider performing a DMGT (draft OECD TG). Given the absence or equivocal nature of the *in vitro* mechanistic data, it might also be helpful to conduct an *in vitro* screen for JH activity.	If a new *in vitro* mechanistic assay is conducted, note that a negative result does not mean that the test material has no JH activity. The DMGT will show if sex ratio bias towards males carries over into the F2 generation and some regulatory authorities may consider that this provides sufficient information on adverse apical effects in crustaceans. However, as *Daphnia* are parthenogenetic, it would be desirable to perform an additional apical test with sexually reproducing crustaceans (e.g. the Harpacticoid Copepod Development and Reproduction Test – OECD GD 201 and/or the Sediment-Water Chironomid Life Cycle Toxicity Test – OECD TG 233). It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

Scenarios	Result of OECD TG 211 (male endpoint only)	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
J	–	+	+	The test chemical is probably a JH agonist without activity in crustaceans, although it is possible that *Daphnia magna* responds atypically in this case.	Some regulatory authorities may conclude that no further evidence is required. However, it might be desirable to obtain data from insects (e.g. the Sediment-Water Chironomid Life Cycle Toxicity Test – OECD TG 233) if these are not already available.	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is a JH agonist.
K	–	+	–	The test chemical is likely to have JH activity; however, without demonstrating sufficient activity to disrupt physiological processes *in vivo*.	If there is no activity in crustaceans or insects, further evidence is probably not needed.	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is a JH agonist.
L	–	+	Eq/0	The test chemical is likely to have JH activity; however, without demonstrating sufficient activity to disrupt physiological processes *in vivo*.	Some regulatory authorities may conclude that no further evidence is required, but if insect data are absent, it might be desirable to conduct a Sediment-Water Chironomid Life Cycle Toxicity Test – OECD TG 233.	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is a JH agonist. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
M	–	–	+	The test chemical is probably without JH activity in crustaceans, although it is possible that *Daphnia magna* responds atypically in this case.	Some regulatory authorities may conclude that no further evidence is required. However, it might be desirable to obtain data from insects (e.g. the Sediment-Water Chironomid Life Cycle Toxicity Test – OECD TG 233) if these are not already available.	The lack of *in vitro* JH activity is not evidence against any JH activity, due to the limited nature of current *in vitro* JH screens. However, it is possible that the existing effects may not be due to JH activity.
N	–	–	–	The test chemical is probably without JH activity in arthropods.	No further action is necessary.	–
O	–	–	Eq/0	The test chemical is probably without JH activity in arthropods.	Some regulatory authorities may conclude that no further evidence is required. However, it might be desirable to obtain data from insects (e.g. the Sediment-Water Chironomid Life Cycle Toxicity Test – OECD TG 233) if these are not already available.	The lack of *in vitro* JH activity is not evidence against any JH activity, due to the limited nature of current *in vitro* JH screens. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

Scenarios	Result of OECD TG 211 (male endpoint only)	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
P	–	Eq/0	+	The test chemical is probably without JH activity in crustaceans, although it is possible that *Daphnia magna* responds atypically in this case.	Some regulatory authorities may conclude that no further evidence is required. Also, if clear *in vitro* mechanistic data are missing, it might be desirable to obtain some.	If a new *in vitro* mechanistic assay is conducted, note that a negative result does not mean that the test material has no JH activity. However, it is possible that the existing effects may not be due to JH activity. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
Q	–	Eq/0	–	The test chemical is probably without JH activity in arthropods.	No further action is necessary.	It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
R	–	Eq/0	Eq/0	The test chemical is probably without JH activity in crustaceans and possibly insects.	Some regulatory authorities may conclude that no further evidence is required. However, it might be desirable to obtain data from insects (e.g. the Sediment-Water Chironomid Life Cycle Toxicity Test – OECD TG 233) if these are not already available.	It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

C.2.8. Fish, Early-Life Stage (FELS) Toxicity Test (OECD TG 210)

Status: Assay validated by the OECD.

390. Modality detected/endpoints: This test has no endocrine-specific endpoints. However, there is limited evidence to suggest that some thyroid system disrupters are able to interfere with metamorphosis of the fish embryo to the larva.

Background to the assay

391. This test is widely used as a sub-chronic assay for non-endocrine disrupting (ED) chemicals, and can be used to predict concentrations causing chronic effects on growth and reproduction in fish. It was developed before concerns about endocrine disrupting chemicals (EDCs) arose and cannot be used to identify these chemicals. It exposes fish from immediately post-fertilisation to the larval free-feeding stage (28-60 days post-hatch [dph], depending on species). Permitted species include rainbow trout (*Oncorhynchus mykiss*), fathead minnow (*Pimephales promelas*), zebrafish (*Danio rerio*), medaka (*Oryzias latipes*), sheepshead minnow (*Cyprinodon variegatus*) and silverside (*Menidia* sp.). The main endpoints include mortality, time to hatching, hatching success, growth, morphological abnormalities and abnormal behaviour.

392. Although the test does not have endpoints that specifically respond to EDCs alone, there are limited data which show that it is responsive to certain thyroid-disrupting chemicals. It is known that thyroid hormone receptors TRα and TRβ are both present in fish early embryos and larvae (Power et al., 2001), and that maternally derived thyroxine (T4) is important for thyroid-dependent processes in fish early life stages (Nelson et al., 2014). One of these processes is swimbladder inflation, an endpoint which can be recorded in the FELS test, and which is vital for the survival of fish fry. It has been shown, for example, that fathead minnow embryos exposed to a thyroid peroxidase (TPO) inhibitor (2-mercaptobenzothiazole) do not develop inflated swimbladders, probably because inhibition of TPO leads to decreased thyroid hormone synthesis (Villeneuve et al., 2013; Nelson et al., 2014). Also, Liu and Chan (2002) have shown that metamorphosis from embryo to larva in zebrafish is arrested by exposure to amiodarone (a TR antagonist) and by the goitrogen methimazole. Furthermore, Shi et al. (2008) demonstrated that the thyroid disrupter perfluorooctanesulfonic acid (PFOS) is able to delay hatching and cause developmental malformations in zebrafish embryos while upregulating two thyroid-related developmental genes, *hhex* and *pax8*. However, it is important to note that many non-ED chemicals will also cause these types of apical response, but by different mechanisms.

When/why the assay may be used

393. For an existing chemical, it is quite likely that data from a FELS (OECD TG 210) will already be available. If this is the case, indications of damage to the metamorphosis of fish embryos to larvae could be used as supporting data for a case that the chemical may be a thyroid disrupter. However, as stated above, there are many non-EDCs which are also

able to damage fish metamorphosis. Given the limited data (with respect to endocrine disruption) available from a FELS test, it would generally not be appropriate to request this assay especially to evaluate a suspected thyroid-acting ED.

394. Caution should be used when negative results are obtained with certain types of chemicals because absorption into the embryo via the chorion may have been impeded. Development of the OECD Fish Embryo Acute Toxicity (FET) Test (OECD TG 236) with zebrafish showed that this applies in particular to chemicals with a molecular weight ≥3kDa and a very bulky molecular structure. Absorption of these chemicals will take place at a higher rate after hatching, but delayed hatch may therefore also protect the embryo from other forms of toxicity. Although it is known that fish embryos have some metabolic capacity (e.g. Weigt et al. [2011]), this may be less efficient than in juveniles and adults, so use of the test with EDCs that require metabolic activation may give false negatives.

395. In order to provide information relevant for assessing whether or not a chemical may fulfil the WHO/IPCS (2002) definition of an ED, the study design has to be sufficiently robust to demonstrate the presence or absence of effects. In the dose selection, the investigator should also consider and ensure that data generated are adequate to fulfil the regulatory requirement across OECD countries as appropriate (e.g. hazard and risk assessment and labelling, ED assessment, etc.). The top dose or concentration should be sufficiently high to give clear systemic (i.e. non endocrine-specific) toxicity in order to ensure that a wide range of exposures (high to low) is tested. However, endocrine effects observed solely in the presence of clear systemic toxicity should be interpreted with caution and may be disregarded when sufficiently justified to be caused by secondary effects which are unlikely to be due to endocrine activity. The reason for this advice is a concern that some endocrine active substance (EAS) sensitive assays are being run at doses/concentrations of EASs that are too low to trigger direct impacts on the endocrine system. This guidance document is not the place to address this issue directly, but it should be considered when EAS-sensitive test guidelines (TGs) are revised in the future. In addition, the number and spacing of dose/concentration levels should also be adequate to fulfil the objectives of the study (e.g. to demonstrate dose response relationships if this is required).

Existing data to be considered

396. Existing data available before consideration of OECD TG 210 might include *in vivo* results obtained with other vertebrates (e.g. a positive *in vivo* assay with amphibians, for example OECD TG 231; or positive findings for thyroid endpoints in mammalian repeat dose toxicity studies, for example OECD TG 407), or one or more of a range of *in silico* or *in vitro* results which suggest that thyroid disruption may occur *in vivo* (but note the limitations of this approach, given that validated *in silico* and *in vitro* screens for thyroid activity are not yet available). Such indicators of possible thyroid activity might include quantitative structure activity relationship (QSAR) predictions of thyroid activity, "read-across" from *in vivo* results obtained with structurally related chemicals or positive results from an *in vitro* screen for thyroid agonist/antagonist activity.

Scenarios: Positive and negative results combined with existing data

397. The scenarios (A to R) presented in Table C.2.8 represent all the possibilities of positive or negative results in combination with the presence or absence of existing data. The action taken will also depend on the regulatory environment, but the considerations given here are generally science based. Wherever possible, the recommended "next step which could be taken" avoids unnecessary animal testing. However, sometimes conducting an animal

test will be indicated and then the relevance of species, strain and exposure route should always be considered. Further considerations specific to each scenario are given in the table.

398. Positive results obtained with one or more of the endpoints (Table C.2.8, Scenarios A-I) will not result in a conclusion that the test chemical is a possible thyroid-acting ED *in vivo*, although this may strengthen such a case. In some cases, this may need to be followed up with more comprehensive testing to show whether adverse apical effects related to thyroid impacts occur in sensitive species such as amphibians at any part of the life cycle (and hence to discover whether the chemical is an ED acting through thyroid mechanisms). In other words, a positive result in OECD TG 210 alone will not trigger further testing. Existing data suggesting endocrine activity will strengthen the case for additional testing.

399. The situation in which OECD TG 210 gives a negative result (Table C.2.8, Scenarios J-R) needs careful consideration of any existing data. If the weight of evidence of these data suggests that the chemical is thyroid-active both *in vitro* and *in vivo* in other species (Scenario J), then the probability is that OECD TG 210 may simply be insufficiently responsive in that case, or fish in general may be unresponsive. In some of these circumstances, it might be appropriate to conduct an Amphibian Metamorphosis Assay (AMA; OECD TG 231) or a Larval Amphibian Growth and Development Assay (LAGDA; OECD TG 241).

400. If OECD TG 210 and existing *in vivo* data are all negative, but *in vitro* data reveal some endocrine activity (Scenario K), the probability is that the test chemical is not sufficiently potent to produce endocrine effects *in vivo* in fish, or it may be rapidly metabolised. In such a situation, further testing will generally not be necessary. If the chemical is known to bioaccumulate slowly, it may be that exposures in the *in vivo* tests were not of sufficient duration, in which case longer term testing might be justified. If existing information, including QSAR predictions, *in vivo* mammalian data and/or *in vitro* data reveal thyroid activity, consideration should be given to conducting the AMA (OECD TG 231) or LAGDA (OECD TG 241), unless exposure of the aquatic environment can be excluded.

401. On the other hand, if OECD TG 210 and the *in vitro* tests are negative, but there are positive existing *in vivo* data (Scenario M), the chemical may not be thyroid-active in fish. In this situation, the relevant existing non-test and *in vitro* or *in vivo* data should be used to guide decisions about whether to conduct any further *in vivo* testing for thyroid activity with the AMA, or the LAGDA if a positive AMA is already available.

402. Finally, a negative OECD TG 210 test, set against a background of negative *in vitro* and *in vivo* data (Scenario N), suggests that the test chemical is not a possible thyroid-acting ED in fish or other vertebrates, and no further testing for thyroid modes of action (MOA) will generally be necessary.

403. In each of the above scenarios, it is possible that existing data will be equivocal, or there may be no existing data (Scenarios C, F-I, L and O-R). This will weaken the conclusions which can be drawn about a negative OECD TG 210 test, and this is reflected in Table C.2.8. However, a lack of mechanistic data on thyroid activity should usually be rectified before any further *in vivo* testing is finally rejected. Indeed, as a general principle, it is desirable to obtain mechanistic data from non-testing and/or *in vitro* testing approaches before any *in vivo* testing which targets endocrine disruption. On the other hand, if OECD TG 210 is positive but existing data are equivocal, this does not necessarily indicate that the test chemical is thyroid-active, and further *in vitro* or *in vivo* testing may be desirable.

There is also the possibility that equivocal mechanistic data may be the result of multiple modes of endocrine action. Under some circumstances, two opposite modes of simultaneous action (e.g. thyroidogenic and anti-thyroidogenic) could, depending on dose, lead to a minimisation or abolition of adverse effects, while in others two different MOA could potentially reinforce effects on certain apical endpoints and/or modify the typical adverse outcome signs related to certain ED MOA. If multiple MOA are suspected, either from the existing results or based on QSAR/read-across/integrated approaches, this situation should be investigated further if needed for regulatory decision making.

404. The scenario in which the results of OECD TG 210 are themselves equivocal has not been dealt with in Table C.2.8, for reasons of brevity. In this context, an equivocal result might be an inconsistent concentration-response (e.g. no effect at a high concentration but effects at a lower concentration), or a result which borders on statistical significance. Without knowing the exact circumstances, reliable advice cannot be given, but the opinions of an experienced ecotoxicologist should be sought. Clearly, however, such equivocal results do not necessarily rule out the existence of *in vivo* thyroid activity. If convincing reasons for false negatives are suspected (e.g. the test chemical is too bulky to be absorbed via the chorion or it requires metabolic activation), the results from OECD TG 210 should not be considered further but other *in vivo* testing may be indicated.

405. In summary, positive results in the OECD TG 210 test may support the case that a chemical is a possible thyroid disrupter, but cannot on their own be used to reach such a conclusion. More thyroid-specific *in vivo* testing would then be necessary to produce a long-term NOEC/ECx and/or to confirm whether or not the chemical is an actual thyroid disrupter with adverse effects *in vivo*. For suspected thyroid-active chemicals, the best available apical test is the LAGDA (OECD TG 241). Negative results in OECD TG 210 do not necessarily mean that the chemical is not a possible thyroid disrupter – a judgement about its endocrine disruption potential and the possible need for additional testing will have to be made based on a weight of evidence evaluation of existing non-test, *in vitro* and *in vivo* data.

References

Liu, Y.-W. and W.K. Chan (2002), "Thyroid hormones are important for embryonic to larval transitory phase in zebrafish", *Differentiation*, Vol. 70/1, pp. 36-45, https://doi.org/10.1046/j.1432-0436.2002.700104.x.

Nelson, K. et al. (2014), "Evaluation of hypothesized adverse outcome pathway linking thyroid peroxidase to fish early life stage toxicity", presentation, Society of Environmental Toxicology and Chemistry, Vancouver, British Columbia, Canada, 9-13 November 2014.

Power, D.M. et al. (2001), "Thyroid hormones in growth and development of fish", *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology*, Vol. 130/4, pp. 447-459, https://doi.org/10.1016/S1532-0456(01)00271-X.

Shi, X. et al. (2008), "Developmental toxicity and alteration of gene expression in zebrafish embryos exposed to PFOS", *Toxicology and Applied Pharmacology*, Vol. 230/1, pp. 23-32, https://doi.org/10.1016/j.taap.2008.01.043.

Villeneuve, S. et al. (2013), "Investigating alternatives to the Fish Early-Life Stage Test: A strategy for discovering and annotating adverse outcome pathways for early fish development", *Environmental Toxicology and Chemistry*, Vol. 33/1, pp. 158-169, https://doi.org/10.1002/etc.2403.

Weigt, S. et al. (2011), "Zebrafish (*Danio rerio*) embryos as a model for testing proteratogens", *Toxicology*, Vol. 281/1-3, pp. 25-36, https://doi.org/10.1016/j.tox.2011.01.004.

WHO/IPCS (2002), "Global assessment of the state-of-the-science of endocrine disrupters", Damstra, T. et al. (eds.) WHO/PCS/EDC/02.2, World Health Organization, Geneva, www.who.int/ipcs/publications/new_issues/endocrine_disruptors/en.

Table C.2.8. **Fish, Early-Life Stage (FELS) Toxicity Test (OECD TG 210): Guidance for scenarios of combinations of results with existing data**

This table represents possible conclusions to be drawn from assay data, and a next step which could be taken if further evidence is required about possible endocrine disrupting properties and/or effects. The guidance offered is not meant to be prescriptive, but provides science-based considerations. It encourages the use of all available data and expert judgement in a weight of evidence approach. Regional and national interpretation of results and "next steps" may vary.

The conclusions are grouped into a series of scenarios (A-R), each scenario representing a different combination of assay results, existing *in vitro* data and existing *in vivo* data. The symbol "+" indicates that the data in question represent a positive result, «-" indicates a negative result, and "Eq/0" indicates that the data are either equivocal or are not available. Note that there are no endpoints in this assay considered to be diagnostic of an estrogen/androgen/thyroid/steroidogenesis (E,A,T,S) modality.

Existing results: * "Mechanism (*in vitro* mechanistic data)" assumes that mechanistic data are available from thyroid hormone receptor (TR) and other assays concerning mechanisms of thyroid disruption, although these are not in common use. In practice, data from all assays may not be available and therefore this must be taken into account when deciding on the "next step". Quantitative structure activity relationship (QSAR) predictions of estrogen and androgen binding/activation may be made for some substances. There is no evidence at present that equivalent *in vitro* assays with systems derived from fish offer advantages over their mammalian counterparts.

Existing results: ** "Effects (*in vivo* effects of concern)" assumes effects have been observed in other *in vivo* screens/tests which give rise to concern that the test chemical may be an endocrine disrupter.

Scenarios	Result of OECD TG 210 assay (FELS)	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
A	+	+	+	Limited evidence for *in vivo* thyroid activity with potential adverse effects (developmental/growth toxicity) in fish, plus thyroid effects in other species	Consider performing an Amphibian Metamorphosis Assay (AMA), or a Larval Amphibian Growth and Development Assay (LAGDA) if a positive AMA result is already available, or if it is regarded as *a priori* likely that an AMA would be positive.	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is a thyroid (ant)agonist.
B	+	+	–	Weak evidence for *in vivo* thyroid activity with potential adverse effects (developmental/growth toxicity) in fish or other species.	Consider performing an AMA.	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is a thyroid (ant)agonist.
C	+	+	Eq/0	Weak evidence for *in vivo* thyroid activity with potential adverse effects (developmental/growth toxicity) in fish or other species.	Consider performing an AMA.	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is a thyroid (ant)agonist. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple modes of action (MOA). If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
D	+	–	+	Limited evidence for endocrine activity which may be thyroid-related, with potential adverse effects (developmental/growth toxicity) in fish, plus thyroid effects in other species.	Consider performing an AMA, or a LAGDA if a positive AMA result is already available.	The lack of *in vitro* thyroid activity is not evidence against any thyroid activity, due to the limited nature of current *in vitro* thyroid screens.
E	+	–	–	No evidence for endocrine activity, but damaged metamorphosis in the Fish, Early-Life Stage (FELS) assay could be caused by thyroid disruption.	Consider performing *in vitro* thyroid assays, quantitative structure activity relationship (QSAR) predictions or read-across if these have not already been conducted.	
F	+	–	Eq/0	No evidence for endocrine activity, but damaged metamorphosis in the FELS assay could be caused by thyroid disruption.	Consider performing *in vitro* thyroid assays if these have not already been conducted. If these are positive, it may be desirable to conduct an AMA.	It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

Scenarios	Result of OECD TG 210 assay (FELS)	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
G	+	Eq/0	+	Weak evidence for endocrine activity which may be thyroid-related, with potential adverse effects (developmental/growth toxicity) in fish, plus thyroid effects in other species.	Consider performing a new *in vitro* thyroid assay for (ant)agonistic activity, or QSAR predictions or read-across.	If a new *in vitro* mechanistic assay is conducted, note that a negative result does not mean that the test material has no thyroid activity. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
H	+	Eq/0	–	No evidence for endocrine activity, but damaged metamorphosis in the FELS assay could be caused by thyroid disruption.	Consider performing a new *in vitro* thyroid assay for (ant)agonistic activity, or QSAR predictions or read-across.	If a new *in vitro* mechanistic assay is conducted, note that a negative result does not mean that the test material has no thyroid activity. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
I	+	Eq/0	Eq/0	No evidence for endocrine activity, but damaged metamorphosis in the FELS assay could be caused by thyroid disruption.	Consider performing a new *in vitro* thyroid assay for (ant)agonistic activity. If this is positive, it may be desirable to conduct an AMA.	If a new *in vitro* mechanistic assay is conducted, note that a negative result does not mean that the test material has no thyroid activity. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
J	–	+	+	The test chemical may be a thyroid (ant)agonist without activity in fish.	Some regulatory authorities may conclude that no further evidence is required, but it might be desirable to conduct an AMA, or a LAGDA if a positive AMA is already available.	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is a thyroid (ant)agonist.
K	–	+	–	The test chemical may a thyroid (ant)agonist without activity *in vivo*.	If there is no activity in fish, amphibians or mammals, further evidence is probably not needed.	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is a thyroid (ant)agonist.

Scenarios	Result of OECD TG 210 assay (FELS)	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
L	–	+	Eq/0	The test chemical may a thyroid (ant)agonist without activity *in vivo*.	If there is no activity in fish or mammals, further evidence is probably not needed. However, if the equivocal or absent *in vivo* data relate to amphibians, it may be desirable to repeat an AMA.	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is a thyroid (ant)agonist. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
M	–	–	+	The test chemical may be without thyroid activity in fish.	Some regulatory authorities may conclude that no further evidence is required, but it might be helpful to perform an AMA, or a LAGDA if the positive *in vivo* data are from an AMA.	The lack of *in vitro* thyroid activity is not evidence against any thyroid activity, due to the limited nature of current *in vitro* thyroid screens.
N	–	–	–	The test chemical may be without thyroid activity in fish or other taxa.	No further action is necessary.	–
O	–	–	Eq/0	The test chemical may be without thyroid activity in fish or other taxa.	Some regulatory authorities may conclude that no further evidence is required, but if mammalian or amphibian data are absent, it might be desirable to conduct a thyroid-responsive rodent screen (e.g. rat pubertal) or an AMA.	The lack of *in vitro* thyroid activity is not evidence against any thyroid activity, due to the limited nature of current *in vitro* thyroid screens. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
P	–	Eq/0	+	The test chemical may be without thyroid activity in fish.	Some regulatory authorities may conclude that no further evidence is required, but it might be considered worthwhile to conduct an(other) *in vitro* thyroid assay, QSAR prediction or read-across.	If a new *in vitro* mechanistic assay is conducted, note that a negative result does not mean that the test material has no thyroid activity. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

Scenarios	Result of OECD TG 210 assay (FELS)	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
Q	–	Eq/0	–	The test chemical may be without thyroid activity in fish or other taxa.	No further action is necessary.	It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
R	–	Eq/0	Eq/0	The test chemical may be without thyroid activity in fish.	Some regulatory authorities may conclude that no further evidence is required, but it might be considered worthwhile to conduct an(other) *in vitro* thyroid assay, or QSAR predictions or read-across.	If a new *in vitro* mechanistic assay is conducted, note that a negative result does not mean that the test material has no thyroid activity. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

C.2.9. Fish Sexual Development Test (FSDT) (OECD TG 234)

Status: Assay validated by the OECD.

406. Modality detected/endpoints: estrogens (♀ and ♂ VTG ↑; phenotypic sex ratio ♀↑); anti-estrogens (♀ VTG ↓; phenotypic sex ratio ♂↑; sexually undifferentiated fish ↑); androgens (phenotypic sex ratio ♂↑; ♀ VTG ↓); anti-androgens (intersex fish ↑; ♀ VTG ↑; phenotypic sex ratio ♀↑); aromatase inhibitors (♀ VTG ↓; phenotypic sex ratio ♂↑); (optional endpoints – gonadal histopathology; genetic sex in medaka and stickleback). OECD TG 234 (FSDT) has been fully validated for Japanese medaka, zebrafish and stickleback. The test may also be responsive to certain thyroid-disrupting chemicals. It is known that thyroid hormone receptors TRα and TRβ are both present in fish early embryos and larvae (Power et al., 2001), and that maternally derived thyroxine (T4) is important for thyroid-dependent processes in fish early life stages (Nelson et al., 2014). One of these processes is swimbladder inflation, an endpoint which could be recorded in the FSDT test, and which is vital for the survival of fish fry. It has been shown, for example, that fathead minnow embryos exposed to a thyroid peroxidase (TPO) inhibitor (2-mercaptobenzothiazole) do not develop inflated swimbladders, probably because inhibition of TPO leads to decreased thyroid hormone synthesis (Villeneuve et al., 2013; Nelson et al., 2014). Also, Liu and Chan (2002) have shown that metamorphosis from embryo to larva in zebrafish is arrested by exposure to amiodarone (a TR antagonist) and by the goitrogen methimazole. Furthermore, Shi et al. (2008) demonstrated that the thyroid disrupter perfluorooctanesulfonic acid (PFOS) is able to delay hatching and cause developmental malformations in zebrafish embryos while upregulating two thyroid-related developmental genes, *hhex* and *pax8*. However, it is important to note that many non-ED chemicals will also cause these types of apical response, but by different mechanisms.

Background to the assay

407. This partial life cycle assay could potentially be used as a screen for the types of *in vivo* endocrine disruption activity in fish which are listed above (although it is considerably more expensive and time-consuming than the OECD TG 229/230 or EASZY screens), but should generally be used as a test which can also provide apical information of use in hazard identification/characterisation. It includes an endpoint (altered sex ratio), which is indicative of endocrine action, but more importantly indicates that adverse apical effects on sexual development are occurring. Major effects on phenotypic sex ratio would be expected to damage the ability of a fish population to reproduce itself, although small effects may be tolerated, but it is not possible to define the precise change in sex ratio beyond which adverse effects will occur unless specific information about a particular population is available. It should be noted that if the assay gives a positive result, this may be due to a positive indicator of hormonal activity (e.g. vitellogenin [VTG]), a positive for biased sex ratio, or a positive for both types of endpoint. Each of these three possible combinations of positive response should be considered separately (although the distinctions between indicators of hormonal activity and apical effects are not as clear in

OECD TG 234 [FSDT] as in other tests because it is acknowledged that sex ratio is both an apical endpoint [relevant for populations] as well as a biomarker endpoint [indicative of mode of action]), so they have been listed individually as points 1, 2 and 3 in the possible conclusions column of Table C.2.9.

408. If only three test concentrations are employed, a reliable NOEC or ECx for biased sex ratio may not be obtainable, so it may be desirable to use at least five test concentrations. However, if the test is used for hazard identification/characterisation, the stickleback should not be used because the validation data available so far show that in this species alterations of phenotypic sex ratio by test substances are uncommon. It should be noted that simultaneous measurement of both phenotypic and genotypic sex ratio (currently only possible in medaka and stickleback) will increase the statistical power of the test. However, power analysis of the validation results was used to prepare a test design providing sufficient power to detect changes in both sex ratio and VTG for the currently validated species. The ability of a substance with a suspected specific endocrine mechanism to change the sex ratio of fish should be considered during the choice of fish test species because some species are more susceptible to sex ratio changes caused by a specific endocrine mechanism than others. For example, zebrafish sex ratio is very sensitive to androgen agonists. Power analyses indicate that adequate power can be achieved with zebrafish as long as sufficient replication and fish per replicate are used (OECD, 2012). Given the high degree of endocrine system conservation across the vertebrates, adverse endocrine-linked effects in the FSDT may also indicate the possibility of related activity in other organisms such as amphibians, reptiles, birds or mammals.

When/why the assay may be used

409. Although OECD TG 234 (FSDT) could, in principle, be used at any stage in the hazard assessment process, the most likely use scenario will be when there are already some *in vitro* or *in vivo* screening data available about the possible endocrine disrupting properties of a chemical. It is unlikely that no other existing endocrine-relevant data will be available (i.e. if TG 234 has been used as a primary screen), but in that case a positive result in TG 234 should ideally be followed up with relevant *in vitro* screening to confirm the suspected mode of action (MOA) before any other *in vivo* testing is considered.

410. In order to provide information relevant for assessing whether or not a chemical may fulfil the WHO/IPCS (2002) definition of an endocrine disrupter (ED), the study design has to be sufficiently robust to demonstrate the presence or absence of effects. In the dose selection, the investigator should also consider and ensure that data generated are adequate to fulfil the regulatory requirement across OECD countries as appropriate (e.g. hazard and risk assessment and labelling, ED assessment, etc.). The top dose or concentration should be sufficiently high to give clear systemic (i.e. non endocrine-specific) toxicity in order to ensure that a wide range of exposures (high to low) is tested. However, endocrine effects observed solely in the presence of clear systemic toxicity should be interpreted with caution and may be disregarded when sufficiently justified to be caused by secondary effects which are unlikely to be due to endocrine activity. The reason for this advice is a concern that some endocrine active substance (EAS) sensitive assays are being run at doses/concentrations of EASs that are too low to trigger direct impacts on the endocrine system. This guidance document is not the place to address this issue directly, but it should be considered when EAS-sensitive test guidelines (TGs) are revised in the future. In addition, the number and spacing of dose/concentration levels should also be adequate to

fulfil the objectives of the study (e.g. to demonstrate dose response relationships if this is required).

Existing data to be considered

411. Given the commonality of endocrine mechanisms in the vertebrates, relevant existing data available before deployment of OECD TG 234 (FSDT) might include *in vivo* results obtained with other vertebrates (e.g. a positive Uterotrophic Assay with rodents; positive findings for endocrine endpoints in mammalian repeat dose toxicity or reproductive studies; or a positive result in the fish assays OECD TG 229 or 230), or one or more of a range of *in silico* or *in vitro* results which suggest that the modalities indicated above may occur *in vivo*. Such indicators of possible *in vivo* activity might include quantitative structure activity relationship (QSAR) predictions of endocrine activity, high throughput screening (HTS) data, "read-across" from *in vivo* results obtained with structurally related chemicals, or positive results from an *in vitro* screen for estrogen or androgen receptor-mediated activity, or for effects on steroidogenesis (especially aromatase inhibition). Further strong indication of *in vivo* estrogenic activity may also be available from an EASZY Assay with transgenic zebrafish embryos. Conduct of OECD TG 234 (FSDT) would be particularly relevant if the test chemical is suspected to act primarily on the sexual development phase of the fish life cycle (as opposed to the reproductive phase), because it provides apical information on phenotypic sex ratio which is fixed during the fry or juvenile stages of the species used in this test.

Scenarios: Positive and negative results combined with existing data

412. The scenarios (A to R) presented in Table C.2.9 represent all the possibilities of positive or negative results in combination with the presence or absence of existing data. The action taken will also depend on the regulatory environment, but the considerations given here are generally science based. Wherever possible, the recommended "next step which could be taken" avoids unnecessary animal testing. However, sometimes conducting an animal test will be indicated and then the relevance of species, strain and exposure route should always be considered. Further considerations specific to each scenario are given in the table.

413. Positive results obtained with one or more of OECD TG 234 (FSDT) indicators of hormonal activity but not with apical endpoints (Table C.2.9, Scenarios A-I, sub-section 2) result in the conclusion that the test chemical is a potential ED *in vivo*. If both an indicator of hormonal activity and sex ratio[1] give a correlated response (Table C.2.9, Scenarios A-I, sub-section 1), this provides evidence that the chemical is probably an actual ED (i.e. it causes adverse effects through an endocrine mechanism) if adverse population effects are expected as a consequence. If only sex ratio responds (Table C.2.9, Scenarios A-I, sub-section 3), it indicates that the chemical is probably an ED, but before drawing that conclusion, existing *in vitro* and *in vivo* data should be considered and a weight of evidence assessment carried out.

414. As indicated above, an effect on sex ratio in OECD TG 234 (FSDT) shows that the test chemical causes an adverse apical effect, is a developmental toxicant and is probably also an ED (assuming that the concentration giving this response is not sufficiently high to cause systemic toxicity). If these results are combined with positive indicators of hormonal activity and/or positive *in vitro* screening assay data, some regulatory authorities may consider that this is sufficient to show the chemical is an ED, and/or that the information could be

used for hazard identification/characterisation (providing sufficient concentrations have been tested to give an acceptably precise no-observed-effect-concentration [NOEC] or x% effect concentration [ECx]). Other authorities might nevertheless require further data to demonstrate that adverse effects at lower concentrations do not occur during the reproductive phase of the life cycle, and in these circumstances, conduct of a fish life cycle test (Medaka Extended One-Generation Reproduction Test [MEOGRT] – OECD TG 240), or Zebrafish Extended One-Generation Reproduction Test [ZEOGRT]) would be appropriate. In principle, an extended version of OECD TG 229 (i.e. a Fish Reproduction Partial Life Cycle Assay) might also address this issue, but a suitable protocol for this has not been validated. Additional testing of this type might also be required if an indicator or indicators of hormonal activity in OECD TG 234 (FSDT), but not sex ratio, respond positively. Existing data suggesting endocrine activity would strengthen the case for any additional testing still further.

415. A situation in which OECD TG 234 (FSDT) gives a negative result needs careful consideration of any existing data. If these data suggest that the chemical is endocrine active both *in vitro* and *in vivo* (Table C.2.9, Scenario J), then the probability is that OECD TG 234 (FSDT) is simply insufficiently sensitive, perhaps because the main MOA acts during the reproductive phase of the life cycle. It might then be appropriate to conduct a fish life cycle test (MEOGRT – OECD TG 240, or ZEOGRT) to confirm that there is no adverse endocrine activity in fish.

416. If OECD TG 234 (FSDT) and existing *in vivo* data are all negative, but *in vitro* data reveal some endocrine activity (Scenario K), the probability is that the test chemical is not sufficiently potent to produce endocrine effects *in vivo* in fish, or it may be rapidly metabolised. In such a situation, further testing is probably not necessary. However, if there is good reason to believe that the reproductive part of the life cycle may be more responsive than sexual development, consider conducting OECD TG 229 or a life cycle test (MEOGRT – OECD TG 240, or ZEOGRT).

417. Furthermore, if OECD TG 234 (FSDT) and the *in vitro* tests are negative, but there are positive existing *in vivo* data (Scenario M), the chemical is probably not an ED acting on fish sexual development, but it may act via MOA not covered by the *in vitro* screens, or it may be more potent in species or life stages that have not been tested. In this situation, the existing *in vivo* data should be used to guide decisions about whether to conduct any further testing, either for modalities such as thyroid activity (e.g. OECD TG 231, or *Xenopus* Embryonic Thyroid Signalling Assay [XETA]), or including other life stages represented in OECD TG 229 or the MEOGRT/ZEOGRT.

418. Finally, a negative OECD TG 234 (FSDT), set against a background of negative *in vitro* and *in vivo* data (Scenario N), suggests that the test chemical is probably not an ED acting on sexual development in fish, and no further testing for estrogenic, anti-estrogenic, androgenic or steroidogenic MOA should generally be considered unless there is reason to believe that reproduction may be more responsive than development. It remains possible that the chemical has thyroid activity, but this is unlikely if OECD TG 231 or the XETA are one of the negative *in vivo* assays.

419. In each of the above scenarios, it is possible that existing data will be equivocal, or there may be no existing data (Scenarios C, F-I, L and O-R). This will weaken the conclusions which can be drawn about a negative OECD TG 234 (FSDT). However, a lack of mechanistic data on endocrine activity should ideally be rectified before any further *in vivo* testing is finally rejected. On the other hand, if OECD TG 234 (FSDT) is positive, further *in vivo* testing may be needed, even if all existing data are equivocal, or if there are

no existing data. Again, however, it will always be desirable to obtain some mechanistic information before conducting further *in vivo* testing. There is also the possibility that equivocal mechanistic data may be the result of multiple modes of endocrine action. Under some circumstances, two opposite modes of simultaneous action (e.g. estrogenic and anti-estrogenic) could, depending on dose, lead to a minimisation or abolition of adverse effects, while in others two different MOA (e.g. estrogenic and anti-androgenic) could potentially reinforce effects on certain apical endpoints. If multiple MOA are suspected, either from the existing results or based on QSAR/read-across/integrated approaches, this situation should be investigated further if needed for regulatory decision making.

420. The scenario in which the results of OECD TG 234 (FSDT) are themselves equivocal has not been dealt with in Table C.2.9, for reasons of brevity. In this context, an equivocal result might be a non-monotonic concentration-response (e.g. no effect at a high concentration but effects at a lower concentration), or a result which borders on statistical significance. Without knowing the exact circumstances, reliable advice cannot be given, but the opinions of an experienced ecotoxicologist should be sought. Clearly, however, such equivocal results do not necessarily rule out the existence of *in vivo* endocrine activity. For example, an effect on sex ratio might just fail to reach a statistically significant level due to a random imbalance in the control sex ratio. If these or other possible reasons for false negatives are suspected with good reason, the test could be repeated (e.g. conduct it at lower concentrations which avoid systemic toxicity), or a more appropriate version of it (e.g. more fish per replicate) could be conducted.

421. In summary, an adverse apical response (i.e. biased sex ratio) in OECD TG 234 (FSDT) indicates that a chemical is a probable ED. A combination of biased sex ratio and a positive endocrine-responsive mechanistic endpoint (e.g. VTG) is even stronger evidence that the chemical is an actual ED. If sufficient test concentrations have been tested, this will allow a precise NOEC or ECx to be calculated. In such cases, some regulatory authorities may consider that no more data are required, while others may wish to investigate whether the reproductive stage of the life cycle is even more sensitive than the developmental part. On the other hand, negative results in OECD TG 234 (FSDT) do not necessarily mean that the chemical is not an ED – a judgement about this will have to be made in the light of existing *in vitro* and *in vivo* data.

Note

1. Note that sex ratio can be considered as an indicator or biomarker of endocrine activity in its own right, as well as an apical measurement of adverse effects, although some types of non-EDCs may hypothetically be able to affect this endpoint in some species. None of these non-EDCs have yet been found.

References

Liu, Y.-W. and W.K. Chan (2002), "Thyroid hormones are important for embryonic to larval transitory phase in zebrafish", *Differentiation*, Vol. 70/1, pp. 36-45, https://doi.org/10.1046/j.1432-0436.2002.700104.x.

Nelson, K. et al. (2014), "Evaluation of hypothesized adverse outcome pathway linking thyroid peroxidase to fish early life stage toxicity", presentation, Society of Environmental Toxicology and Chemistry, Vancouver, British Columbia, Canada, 9-13 November 2014.

OECD (2012), "Validation report (Phase 2) for the Fish Sexual Development Test for the detection of endocrine active substances (updated 2012)", OECD Series on Testing and Assessment, No. 142, OECD, Paris, www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/JM/MONO(2011)23/REV1&doclanguage=en.

Power, D.M. et al. (2001), "Thyroid hormones in growth and development of fish", *Comparative Biochemistry and Physiology Oart C: Toxicology & Pharmacology*, Vol. 130/4, pp. 447-459, https://doi.org/10.1016/S1532-0456(01)00271-X.

Shi, X. et al. (2008), "Developmental toxicity and alteration of gene expression in zebrafish embryos exposed to PFOS", *Toxicology and Applied Pharmacology*, Vol. 230/1, pp. 23-32, https://doi.org/10.1016/j.taap.2008.01.043.

Villeneuve, S. et al. (2013), "Investigating alternatives to the fish early-life stage test: A strategy for discovering and annotating adverse outcome pathways for early fish development", *Environmental Toxicology and Chemistry*, Vol. 33/1, pp. 158-169, https://doi.org/10.1002/etc.2403.

WHO/IPCS (2002), "Global assessment of the state-of-the-science of endocrine disrupters", Damstra, T. et al. (eds.) WHO/PCS/EDC/02.2, World Health Organization, Geneva, www.who.int/ipcs/publications/new_issues/endocrine_disruptors/en.

Table C.2.9. **Fish Sexual Development Test (FSDT) (OECD TG 234): Guidance for scenarios of combinations of results with existing data**

This table represents possible conclusions to be drawn from assay data, and a next step which could be taken if further evidence is required about possible endocrine disrupting properties and/or effects. The guidance offered is not meant to be prescriptive, but provides science-based considerations. It encourages the use of all available data and expert judgement in a weight of evidence approach. Regional and national interpretation of results and "next steps" may vary.

The conclusions are grouped into a series of scenarios (A-R), each scenario representing a different combination of assay results, existing *in vitro* data and existing *in vivo* data. The symbol "+" indicates that the data in question represent a positive result, «-" indicates a negative result, and "Eq/0" indicates that the data are either equivocal or are not available.

The assay under discussion could either be positive for both apical and indicators of endocrine activity endpoints, or positive just for apical endpoints, or positive just for indicators of endocrine activity. However, note that sex ratio could in most cases be considered as both an indicator of endocrine activity and an apical endpoint, and as yet, no chemicals have been found which are able to alter sex ratios by way of mechanisms other than endocrine disruption. For each scenario, each of these three possibilities is addressed separately in the possible conclusions column.

Existing results: * "Mechanism (*in vitro* mechanistic data)" assumes that mechanistic data are available from estrogen receptor (ER-), androgen receptor (AR-) and steroidogenesis-based assays (Level 2). Thyroid hormone receptor (TR) and other assays concerning mechanisms of thyroid disruption may be available, but they are not in common use. In practice, data from all assays may not be available and therefore this must be taken into account when deciding on the "next step". Quantitative structure activity relationship (QSAR) predictions of estrogen and androgen binding/activation may be made for some substances. There is no evidence at present that equivalent *in vitro* assays with systems derived from fish offer advantages over their mammalian counterparts.

Existing results: ** "Effects (*in vivo* effects of concern)" assumes effects have been observed in other *in vivo* screens/tests which give rise to concern that the test chemical may be an endocrine disrupter.

Scenario	Result of TG 234 (FSDT)	Existing results		Possible conclusions: 1) Indicators of endocrine activity and apical endpoints positive 2) Indicators of endocrine activity positive and apical endpoints negative 3) Indicators of endocrine activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
A	+	+	+	1) Strong evidence for adverse effects in fish and other organisms by an endocrine mechanism. 2) Strong evidence for endocrine effects, but uncertainty about whether they are adverse in fish. 3) Strong evidence for adverse effects in fish and other organisms. There is a possibility that the apical endpoint sex ratio is more sensitive to the test chemical than the mechanistic endpoint vitellogenin (VTG), or mechanism may hypothetically not be via direct interaction with estrogen receptor (ER), androgen receptor (AR) or by aromatase inhibition, even though it is noted that currently there is no evidence for sex ratio change in fish caused by other mechanisms than those mentioned here at otherwise non-toxic concentrations of chemicals.	Some regulatory authorities may consider that further evidence is not required, especially if adverse effects have been demonstrated. However, if more evidence is needed about adverse effects in fish, performance of a fish life cycle test (MEOGRT – OECD TG 240, or ZEOGRT) should be considered.	If OECD TG 234 (Fish Sexual Development Test [FSDT]) was only performed with three test concentrations, this may not be sufficiently precise to establish a reliable no-observed-effect-concentration/x% effect concentration [NOEC/ECx]. Also, note that some endocrine disruptors (EDs) may be more toxic to reproduction than to sexual development, in which case TG 234 (FSDT) would be less responsive than a life cycle test (MEOGRT – OECD TG 240, or ZEOGRT).
B	+	+	–	1) Strong evidence for adverse effects in fish by an endocrine mechanism. 2) Strong evidence for endocrine effects in fish, but uncertainty about whether they are adverse. 3) Strong evidence for adverse effects in fish and other organisms. There is a possibility that the apical endpoint sex ratio is more sensitive to the test chemical than the mechanistic endpoint VTG, or mechanism may hypothetically not be via direct interaction with ER, AR or by aromatase inhibition, even though it is noted that currently there is no evidence for sex ratio change in fish caused by other mechanisms than those mentioned here at otherwise non-toxic concentrations of chemicals.	Some regulatory authorities may consider that further evidence is not required, especially if adverse effects have been demonstrated. However, if more evidence is needed about adverse effects in fish, performance of a fish life cycle test (MEOGRT – OECD TG 240, or ZEOGRT) should be considered.	If OECD TG 234 (FSDT) was only performed with three test concentrations, this may not be sufficiently precise to establish a reliable NOEC/ECx. Also, note that some EDs may be more toxic to reproduction than to sexual development, in which case TG 234 (FSDT) would be less responsive than a life cycle test (MEOGRT – OECD TG 240, or ZEOGRT).
C	+	+	Eq/0**	1) Strong evidence for adverse effects in fish by an endocrine mechanism. 2) Strong evidence for endocrine effects in fish, but uncertainty about whether they are adverse. 3) Strong evidence for adverse effects in fish and other organisms. There is a possibility that the apical endpoint sex ratio is more sensitive to the test chemical than the mechanistic endpoint VTG, or mechanism may hypothetically not be via direct interaction with ER, AR or by aromatase inhibition, even though it is noted that currently there is no evidence for sex ratio change in fish caused by other mechanisms than those mentioned here at otherwise non-toxic concentrations of chemicals.	Some regulatory authorities may consider that further evidence is not required, especially if adverse effects have been demonstrated. However, if more evidence is needed about adverse effects in fish, performance of a fish life cycle test (MEOGRT – OECD TG 240, or ZEOGRT) should be considered. This would be particularly helpful given the equivocal *in vivo* effects, or lack of *in vivo* tests, in other taxa.	If OECD TG 234 (FSDT) was only performed with three test concentrations, this may not be sufficiently precise to establish a reliable NOEC/ECx. Also, note that some EDs may be more toxic to reproduction than to sexual development, in which case TG 234 (FSDT) would be less responsive than a life cycle test (MEOGRT – OECD TG 240, or ZEOGRT). It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple modes of action (MOA). If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

Scenario	Result of TG 234 (FSDT)	Existing results		Possible conclusions: 1) Indicators of endocrine activity and apical endpoints positive 2) Indicators of endocrine activity positive and apical endpoints negative 3) Indicators of endocrine activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
D	+	–	+	1) Strong evidence for adverse effects in fish and other organisms, possibly by an unknown endocrine mechanism. 2) Medium-strong evidence for endocrine effects in fish, but they do not appear to be adverse. 3) Strong evidence for adverse effects in fish and other organisms. There is a possibility that the apical endpoint sex ratio is more sensitive to the test chemical than the mechanistic endpoint VTG, or mechanism may hypothetically not be via direct interaction with ER, AR or by aromatase inhibition, even though it is noted that currently there is no evidence for sex ratio change in fish caused by other mechanisms than those mentioned here at otherwise non-toxic concentrations of chemicals.	Some regulatory authorities may consider that further evidence is not required, especially if adverse effects have been demonstrated. However, if more evidence is needed about adverse effects in fish, performance of a fish life cycle test (MEOGRT – OECD TG 240, or ZEOGRT) should be considered.	If OECD TG 234 (FSDT) was only performed with three test concentrations, this may not be sufficiently precise to establish a reliable NOEC/ECx. Also, note that some EDs may be more toxic to reproduction than to sexual development, in which case TG 234 (FSDT) would be less responsive than a life cycle test (MEOGRT – OECD TG 240, or ZEOGRT). If *in vitro* data are negative or equivocal, it might be unsafe to conclude that an effect on sex ratio was definitely caused by endocrine disruption, although this is the most probable explanation, especially if endocrine disruption has been shown in other species.
E	+	–	–	1) Strong evidence for adverse effects in fish, possibly by an unknown endocrine mechanism. 2) Medium-strong evidence for endocrine effects in fish, but they do not appear to be adverse. 3) Strong evidence for adverse effects in fish and other organisms. There is a possibility that the apical endpoint sex ratio is more sensitive to the test chemical than the mechanistic endpoint VTG, or mechanism may hypothetically not be via direct interaction with ER, AR or by aromatase inhibition, even though it is noted that currently there is no evidence for sex ratio change in fish caused by other mechanisms than those mentioned here at otherwise non-toxic concentrations of chemicals.	Some regulatory authorities may consider that further evidence is not required, especially if adverse effects have been demonstrated. However, if more evidence is needed about adverse effects in fish, performance of a fish life cycle test (MEOGRT – OECD TG 240, or ZEOGRT) should be considered.	If OECD TG 234 (FSDT) was only performed with three test concentrations, this may not be sufficiently precise to establish a reliable NOEC/ECx. Also, note that some EDs may be more toxic to reproduction than to sexual development, in which case TG 234 (FSDT) would be less responsive than a life cycle test (MEOGRT – OECD TG 240, or ZEOGRT). If *in vitro* data are negative or equivocal, it might be unsafe to conclude that an effect on sex ratio was definitely caused by endocrine disruption, although this is the most probable explanation, especially if endocrine disruption has been shown in other species.

Scenario	Result of TG 234 (FSDT)	Existing results		Possible conclusions: 1) Indicators of endocrine activity and apical endpoints positive 2) Indicators of endocrine activity positive and apical endpoints negative 3) Indicators of endocrine activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
F	+	–	Eq/0	1) Strong evidence for adverse effects in fish, possibly by an unknown endocrine mechanism. 2) Medium-strong evidence for endocrine effects in fish, but they do not appear to be adverse. 3) Strong evidence for adverse effects in fish and other organisms. There is a possibility that the apical endpoint sex ratio is more sensitive to the test chemical than the mechanistic endpoint VTG, or mechanism may hypothetically not be via direct interaction with ER, AR or by aromatase inhibition, even though it is noted that currently there is no evidence for sex ratio change in fish caused by other mechanisms than those mentioned here at otherwise non-toxic concentrations of chemicals.	Some regulatory authorities may consider that further evidence is not required, especially if adverse effects have been demonstrated. However, if more evidence is needed about adverse effects in fish, performance of a fish life cycle test (MEOGRT – OECD TG 240, or ZEOGRT) should be considered. This would be particularly helpful given the equivocal *in vivo* effects, or lack of *in vivo* tests, in other taxa.	If OECD TG 234 (FSDT) was only performed with three test concentrations, this may not be sufficiently precise to establish a reliable NOEC/ECx. Also, note that some EDs may be more toxic to reproduction than to sexual development, in which case TG 234 (FSDT) would be less responsive than a life cycle test (MEOGRT – OECD TG 240, or ZEOGRT). If *in vitro* data are negative or equivocal, it might be unsafe to conclude that an effect on sex ratio was definitely caused by endocrine disruption, although this seems the most probable explanation, especially if endocrine disruption has been shown in other species. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
G	+	Eq/0	+	1) Strong evidence for adverse effects in more than one organism, possibly by an unknown endocrine mechanism. 2) Medium-strong evidence for endocrine effects, but they do not appear to be adverse in fish. 3) Strong evidence for adverse effects in fish and other organisms. There is a possibility that the apical endpoint sex ratio is more sensitive to the test chemical than the mechanistic endpoint VTG, or mechanism may hypothetically not be via direct interaction with ER, AR or by aromatase inhibition, even though it is noted that currently there is no evidence for sex ratio change in fish caused by other mechanisms than those mentioned here at otherwise non-toxic concentrations of chemicals.	Some regulatory authorities may consider that further evidence is not required, especially if adverse effects have been demonstrated. However, if more evidence is needed about adverse effects in fish, performance of a fish life cycle test (MEOGRT – OECD TG 240, or ZEOGRT) should be considered. Given uncertainty about the mechanism of action, any further *in vivo* testing should be preceded by *in vitro* mechanistic studies.	If OECD TG 234 (FSDT) was only performed with three test concentrations, this may not be sufficiently precise to establish a reliable NOEC/ECx. Also, note that some EDs may be more toxic to reproduction than to sexual development, in which case TG 234 (FSDT) would be less responsive than a life cycle test (MEOGRT – OECD TG 240, or ZEOGRT). If *in vitro* data are negative or equivocal, it might be unsafe to conclude that an effect on sex ratio was definitely caused by endocrine disruption, although this seems the most probable explanation, especially if endocrine disruption has been shown in other species. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

Scenario	Result of TG 234 (FSDT)	Existing results		Possible conclusions: 1) Indicators of endocrine activity and apical endpoints positive 2) Indicators of endocrine activity positive and apical endpoints negative 3) Indicators of endocrine activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
H	+	Eq/0	–	1) Strong evidence for adverse effects in fish, possibly by an unknown endocrine mechanism. 2) Medium-strong evidence for endocrine effects in fish, but they do not appear to be adverse. 3) Strong evidence for adverse effects in fish and other organisms. There is a possibility that the apical endpoint sex ratio is more sensitive to the test chemical than the mechanistic endpoint VTG, or mechanism may hypothetically not be via direct interaction with ER, AR or by aromatase inhibition, even though it is noted that currently there is no evidence for sex ratio change in fish caused by other mechanisms than those mentioned here at otherwise non-toxic concentrations of chemicals.	Some regulatory authorities may consider that further evidence is not required, especially if adverse effects have been demonstrated. However, if more evidence is needed about adverse effects in fish, performance of a fish life cycle test (MEOGRT – OECD TG 240, or ZEOGRT) should be considered. Given uncertainty about the mechanism of action, any further *in vivo* testing should be preceded by *in vitro* mechanistic studies.	If OECD TG 234 (FSDT) was only performed with three test concentrations, this may not be sufficiently precise to establish a reliable NOEC/ECx. Also, note that some EDs may be more toxic to reproduction than to sexual development, in which case TG 234 (FSDT) would be less responsive than a life cycle test (MEOGRT – OECD TG 240, or ZEOGRT). If *in vitro* data are negative or equivocal, it might be unsafe to conclude that an effect on sex ratio was definitely caused by endocrine disruption, although this seems the most probable explanation, especially if endocrine disruption has been shown in other species. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
I	+	Eq/0	Eq/0	1) Strong evidence for adverse effects in fish, possibly by an unknown endocrine mechanism. 2) Moderate-strong evidence for endocrine effects in fish, but they do not appear to be adverse. 3) Strong evidence for adverse effects in fish and other organisms. There is a possibility that the apical endpoint sex ratio is more sensitive to the test chemical than the mechanistic endpoint VTG, or mechanism may hypothetically not be via direct interaction with ER, AR or by aromatase inhibition, even though it is noted that currently there is no evidence for sex ratio change in fish caused by other mechanisms than those mentioned here at otherwise non-toxic concentrations of chemicals.	Some regulatory authorities may consider that further evidence is not required, especially if adverse effects have been demonstrated. However, if more evidence is needed about adverse effects in fish, performance of a fish life cycle test (MEOGRT – OECD TG 240, or ZEOGRT) should be considered. Given uncertainty about the mechanism of action, any further *in vivo* testing should be preceded by *in vitro* mechanistic studies.	If OECD TG 234 (FSDT) was only performed with three test concentrations, this may not be sufficiently precise to establish a reliable NOEC/ECx. Also, note that some EDs may be more toxic to reproduction than to sexual development, in which case TG 234 (FSDT) would be less responsive than a life cycle test (MEOGRT – OECD TG 240, or ZEOGRT). If *in vitro* data are negative or equivocal, it might be unsafe to conclude that an effect on sex ratio was definitely caused by endocrine disruption, although this seems the most probable explanation, especially if endocrine disruption has been shown in other species. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity, lack of metabolic activation or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information

Scenario	Result of TG 234 (FSDT)	Existing results		Possible conclusions: 1) Indicators of endocrine activity and apical endpoints positive 2) Indicators of endocrine activity positive and apical endpoints negative 3) Indicators of endocrine activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
J	–	+	+	The chemical is an ED *in vivo* in other species but does not appear to act on sexual development in fish. If any other fish tests are also negative, fish may not be responsive at all to the test chemical.	Some regulatory authorities may consider that further evidence is not required. However, if it is suspected that the reproductive part of the life cycle may be responsive, consider conducting OECD TG 229 or a fish life cycle test (MEOGRT – OECD TG 240, or ZEOGRT).	As OECD TG 229 only uses three test concentrations and exposes fish for just three weeks, an extended version which runs more concentrations for longer would provide more comprehensive data about interference with reproduction than OECD TG 229 unmodified. However, an agreed protocol for such an extended test is not available, so an option would be to run a fish life cycle test (MEOGRT – OECD TG 240, or ZEOGRT).
K	–	+	–	Despite the *in vitro* mechanistic data for potential endocrine activity, there is no evidence for endocrine disruption *in vivo*. This may be because the chemical is quickly transformed/degraded to an inactive metabolite, or because it only interacts very weakly with the endocrine system. However, it is also possible that the chemical only acts on the reproductive part of the fish life cycle which is not exposed in OECD TG 234 (FSDT).	Some regulatory authorities may consider that sufficient data are available to show that the chemical is not an ED *in vivo*. However, if it is suspected that the reproductive part of the life cycle may be responsive, consider conducting either OECD TG 229 or a fish life cycle test (MEOGRT – OECD TG 240, or ZEOGRT).	As OECD TG 229 only uses three test concentrations and exposes fish for just three weeks, an extended version which runs more concentrations for longer would provide more comprehensive data about interference with reproduction than OECD TG 229 unmodified. However, an agreed protocol for such an extended test is not available, so an option would be to run a fish life cycle test (MEOGRT – OECD TG 240, or ZEOGRT).
L	–	+	Eq/0	The chemical may not be an ED *in vivo*, but the confidence in this conclusion is relatively low as there is only one unequivocal *in vivo* test result (the negative OECD TG 234). However, it is also possible that the chemical only acts on the reproductive part of the fish life cycle which is not exposed in TG 234 (FSDT).	Some regulatory authorities may consider that sufficient data are available to show that the chemical is not an ED *in vivo*. However, such a conclusion is not well supported. If it is suspected that the reproductive part of the life cycle may be responsive, consider conducting either OECD TG 229 or a fish life cycle test (MEOGRT – OECD TG 240, or ZEOGRT).	As OECD TG 229 only uses three test concentrations and exposes fish for just three weeks, an extended version which runs more concentrations for longer would provide more comprehensive data about interference with reproduction than OECD TG 229 unmodified. However, an agreed protocol for such an extended test is not available, so an option would be to run a fish life cycle test (MEOGRT – OECD TG 240, or ZEOGRT). It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

Scenario	Result of TG 234 (FSDT)	Existing results		Possible conclusions: 1) Indicators of endocrine activity and apical endpoints positive 2) Indicators of endocrine activity positive and apical endpoints negative 3) Indicators of endocrine activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
M	–	–	+	The chemical is probably not an ED acting on sexual development in fish, but it does have endocrine activity in other species. However, it may act through MOA not covered by the available *in vitro* assays, or it may be more potent in a fish species other than that tested. It is also possible that the chemical only acts on the reproductive part of the fish life cycle which is not exposed in OECD TG 234 (FSDT), although such action is presumably not via one of the mechanisms mentioned above.	Some regulatory authorities may consider that sufficient evidence is available. However, if it is suspected that the reproductive part of the life cycle may be responsive, consider conducting either OECD TG 229 or a fish life cycle test (MEOGRT – OECD TG 240, or ZEOGRT), possibly using a different species to that employed in OECD TG 234 (FSDT). If thyroid activity is suspected, it may be helpful to conduct an Amphibian Metamorphosis Assay (TG 231) or *Xenopus* Embryonic Thyroid Signalling Assay.	As OECD TG 229 only uses three test concentrations and exposes fish for just three weeks, an extended version which runs more concentrations for longer would provide more comprehensive data about interference with reproduction than OECD TG 229 unmodified. However, an agreed protocol for such an extended test is not available, so an option would be to run a fish life cycle test (MEOGRT – OECD TG 240, or ZEOGRT).
N	–	–	–	The chemical is probably not an ED acting on sexual development in fish, or *in vivo* in other species. It is possible that the chemical is able to interfere with the reproductive part of the fish life cycle but the probability of this is low given the apparent absence of estrogenic, androgenic or steroidogenic properties *in vitro* or *in vivo*.	Some regulatory authorities may consider that sufficient data are available to show that the chemical is not an ED *in vivo*. However, if it is suspected that the reproductive part of the life cycle may be responsive, consider conducting either OECD TG 229 or a fish life cycle test (MEOGRT – OECD TG 240, or ZEOGRT).	As OECD TG 229 only uses three test concentrations and exposes fish for just three weeks, an extended version which runs more concentrations for longer would provide more comprehensive data about interference with reproduction than OECD TG 229 unmodified. However, an agreed protocol for such an extended test is not available, so an option would be to run a fish life cycle test (MEOGRT – OECD TG 240, or ZEOGRT).
O	–	–	Eq/0	The chemical is probably not an ED acting on sexual development in fish. It is possible that the chemical is able to interfere with the reproductive part of the fish life cycle, but the probability of this is low given the apparent absence of estrogenic, androgenic or steroidogenic properties.	Some regulatory authorities may consider that sufficient data are available to show that the chemical is not an ED *in vivo*. However, if it is suspected that the reproductive part of the life cycle may be responsive, consider conducting either OECD TG 229 or a fish life cycle test (MEOGRT – OECD TG 240, or ZEOGRT).	As OECD TG 229 only uses three test concentrations and exposes fish for just three weeks, an extended version which runs more concentrations for longer would provide more comprehensive data about interference with reproduction than OECD TG 229 unmodified. However, an agreed protocol for such an extended test is not available, so an option would be to run a fish life cycle test (MEOGRT – OECD TG 240, or ZEOGRT). It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

Scenario	Result of TG 234 (FSDT)	Existing results		Possible conclusions: 1) Indicators of endocrine activity and apical endpoints positive 2) Indicators of endocrine activity positive and apical endpoints negative 3) Indicators of endocrine activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
P	–	Eq/0	+	The chemical is probably not an ED acting on sexual development in fish, but confidence in this conclusion is low given the lack of comprehensive *in vitro* data and the availability of positive existing *in vivo* data. However, it is possible that the chemical only acts on the reproductive part of the fish life cycle which is not exposed in OECD TG 234 (FSDT).	Some regulatory authorities may consider that sufficient evidence is available. However, if it is suspected that the reproductive part of the life cycle may be responsive, consider conducting either OECD TG 229 or a fish life cycle test (MEOGRT – OECD TG 240, or ZEOGRT). However, it would be desirable to obtain comprehensive mechanistic data before possibly proceeding to further *in vivo* testing.	As OECD TG 229 only uses three test concentrations and exposes fish for just three weeks, an extended version which runs more concentrations for longer would provide more comprehensive data about interference with reproduction than OECD TG 229 unmodified. However, an agreed protocol for such an extended test is not available, so an option would be to run a fish life cycle test (MEOGRT – OECD TG 240, or ZEOGRT). It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
Q	–	Eq/0	–	The chemical is probably not an ED acting on sexual development in fish, or *in vivo* on other species, but the lack of more predictive mechanistic data are a concern, even though the existing *in vivo* data are negative. It is nevertheless possible that the chemical is able to interfere with the reproductive part of the fish life cycle.	Some regulatory authorities may consider that sufficient data are available to show that the chemical is not an ED *in vivo*. However, if it is suspected that the reproductive part of the life cycle may be responsive, consider conducting either OECD TG 229 or a fish life cycle test (MEOGRT – OECD TG 240, or ZEOGRT). It would be desirable to obtain comprehensive mechanistic data before any further *in vivo* testing.	As OECD TG 229 only uses three test concentrations and exposes fish for just three weeks, an extended version which runs more concentrations for longer would provide more comprehensive data about interference with reproduction than OECD TG 229 unmodified. However, an agreed protocol for such an extended test is not available, so an option would be to run a fish life cycle test (MEOGRT – OECD TG 240, or ZEOGRT). It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
R	–	Eq/0	Eq/0	The chemical may not be an ED acting on sexual development in fish, but confidence in this conclusion is low given the lack of comprehensive *in vitro* and existing *in vivo* data. It is nevertheless possible that the chemical is able to interfere with the reproductive part of the fish life cycle.	Some regulatory authorities may consider that sufficient data are available to show that the chemical is not an ED *in vivo*. However, if it is suspected that the reproductive part of the life cycle may be responsive, consider conducting either OECD TG 229 or a fish life cycle test (MEOGRT – OECD TG 240, or ZEOGRT). However, it would be desirable to obtain comprehensive mechanistic data before any further *in vivo* testing.	As OECD TG 229 only uses three test concentrations and exposes fish for just three weeks, an extended version which runs more concentrations for longer would provide more comprehensive data about interference with reproduction than OECD TG 229 unmodified. However, an agreed protocol for such an extended test is not available, so an option would be to run a fish life cycle test (MEOGRT – OECD TG 240, or ZEOGRT). It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

C.2.10. Larval Amphibian Growth and Development Assay (LAGDA) (OECD TG 241)

Status: Assay validated by the OECD.

422. Modality detected/endpoints: OECD TG 241 has three endpoints indicating generalised toxicity (mortality, abnormal behaviour and growth), and several providing specific information about endocrine disruption or impaired reproduction (histopathology of thyroid, gonads, kidney and liver, time to metamorphosis [NF stage 62]; secondary sex characteristics (nuptial pads); vitellogenin (optional); genetic and phenotypic sex ratio). Most of these specific endocrine endpoints are likely to respond to interference with the hypothalamic/pituitary/gonadal (HPG) axis, while thyroid histopathology and time to metamorphosis may respond to interference with the hypothalamic/pituitary/thyroid axis (as may the "generalised toxicity" indicator, growth).

Background to the assay

423. This assay is a partial life cycle test with the clawed frog *Xenopus laevis*. The LAGDA was performed adequately to evaluate apical effects of chronic exposure to two endocrine-active compounds (Haselman et al., 2016a; 2016b). It starts with NF stage 8 F0 larvae and ends 10 weeks after the median time that controls take to reach NF stage 62 F0 juveniles (typically a total of 16 weeks). In essence, therefore, it covers the stages of larval/juvenile growth and sexual development, but not those of reproduction and embryonic development. It could therefore be thought of as the amphibian near-equivalent of the Fish Sexual Development Test (FSDT – OECD TG 234), although it also includes endpoints that are specifically responsive to thyroid disrupters. It does not include all processes which may respond to estrogen/androgen/thyroid/steroidogenesis (E,A,T,S) endocrine disruptors (EDs) (especially reproduction), and it is currently unknown whether the LAGDA is therefore less responsive to some of these chemicals than an amphibian life cycle test (a standardised protocol which is not available). It may ultimately be concluded that a fish life cycle test may be a more suitable test than the LAGDA or an amphibian life cycle test for long-term evaluation of chemicals with E,A,S properties despite the fact that the LAGDA can detect certain estrogenic active substances. Due to the LAGDA's higher sensitivity towards thyroid (T) properties, this test may be the preferred testing choice for confirming T properties.

424. OECD TG 241 provides a table of endpoints (test guidelines Table 1), some of which are "apical", while others should more properly be considered as indicators of hormonal activity. Probably the only true apical endpoints which could be used for hazard identification/characterisation (because they can be related directly to adverse effects on populations) are mortality, growth and phenotypic/genotypic sex ratio. The latter two are likely to be responsive to some EDs, but growth may also respond to certain other chemicals. On the other hand, indicators of hormonal activity of use in diagnosing the effects of EDs include gonad and thyroid histopathology, liver-somatic index, time to metamorphosis, and vitellogenin (VTG). Time to metamorphosis can also arguably be

considered as an apical endpoint with potential implications at the population level. The endpoints will be grouped in this way for the purposes of this document.

425. Consequently, if the assay gives a positive result, this may be due to a combination of a positive indicator of hormonal activity (gonad and thyroid histopathology, liver-somatic index, time to metamorphosis, and vitellogenin) and a positive apical endpoint (sex ratio and possibly growth), or a positive for an indicator of hormonal activity alone, or for an apical endpoint alone. Each of these possible combinations of positive response should be considered separately (although the distinctions between indicators of hormonal activity and apical effects are not always clear), so they have been listed individually as points 1, 2 and 3 in the possible conclusions column of Table C.2.10. Given the high degree of endocrine system conservation across the vertebrates, adverse endocrine-linked effects in the LAGDA may also indicate the possibility of related activity in other organisms such as fish, reptiles, birds or mammals.

When/why the assay may be used

426. Although the LAGDA could, in principle, be used at any stage in the hazard assessment process, the most likely use scenario will be when there are some data available about the possible thyroid disrupting properties of a chemical, or if the chemical is suspected of having (anti)estrogenic or (anti)androgenic properties. Thus, there are likely to be data available from *in vitro* mechanistic screens, as well as *in vivo* non-mammalian wildlife screens such as OECD TG 229, TG 230, TG 231, *Xenopus* Embryonic Thyroid Signalling Assay (XETA) or EASZY. Furthermore, a number of mammalian (rat) assays (which may have been performed before any non-mammalian wildlife testing) are sensitive to thyroid disruption, including the pubertal assay (male or female), the enhanced repeat dose assay (OECD TG 407), and the intact male screening assay. Rodent screening assays (e.g. the Hershberger or Uterotrophic Bioassays) with responsiveness to other EDs (e.g. androgens or estrogens) may also have been conducted.

427. It is unlikely that no endocrine-relevant data will available before the LAGDA is deployed (i.e. the LAGDA has been used as a primary screen), but in that case a positive result in the LAGDA could be followed up with relevant *in vitro* assays to investigate the suspected mode of action (MOA). However, it should be noted that while *in vitro* assays are available for estrogens, androgens and steroidogenesis inhibitors, they additionally exist only for thyroid agonists and antagonists (e.g. GH_3 rat pituitary somatotroph cell proliferation; solid state thyroid receptor binding assays; transfected reporter gene assays in yeast or mammalian cell lines), while thyroid disruption can occur at other points in the endocrine system for which *in vitro* assays do not exist, or are still at the research stage (e.g. FRTL-5 rat cell lines sensitive to iodide uptake inhibitors). Furthermore, none of these *in vitro* thyroid assays have yet been validated and standardised at the international level, although several are in development.

428. In order to provide information relevant for assessing whether or not a chemical may fulfil the WHO/IPCS (2002) definition of an ED, the study design has to be sufficiently robust to demonstrate the presence or absence of effects. In the dose selection, the investigator should also consider and ensure that data generated are adequate to fulfil the regulatory requirement across OECD countries as appropriate (e.g. hazard and risk assessment and labelling, ED assessment, etc.). The top dose or concentration should be sufficiently high to give clear systemic (i.e. non endocrine-specific) toxicity in order to ensure that a wide range of exposures (high to low) is tested. However, endocrine effects observed solely in the presence of clear systemic toxicity should be interpreted with caution

and may be disregarded when sufficiently justified to be caused by secondary effects which are unlikely to be due to endocrine activity. The reason for this advice is a concern that some endocrine active substance (EAS) sensitive assays are being run at doses/concentrations of EASs that are too low to trigger direct impacts on the endocrine system. This guidance document is not the place to address this issue directly, but it should be considered when EAS-sensitive test guidelines (TGs) are revised in the future. In addition, the number and spacing of dose/concentration levels should also be adequate to fulfil the objectives of the study (e.g. to demonstrate dose response relationships if this is required).

Existing data to be considered

429. Given the commonality of endocrine mechanisms in the vertebrates, relevant existing data available before deployment of the LAGDA might include *in vivo* results obtained with other vertebrates (e.g. a positive *in vivo* assay with rats or fish – see above), or one or more of a range of *in silico* or *in vitro* results which suggest that estrogenic, androgenic or thyroid disruption may occur *in vivo* (but note the limitations of this approach for thyroid disrupters, as indicated above). Such indicators of possible endocrine activity might include quantitative structure activity relationship (QSAR) predictions, "read-across" from *in vivo* results obtained with structurally related chemicals or positive results from an *in vitro* screen. Further strong indication of *in vivo* estrogenic activity may also be available from an EASZY Assay with transgenic zebrafish embryos, and evidence for thyroid activity could additionally be available from a *Xenopus* Embryonic Thyroid Signalling Assay (XETA).

Scenarios: Positive and negative results combined with existing data

430. The scenarios (A to R) presented in Table C.2.10 represent all the possibilities of positive or negative results in combination with the presence or absence of existing data. The action taken will also depend on the regulatory environment, but the considerations given here are generally science based. Wherever possible, the recommended "next step which could be taken" avoids unnecessary animal testing. However, sometimes conducting an animal test will be indicated and then the relevance of species, strain and exposure route should always be considered. Further considerations specific to each scenario are given in the table.

431. Positive results obtained with an indicator of hormonal activity in the LAGDA but not with apical endpoints (Table C.2.10, Scenarios A-I, sub-section 2) result in the conclusion that the test chemical is probably a potential ED *in vivo*. If both an indicator of hormonal activity and an apical endpoint give a response (Table C.2.10, Scenarios A-I, sub-section 1), this provides evidence that one is dealing with an actual ED with adverse effects *in vivo* if adverse population effects are expected as a consequence. If only an apical endpoint responds (Table C.2.10, Scenarios A-I, sub-section 3), it suggests that the chemical is harmful to growth or sexual development, but is not necessarily an ED (although existing positive *in vitro* data, or positive *in vivo* data from other species, would have to be weighed against this conclusion).

432. The situation in which a LAGDA gives a negative result (Table C.2.10, Scenarios J-R) needs careful consideration of any existing data. If these data suggest that the chemical is endocrine active both *in vitro* and *in vivo* (Scenario J), then it is possible that the LAGDA is simply insufficiently sensitive (perhaps because it does not include reproduction).

Depending on the robustness of the existing data, it might therefore be appropriate to conduct an amphibian life cycle test, although a protocol for one has not been standardised or validated.

433. If the LAGDA and existing *in vivo* data are all negative, but *in vitro* data reveal some endocrine activity (Scenario K), the probability is that the test chemical is not sufficiently potent to produce effects *in vivo* in amphibians or other organisms, or it may be rapidly metabolised. In such a situation, further testing is probably not necessary.

434. On the other hand, if the LAGDA and the *in vitro* tests are negative (Scenario M), but there are positive existing *in vivo* data, the nature of those existing data should be considered. Unless the existing data are from another amphibian, the chemical is probably not an ED acting on amphibian growth or development, but it may act via MOA not covered by the *in vitro* screens, or it may be more potent in species or life stages that have not been tested. In this situation, the existing *in vivo* data should be used to guide decisions about whether to conduct any further testing.

435. Finally, a negative LAGDA, set against a background of negative *in vitro* and *in vivo* data (Scenario N), suggests that the test chemical is not a possible E,A,T,S ED, and further action is unnecessary.

436. In each of the above scenarios, it is possible that existing data will be equivocal, or there may be no existing data. This will weaken the conclusions which can be drawn about a negative LAGDA, and this is reflected in Table C.2.10. However, a lack of *in vitro* mechanistic data should ideally be rectified before any further *in vivo* testing is finally rejected, although as indicated above, many thyroid modalities are not detectable in *in vitro* screens. On the other hand, if the LAGDA is positive, further *in vivo* testing would not generally be needed unless it is suspected that the chemical acts primarily on reproduction. Again, however, it may be useful to obtain some mechanistic information before conducting further *in vivo* testing, although note that a validated amphibian life cycle protocol is unavailable. A possible substitute for the latter might be a fish life cycle test (either the MEOGRT or ZEOGRT), although the responsiveness of such a procedure to thyroid disrupters is unknown. There is also the possibility that equivocal mechanistic data may be the result of multiple modes of endocrine action. Under some circumstances, two opposite modes of simultaneous action (e.g. estrogenic and anti-estrogenic) could, depending on dose, lead to a minimisation or abolition of adverse effects, while in others two different MOA (e.g. estrogenic and anti-androgenic) could potentially reinforce effects on certain apical endpoints. If multiple MOA are suspected, either from the existing results or based on QSAR/read-across/integrated approaches, this situation should be investigated further if needed for regulatory decision making.

437. The scenario in which the results of a LAGDA are themselves equivocal has not been dealt with in Table C.2.10, for reasons of brevity. In this context, an equivocal result might be an inconsistent concentration-response (e.g. no effect at a high concentration but effects at a lower concentration), or a result which borders on statistical significance. Without knowing the exact circumstances, reliable advice cannot be given, but the opinions of an experienced ecotoxicologist should be sought. Clearly, however, such equivocal results do not necessarily rule out the existence of *in vivo* endocrine activity. For example, thyroid histopathology at a high concentration might be masked by any systemic toxicity, while growth measurements might just fail to reach a statistically significant level due to unexpectedly high variability. If these or other possible reasons for false negatives are suspected with good reason, the test could be repeated (e.g. conduct it at lower

concentrations which avoid systemic toxicity), or a more appropriate version of it (e.g. more larvae per replicate) could be designed and conducted.

438. In summary, positive indicators of hormonal activity in the LAGDA indicate that a chemical is a potential ED via one of several modalities, while a combination of positive indicators of hormonal activity and positive apical results suggest that it is an actual ED (especially if the two types of response are causally related). However, if an apical endpoint alone responds, the chemical may not be an ED (although existing data may help to inform this decision). Negative results in the LAGDA do not necessarily mean that the chemical is not an ED – a judgement about possible endocrine disruption and the possible need for additional testing will have to be made in the light of existing *in vitro* and *in vivo* data.

References

Haselman, J.T. et al. (2016a), "Development of the Larval Amphibian Growth and Development Assay: Effects of chronic 4-tert-octylphenol or 17β-trenbolone exposure in *Xenopus laevis* from embryo to juvenile", *Journal of Applied Toxicology*, Vol. 36/12, pp. 1639-1650, https://doi.org/10.1002/jat.3330.

Haselman, J.T. et al. (2016b), "Development of the Larval Amphibian Growth and Development Assay: Effects of benzophenone-2 exposure in *Xenopus laevis* from embryo to juvenile", *Journal of Applied Toxicology*, Vol. 36/12, pp. 1651-1661, https://doi.org/10.1002/jat.3336.

WHO/IPCS (2002), "Global assessment of the state-of-the-science of endocrine disrupters", Damstra, T. et al. (eds.) WHO/PCS/EDC/02.2, World Health Organization, Geneva, www.who.int/ipcs/publications/new_issues/endocrine_disruptors/en.

Table C.2.10. **Larval Amphibian Growth and Development Assay (LAGDA) (OECD TG 241): Guidance for scenarios of combinations of results with existing data**

This table represents possible conclusions to be drawn from assay data, and a next step which could be taken if further evidence is required about possible endocrine disrupting properties and/or effects. The guidance offered is not meant to be prescriptive, but provides science-based considerations. It encourages the use of all available data and expert judgement in a weight of evidence approach. Regional and national interpretation of results and "next steps" may vary.

The conclusions are grouped into a series of scenarios (A-R), each scenario representing a different combination of assay results, existing *in vitro* data and existing *in vivo* data. The symbol "+" indicates that the data in question represent a positive result, «-" indicates a negative result, and "Eq/0" indicates that the data are either equivocal or are not available.

Existing results: * "Mechanism (*in vitro* mechanistic data)" assumes that mechanistic data are available from thyroid hormone receptor (TR) and other assays concerning mechanisms of thyroid disruption although these are not yet in common use. In practice, data from all assays may not be available and therefore this must be taken into account when deciding on the "next step". Quantitative structure activity relationship (QSAR) predictions of TR binding/activation may be made for some substances.

Existing results: ** "Effects (*in vivo* effects of concern)" assumes effects have been observed in other *in vivo* screens/tests which give rise to concern that the test chemical may be an thyroid disrupter.

The assay under discussion could either be positive for both apical endpoints and indicators of endocrine activity, or positive just for an apical endpoints or indicators of endocrine activity. For each scenario, each of these three possibilities is addressed separately in the possible conclusions column.

Scenarios	Result of LAGDA	Existing results		Possible conclusions: 1) Indicators of endocrine activity and apical endpoints positive 2) Indicators of endocrine activity positive 3) Apical endpoint positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
A	+	+	+	1) Strong evidence for *in vivo* endocrine activity with adverse effects (on growth or sexual development) in amphibians, and effects in other species. 2) Strong evidence for *in vivo* endocrine activity in amphibians and other species. 3) Strong evidence for adverse effects on growth or sexual development in amphibians, and effects in other species, but possibly not via an endocrine mechanism in the case of growth.	Regulatory authorities may consider that further data from amphibians are not required. However, see right-hand column.	The Larval Amphibian Growth and Development Assay (LAGDA) does not cover the reproductive phase of the life cycle, but a life cycle test which could be used to address any concerns about reproduction is not currently available, although a fish life cycle test (MEOGRT) could provide useful information.
B	+	+	–	1) Strong evidence for *in vivo* endocrine activity with adverse effects (on growth or sexual development) in amphibians. 2) Strong evidence for *in vivo* endocrine activity in amphibians. 3) Strong evidence for adverse effects on growth or sexual development in amphibians, but possibly not via an endocrine mechanism in the case of growth.	Regulatory authorities may consider that further data from amphibians are not required. However, see right-hand column.	The LAGDA does not cover the reproductive phase of the life cycle, but a life cycle test which could be used to address any concerns about reproduction is not currently available, although a fish life cycle test (MEOGRT) could provide useful information.
C	+	+	Eq/0	1) Strong evidence for *in vivo* endocrine activity with adverse effects (on growth or sexual development) in amphibians. 2) Strong evidence for *in vivo* endocrine activity in amphibians. 3) Strong evidence for adverse effects on growth or sexual development in amphibians, but possibly not via an endocrine mechanism in the case of growth.	Regulatory authorities may consider that further data from amphibians are not required. However, see right-hand column.	The LAGDA does not cover the reproductive phase of the life cycle, but a life cycle test which could be used to address any concerns about reproduction is not currently available, although a fish life cycle test (MEOGRT) could provide useful information. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple modes of action (MOA). If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
D	+	–	+	1) Moderate evidence for *in vivo* endocrine activity with adverse effects (on growth or sexual development) in amphibians and other species, but possibly not via an estrogen/androgen/thyroid/steroidogenesis (E,A,T,S) mechanism. 2) Moderate evidence for *in vivo* endocrine activity in amphibians and other species, but possibly not via an E,A,T,S mechanism. 3) Strong evidence for adverse effects on growth or sexual development in amphibians and other species, but probably not via an endocrine mechanism in the case of growth.	Regulatory authorities may consider that further data from amphibians are not required. However, see right-hand column.	The LAGDA does not cover the reproductive phase of the life cycle, but a life cycle test which could be used to address any concerns about reproduction is not currently available, although a fish life cycle test (MEOGRT) could provide useful information.

Scenarios	Result of LAGDA	Existing results		Possible conclusions: 1) Indicators of endocrine activity and apical endpoints positive 2) Indicators of endocrine activity positive 3) Apical endpoint positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
E	+	–	–	1) Strong evidence for *in vivo* endocrine activity with adverse effects (on growth or sexual development) in amphibians, but possibly not via an E,A,T,S mechanism. 2) Strong evidence for *in vivo* endocrine activity in amphibians, but possibly not via an E,A,T,S mechanism. 3) Strong evidence for adverse effects on growth or sexual development in amphibians, but probably not via an endocrine mechanism in the case of growth.	Regulatory authorities may consider that further data from amphibians are not required. However, see right-hand column..	The LAGDA does not cover the reproductive phase of the life cycle, but a life cycle test which could be used to address any concerns about reproduction is not currently available, although a fish life cycle test (MEOGRT) could provide useful information.
F	+	–	Eq/0	1) Strong evidence for *in vivo* endocrine activity with adverse effects (on growth or sexual development) in amphibians, but possibly not via an E,A,T,S mechanism. 2) Strong evidence for *in vivo* endocrine activity in amphibians, but possibly not via an E,A,T,S mechanism. 3) Strong evidence for adverse effects on growth or sexual development in amphibians, but probably not via an endocrine mechanism in the case of growth.	Regulatory authorities may consider that further data from amphibians are not required. However, see right-hand column.	The LAGDA does not cover the reproductive phase of the life cycle, but a life cycle test which could be used to address any concerns about reproduction is not currently available, although a fish life cycle test (MEOGRT) could provide useful information.. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
G	+	Eq/0	+	1) Strong evidence for *in vivo* endocrine activity with adverse effects (on growth or sexual development) in amphibians and other species, but possibly not via an E,A,T,S mechanism. 2) Strong evidence for *in vivo* endocrine activity in amphibians and other species, but possibly not via an E,A,T,S mechanism. 3) Strong evidence for adverse effects on growth or sexual development in amphibians and other species, but probably not via an endocrine mechanism in the case of growth.	It would be desirable to obtain some unequivocal mechanistic data to confirm whether or not an E,A,T,S mechanism is operating. Regulatory authorities may consider that further data from amphibians are not required. However, see right-hand column.	The LAGDA does not cover the reproductive phase of the life cycle, but a life cycle test which could be used to address any concerns about reproduction is not currently available, although a fish life cycle test (MEOGRT) could provide useful information. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

Scenarios	Result of LAGDA	Existing results		Possible conclusions: 1) Indicators of endocrine activity and apical endpoints positive 2) Indicators of endocrine activity positive 3) Apical endpoint positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
H	+	Eq/0	–	1) Strong evidence for *in vivo* endocrine activity with adverse effects (on growth or sexual development) in amphibians, but possibly not via an E,A,T,S mechanism. 2) Strong evidence for *in vivo* endocrine activity in amphibians, but possibly not via an E,A,T,S mechanism. 3) Strong evidence for adverse effects on growth or sexual development in amphibians, but possibly not via an endocrine mechanism in the case of growth.	It would be desirable to obtain some unequivocal mechanistic data to confirm whether or not an E,A,T,S mechanism is operating. Regulatory authorities may consider that further data from amphibians are not required. However, see right-hand column.	The LAGDA does not cover the reproductive phase of the life cycle, but a life cycle test which could be used to address any concerns about reproduction is not currently available, although a fish life cycle test (MEOGRT) could provide useful information. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
I	+	Eq/0	Eq/0	1) Strong evidence for *in vivo* endocrine activity with adverse effects (on growth or sexual development) in amphibians, but possibly not via an E,A,T,S mechanism. 2) Strong evidence for *in vivo* endocrine activity in amphibians, but possibly not via an E,A,T,S mechanism. 3) Strong evidence for adverse effects on growth or sexual development in amphibians, but possibly not via an endocrine mechanism in the case of growth.	It would be desirable to obtain some unequivocal mechanistic data to confirm whether or not an E,A,T,S mechanism is operating. Regulatory authorities may consider that further data from amphibians are not required. However, see right-hand column.	The LAGDA does not cover the reproductive phase of the life cycle, but a life cycle test which could be used to address any concerns about reproduction is not currently available, although a fish life cycle test (MEOGRT) could provide useful information. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
J	–	+	+	The test chemical has E,A,T,S activity in other species but not apparently in amphibians, although it is possible that *Xenopus laevis* has responded atypically in this case (e.g. if *X. laevis* responded positively in OECD TG 231).	Regulatory authorities may consider that further data from amphibians are not required. However, see right-hand column.	The LAGDA does not cover the reproductive phase of the life cycle, but a life cycle test which could be used to address any concerns about reproduction is not currently available, although a fish life cycle test (MEOGRT) could provide useful information.
K	–	+	–	The test chemical has E,A,T,S activity *in vitro*, but no apparent activity *in vivo* in amphibians or other species, possibly due to quick degradation/metabolism or failure to reach the active site.	Regulatory authorities may consider that further testing is unnecessary.	–

Scenarios	Result of LAGDA	Existing results		Possible conclusions: 1) Indicators of endocrine activity and apical endpoints positive 2) Indicators of endocrine activity positive 3) Apical endpoint positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
L	–	+	Eq/0	The test chemical has E,A,T,S activity *in vitro*, but no apparent activity *in vivo* in amphibians, possibly due to quick degradation/metabolism or failure to reach the active site.	Regulatory authorities may consider that further testing is unnecessary, but see right-hand column.	Given the presence of E,A,T,S activity *in vitro*, and the absence of reliable *in vivo* data from other species, it might be desirable to run an *in vivo* endocrine screen with fish or mammals. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
M	–	–	+	The test chemical does not apparently have E,A,T,S activity in amphibians, but endocrine activity is present in other species.	Regulatory authorities may consider that further data from amphibians are not required. However, see right-hand column.	The LAGDA does not cover the reproductive phase of the life cycle, but a life cycle test which could be used to address any concerns about reproduction is not currently available.
N	–	–	–	The test chemical does not have E,A,T,S activity in amphibians or other species.	No further action is necessary.	–
O	–	–	Eq/0	The test chemical does not have E,A,T,S activity in amphibians.	No further action is necessary.	It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
P	–	Eq/0	+	The test chemical probably does not have E,A,T,S activity in amphibians, but the uncertain mechanistic data and the presence of endocrine activity in other species reduces confidence in this conclusion. It is possible that *Xenopus laevis* has responded atypically in this case (e.g. if *X. laevis* responded positively in OECD TG 231).	Regulatory authorities may consider that further data from amphibians are not required. However, see right-hand column. Also, if clear *in vitro* mechanistic data are missing, it might be desirable to obtain some.	The LAGDA does not cover the reproductive phase of the life cycle, but a life cycle test which could be used to address any concerns about reproduction is not currently available. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

Scenarios	Result of LAGDA	Existing results		Possible conclusions: 1) Indicators of endocrine activity and apical endpoints positive 2) Indicators of endocrine activity positive 3) Apical endpoint positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
Q	–	Eq/0	–	The test chemical is probably without endocrine activity in amphibians or other taxa, but this conclusion is tentative given the lack of supporting mechanistic data.	If clear *in vitro* mechanistic data are missing, it might be desirable to obtain some.	It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
R	–	Eq/0	Eq/0	The test chemical is probably without endocrine activity in amphibians, but this conclusion is tentative given the lack of supporting data.	Some regulatory authorities may conclude that no further evidence is required, but see right-hand column.	If clear *in vitro* mechanistic data are missing, it may be desirable to obtain some. If these data reveal E,A,T,S activity, it might then be desirable to conduct a fish or rodent screen. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

C.2.11. Avian Reproduction Test (OECD TG 206)

Status: Assay validated by the OECD.

439. Modality detected/endpoints: OECD TG 206 does not contain endpoints which solely respond to endocrine disrupters (EDs), and it has not been specifically validated with EDs. However, some of the endpoints in this apical test are nevertheless potentially affected by estrogen/androgen/thyroid/steroidogenesis (E,A,T,S) EDs. Of particular interest in the context of estrogens, androgens and steroidogenesis disrupters are egg production, embryo viability and hatchability, but other endpoints may also be responsive to some EDs (e.g. growth may respond to some thyroid disrupters; percentage of cracked eggs and egg shell thickness may respond to chemicals interfering with the control of shell deposition).

Background to the assay

440. This assay is designed primarily as an apical test for chemicals with suspected reproductive toxicity, but it is not a life cycle test as it only runs from the stage of pre-laying adults to 14-day-old offspring. Furthermore, only the adults are exposed to the test chemical (via the food), and any effects on sexual development would not be detectable. The endpoints are all apical measures of development, growth or reproduction. Key endpoints which might be affected by EDs include egg production, viability and hatchability. Possible test organisms include mallard duck (*Anas platyrhynchos*), bobwhite quail (*Colinus virginiatus*) and Japanese quail (*Coturnix japonica*).

441. Depending on the species and test objectives, endpoints could include *inter alia* sex ratio (phenotypic and/or genotypic), sex hormones, thyroid hormones, reproductive/thyroid organ weights, gonad histopathology and gross pathology, time to first egg laying, and sexual behaviour. These types of endpoint are all included in the Avian Two-Generation Test (ATGT). However, note that the ATGT does not cover all relevant behaviours and is performed in a precocial species which reacts very differently to embryonic exposure to a test material compared with an altricial species. Given the high degree of endocrine system conservation across the vertebrates, adverse endocrine-linked effects in the Avian Reproduction Test may indicate the possibility of related activity in other organisms such as fish, amphibians, reptiles or mammals.

When/why the assay may be used

442. Although OECD TG 206 could, in principle, be used at any stage in the hazard assessment process, the most likely use scenario will be when there are already some data available to suggest possible endocrine disruption properties. In other words, OECD TG 206 will generally be used to investigate whether such properties result in adverse apical effects on development, growth or reproduction over the reproductive part of the avian life cycle. It would be unlikely to be used if other bird reproduction data are already available. OECD TG 206 could not be used as a primary screen for EDs. Another potential limitation of OECD TG 206 is that the effects of test chemicals may not become fully

apparent during the test because the offspring are not directly dosed, and only receive bioaccumulated material which may be passed from their mothers via the egg.

443. In order to provide information relevant for assessing whether or not a chemical may fulfil the WHO/IPCS (2002) definition of an ED, the study design has to be sufficiently robust to demonstrate the presence or absence of effects. In the dose selection, the investigator should also consider and ensure that data generated are adequate to fulfil the regulatory requirement across OECD countries as appropriate (e.g. hazard and risk assessment and labelling, ED assessment, etc.). The top dose or concentration should be sufficiently high to give clear systemic (i.e. non endocrine-specific) toxicity in order to ensure that a wide range of exposures (high to low) is tested. However, endocrine effects observed solely in the presence of clear systemic toxicity should be interpreted with caution and may be disregarded when sufficiently justified to be caused by secondary effects which are unlikely to be due to endocrine activity. The reason for this advice is a concern that some endocrine active substance (EAS) sensitive assays are being run at doses/concentrations of EASs that are too low to trigger direct impacts on the endocrine system. This guidance document is not the place to address this issue directly, but it should be considered when EAS-sensitive test guidelines (TGs) are revised in the future. In addition, the number and spacing of dose/concentration levels should also be adequate to fulfil the objectives of the study (e.g. to demonstrate dose response relationships if this is required).

Existing data to be considered

444. Existing data available before deployment of OECD TG 206 for ED hazard assessment are likely to include information on possible modes of action (MOA) from quantitative structure activity relationships (QSARs) and/or *in vitro* screens. It would not be advisable to conduct an unmodified OECD TG 206 without mechanistic screening data because it would then not be possible to link any apical effects with endocrine disruption. Given the commonality of endocrine mechanisms in the vertebrates, relevant existing data available before deployment of OECD TG 206 (Avian Reproduction Test) might also include *in vivo* results obtained with other vertebrates (e.g. a positive Uterotrophic Bioassay with rodents; positive findings for endocrine endpoints in mammalian repeat dose toxicity or reproductive studies; or positive result in the fish assays OECD TG 229 or TG 230). As the ethical and financial cost of OECD TG 206 is high, it is important to make full use of existing endocrine-related data, both before the test is begun and during data evaluation.

Scenarios: Positive and negative results combined with existing data

445. The scenarios (A to R) presented in Table C.2.11 represent all the possibilities of positive or negative results in combination with the presence or absence of existing data. The action taken will also depend on the regulatory environment, but the considerations given here are generally science based. Wherever possible, the recommended "next step which could be taken" avoids unnecessary animal testing. However, sometimes conducting an animal test will be indicated and then the relevance of species, strain and exposure route should always be considered. Further considerations specific to each scenario are given in the table.

446. Positive results obtained with one of the OECD TG 206 endpoints which are outside the range of historical controls may result in the conclusion that the test chemical is able to cause adverse effects *in vivo* (Table C.2.11, Scenarios A-I), but not necessarily that it is an

ED. Note that if doubt exists about the test performance (e.g. highly unusual results in controls), a comparison with historical control data with respect to overall test performance might be helpful. However, the nature of these effects and any existing data will require careful consideration. If *in vitro* and/or *in vivo* data already exist which reveal possible endocrine disrupting properties (Scenarios A, B and D), a positive endpoint in OECD TG 206 could lead to a tentative conclusion that the test chemical is an actual ED.

447. If a plausible link of a responding OECD TG 206 endpoint with previously identified endocrine activity can be made, regulatory authorities may conclude that sufficient evidence is available to categorise the chemical as an ED (i.e. interference with the endocrine system has caused adverse effects *in vivo*), and no further information might then be required. However, if a more robust link between adverse effects and an endocrine modality is required (bearing in mind that none of the existing data are likely to have been generated in avian systems), or if possible effects during the sexual development part of the life cycle are suspected, or if the chemical is suspected to cause epigenetic effects, it would be desirable to run an ATGT. Furthermore, if data on hazard are required for an environmental hazard identification/characterisation, an ATGT may also be needed unless the precision of the data from OECD TG 206 (which only uses three test concentrations) are considered adequate for such an assessment. On the other hand, if data from prior endocrine screens and tests are negative (Scenario E), a positive response in OECD TG 206 would not support the hypothesis that the chemical is an ED in birds. It could, of course, still be subjected to an environmental hazard identification/characterisation, but only if sufficient concentrations have been tested to allow derivation of an adequately precise lowest-observed-effect-concentration no-observed-effect-concentration (LOEC/NOEC).

448. The scenarios in which OECD TG 206 gives a negative result (Table C.2.11, Scenarios J-R) lead to a tentative conclusion that the test chemical is not an ED in birds, and this conclusion is strengthened considerably if prior screens have failed to reveal endocrine activity (Scenario N). In the latter circumstances, regulatory authorities may be justified in concluding that no further action is needed. However, if it is thought possible that the sexual development part of the life cycle is sensitive, then conducting an ATGT should be considered. Also, if one or more of those screens was positive (Scenarios J-M and P), the bioconcentration factor of the chemical should be checked. If the bioconcentration factor indicates that the chemical is strongly bioaccumulative, it would also be worth considering conducting an ATGT. If a chemical which screened positive is not bioaccumulative, the probable reasons for lack of effects in OECD TG 206 might be metabolism to an inactive chemical, or failure to reach the active site, and no further action would be indicated.

449. In each of the above scenarios, it is possible that existing data will be equivocal (Table C.2.117, Scenarios C, F-I, L and O-R), or there may be no existing data. This will weaken the conclusions which can be drawn about a positive OECD TG 206, and this is reflected in Table C.2.11. However, as indicated above, it would be undesirable to proceed with OECD TG 206 if prior data on endocrine activity are equivocal or absent. On the other hand, if OECD TG 206 is positive, it would be essential to obtain some reliable mechanistic data before reaching a conclusion about whether or not the chemical is an ED in birds. There is also the possibility that equivocal mechanistic data may be the result of multiple modes of endocrine action. Under some circumstances, two opposite modes of simultaneous action (e.g. estrogenic and anti-estrogenic) could, depending on dose, lead to a minimisation or abolition of adverse effects, while in others two different MOA (e.g. estrogenic and anti-androgenic) could potentially reinforce effects on certain apical endpoints. If multiple MOA are suspected, either from the existing results or based on

QSAR/read-across/integrated approaches, this situation should be investigated further if needed for regulatory decision making.

450. The scenario in which the results of OECD TG 206 are themselves equivocal has not been dealt with in Table C.2.11, for reasons of brevity. In this context, an equivocal result might be an inconsistent concentration-response (e.g. no effect at a high concentration but effects at a lower concentration), or a result which borders on statistical significance. Without knowing the exact circumstances, reliable advice cannot be given, but the opinions of an experienced ecotoxicologist should be sought. However, if prior screens are negative, it is doubtful if further action is needed, because the chemical is unlikely to be an ED. If an endocrine screen is positive, some types of equivocal OECD TG 206 results would have to be taken more seriously. For example, an inconsistent concentration-response would not necessarily rule out the test chemical as an ED in birds. An example of this would be a chemical which causes adverse effects on reproduction at low doses, but reduced reproductive success and ultimately mortality at very high doses, thus potentially giving a U-shaped response curve. Ideally, concentrations causing systemic toxicity of this type should not be tested in OECD TG 206, but such toxicity may have been missed in earlier screens.

451. In summary, positive results in OECD TG 206 indicate that a chemical may be an ED if they can be plausibly linked to an endocrine MOA established on the basis of prior screening. However, more conclusive data in this regard would be obtainable from an ATGT. If screening data are unavailable or negative, it should not be concluded that a positive OECD TG 206 is the result of endocrine disruption. On the other hand, a negative OECD TG 206 combined with negative screening data should lead to a conclusion that a chemical is probably not an ED in birds. A negative OECD TG 206 set against a background of a positive screen might, however, raise concerns if the chemical is strongly bioaccumulative, known to be involved in epigenesis, or suspected of having effects on sexual development, when an ATGT should be considered.

Reference

WHO/IPCS (2002), "Global assessment of the state-of-the-science of endocrine disrupters", Damstra, T. et al. (eds.) WHO/PCS/EDC/02.2, World Health Organization, Geneva, www.who.int/ipcs/publications/new_issues/endocrine_disruptors/en.

Table C.2.11. **Avian Reproduction Test (OECD TG 206): Guidance for scenarios of combinations of results with existing data**

This table represents possible conclusions to be drawn from assay data, and a next step which could be taken if further evidence is required about possible endocrine disrupting properties and/or effects. The guidance offered is not meant to be prescriptive, but provides science-based considerations. It encourages the use of all available data and expert judgement in a weight of evidence approach. Regional and national interpretation of results and "next steps" may vary.

The conclusions are grouped into a series of scenarios (A-R), each scenario representing a different combination of assay results, existing *in vitro* data and existing *in vivo* data. The symbol "+" indicates that the data in question represent a positive result, «-" indicates a negative result, and "Eq/0" indicates that the data are either equivocal or are not available.

Existing results: * "Mechanism (*in vitro* mechanistic data)" assumes that mechanistic data are available from estrogen receptor (ER-), androgen receptor (AR-) and steroidogenesis-based assays (Level 2). Thyroid hormone receptor (TR) and other assays concerning mechanisms of thyroid disruption may be available, but they are not in common use. In practice, data from all assays may not be available and therefore this must be taken into account when deciding on the "next step". Quantitative structure activity relationship (QSAR) predictions of estrogen and androgen binding/activation may be made for some substances. There is no evidence at present that equivalent *in vitro* assays with systems derived from birds offer advantages over their mammalian counterparts.

Existing results: ** "Effects (*in vivo* effects of concern)" assumes effects have been observed in other *in vivo* screens/tests which give rise to concern that the test chemical may be an endocrine disrupter.

Note that although this assay has been used for many years to assess the sub-acute effects of chemicals, and no formal attempt has been made to validate it for use with potential endocrine disruptors (EDs), the United States Environmental Protection Agency (US EPA) has shown that reproduction is a part of the avian life cycle which can be responsive to EDs (https://www.regulations.gov/document?D=EPA-HQ-OPPT-2014-0766-0019). Furthermore, the US EPA has published the Avian Two-Generation Test (ATGT) protocol which contains several ED-specific endpoints, although it has not been internationally validated or harmonised with OECD guidelines.

Scenario	Result of OECD TG 206	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
A	+	+	+	The test chemical is probably an endocrine disruptor (ED) if the modality identified in existing screens/tests can be plausibly linked to the affected endpoint.	Further evidence is probably not required.	If the affected endpoint in OECD TG 206 cannot be plausibly linked to the known modality, the test chemical is unlikely to be an ED in birds. OECD TG 206 cannot detect effects on sexual development and is unlikely to detect effects from long-term bioaccumulation. If these are suspected, an Avian Two-Generation Reproduction Test (ATGT) may reveal them.
B	+	+	–	The test chemical is probably an ED in birds if the modality identified in existing screens/tests can be plausibly linked to the affected endpoint.	Further evidence is probably not required.	If the affected endpoint in OECD TG 206 cannot be plausibly linked to the known modality, the test chemical is unlikely to be an ED in birds. OECD TG 206 cannot detect effects on sexual development and is unlikely to detect effects from long-term bioaccumulation. If these are suspected, an ATGT may reveal them.
C	+	+	Eq/0**	The test chemical is probably an ED in birds if the modality identified in existing screens/tests can be plausibly linked to the affected endpoint.	Further evidence is probably not required.	If the affected endpoint in OECD TG 206 cannot be plausibly linked to the known modality, the test chemical is unlikely to be an ED in birds. OECD TG 206 cannot detect effects on sexual development and is unlikely to detect effects from long-term bioaccumulation. If these are suspected, an ATGT may reveal them. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple modes of action (MOA). If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
D	+	–	+	The test chemical may be an ED, but the negative mechanistic data reduce the confidence in this conclusion. However, if the endocrine disruption effects in existing *in vivo* tests can be plausibly linked to the OECD TG 206 responses, this increases the probability that the chemical is an ED in birds.	Further evidence is probably not required.	If the affected endpoint in OECD TG 206 cannot be plausibly linked to the known modality, the test chemical is unlikely to be an ED in birds. OECD TG 206 cannot detect effects on sexual development and is unlikely to detect effects from long-term bioaccumulation. If these are suspected, an ATGT may reveal them.
E	+	–	–	The test chemical is unlikely to be an ED.	Further evidence is probably not required.	It is possible that the effects observed in OECD TG 206 have been caused by an unknown endocrine mechanism. This would not, however, prevent the chemical being subjected to hazard identification/characterisation. OECD TG 206 cannot detect effects on sexual development and is unlikely to detect effects from long-term bioaccumulation. If these are suspected, an ATGT may reveal them.

Scenario	Result of OECD TG 206	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
F	+	–	Eq/0	The test chemical is unlikely to be an ED, but the relevance of any equivocal existing *in vivo* data to the OECD TG 206 results should be examined.	Further evidence is probably not required.	It is possible that the effects observed in OECD TG 206 have been caused by an unknown endocrine mechanism – equivocal existing *in vivo* data may throw some light on this. The absence of data on a possible endocrine mechanism would, however, not prevent the chemical being subjected to hazard identification/characterisation. OECD TG 206 cannot detect effects on sexual development and is unlikely to detect effects from long-term bioaccumulation. If these are suspected, an ATGT may reveal them. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
G	+	Eq/0	+	The test chemical may be an ED, but the equivocal or absent mechanistic data reduce the confidence in this conclusion. However, if the endocrine disruption effects in existing *in vivo* tests can be plausibly linked to the OECD TG 206 responses, this increases the probability that the chemical is an ED.	If reliable mechanistic data are not available, it would be desirable to obtain some.	The test chemical is probably an ED in birds if a modality identified in the newly commissioned mechanistic screens (see left-hand column), or in the existing *in vivo* data, can be plausibly linked to the affected endpoint. OECD TG 206 cannot detect effects on sexual development and is unlikely to detect effects from long-term bioaccumulation. If these are suspected, an ATGT may reveal them. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
H	+	Eq/0	–	The test chemical may be an ED, but the equivocal or absent mechanistic data reduce the confidence in this conclusion.	If reliable mechanistic data are not available, it would be desirable to obtain some.	The test chemical is probably an ED in birds if a modality identified in the newly commissioned mechanistic screens (see left-hand column) can be plausibly linked to the affected endpoint. OECD TG 206 cannot detect effects on sexual development and is unlikely to detect effects from long-term bioaccumulation. If these are suspected, an ATGT may reveal them. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

Scenario	Result of OECD TG 206	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
I	+	Eq/0	Eq/0	The test chemical may be an ED, but the equivocal or absent mechanistic and *in vivo* data reduce the confidence in this conclusion. Final conclusions about whether a chemical is a potential ED cannot be drawn from the results of this test alone.	If reliable mechanistic data are not available, it would be desirable to obtain some.	The test chemical is probably an ED in birds if a modality identified in the newly commissioned mechanistic screens (see left-hand column) can be plausibly linked to the affected endpoint. OECD TG 206 cannot detect effects on sexual development and is unlikely to detect effects from long-term bioaccumulation. If these are suspected, an ATGT may reveal them. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
J	–	+	+	The chemical is probably not an ED in birds **that acts through the mechanisms tested in the available *in vitro* and *in vivo* studies**.	If the chemical is strongly bioaccumulative, is suspected to affect sexual development or cause epigenetic effects, consider conducting an ATGT.	If any effects in an ATGT can be plausibly linked with mechanistic data, the test chemical is probably an ED in birds.
K	–	+	–	The chemical is probably not an ED **in birds that acts through the mechanisms tested in the available *in vitro* and *in vivo* studies.**	If the chemical is strongly bioaccumulative, is suspected to affect sexual development or cause epigenetic effects, consider conducting an ATGT.	If any effects in an ATGT can be plausibly linked with mechanistic data, the test chemical is probably an ED in birds.
L	–	+	Eq/0	The chemical is probably not an ED in birds **that acts through the mechanisms tested in the available *in vitro* and *in vivo* studies**.	If the chemical is strongly bioaccumulative, is suspected to affect sexual development or cause epigenetic effects, consider conducting an ATGT.	If any effects in an ATGT can be plausibly linked with mechanistic data, the test chemical is probably an ED in birds. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
M	–	–	+	The chemical is probably not an ED in birds **that acts through the mechanisms tested in the available *in vitro* and *in vivo* studies**.	If the chemical is strongly bioaccumulative, is suspected to affect sexual development or cause epigenetic effects, consider conducting an ATGT.	If any effects in an ATGT can be plausibly linked with *in vivo* data which provide information on endocrine disruption properties, the test chemical is probably an ED in birds, but likely not by a mechanism covered by the existing *in vitro* screens.
N	–	–	–	The chemical is probably not an ED in birds **that acts through the mechanisms tested in the available *in vitro* and *in vivo* studies**.	Further evidence is probably not required.	OECD TG 206 cannot detect effects on sexual development and is unlikely to detect effects from long-term bioaccumulation. If these are suspected, an ATGT may reveal them.

Scenario	Result of OECD TG 206	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
O	–	–	Eq/0	The chemical is probably not an ED in birds **that acts through the mechanisms tested in the available *in vitro* and *in vivo* studies**.	Further evidence is probably not required.	OECD TG 206 cannot detect effects on sexual development and is unlikely to detect effects from long-term bioaccumulation. If these are suspected, an ATGT may reveal them. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
P	–	Eq/0	+	The chemical is probably not an ED in birds **that acts through the mechanisms tested in the available *in vitro* and *in vivo* studies**.	If reliable mechanistic data are not available, it would be desirable to obtain some.	If the newly commissioned mechanistic data are positive and the chemical is strongly bioaccumulative, or if developmental effects are suspected, consider conducting an ATGT. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
Q	–	Eq/0	–	The chemical is probably not an ED in birds, but confidence in this conclusion is reduced by the lack of clear mechanistic data.	Further evidence is probably not required, but confidence in the conclusion would be increased by the provision of reliable negative mechanistic data.	If any newly commissioned mechanistic data are positive and the chemical is strongly bioaccumulative, or if developmental effects are suspected, consider conducting an ATGT. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
R	–	Eq/0	Eq/0	The chemical may not be an ED in birds, but confidence in this conclusion is reduced by the lack of clear mechanistic and existing *in vivo* data.	Further evidence is probably not required, but confidence in the conclusion would be increased by the provision of reliable negative mechanistic data.	If any newly commissioned mechanistic data are positive and the chemical is strongly bioaccumulative, or if developmental effects are suspected, consider conducting an ATGT. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

C.2.12. Sediment-Water Chironomid Life-Cycle Toxicity Test (OECD TG 233)

Status: Assay validated by the OECD.

452. Modality detected/endpoints: This long-term *in vivo* assay with the dipteran insect *Chironomus* spp. is responsive to juvenile hormone (JH) (ant)agonists and ecdysteroid (Ec) (ant)agonists which can interfere with such processes as metamorphosis, moulting, growth and reproduction. It exposes the test organisms over two generations. It is important to note, however, that none of the endpoints in this apical test are specifically responsive to JH- or Ec-active chemicals, and the assay will give positive results with many other substances. The lack of internationally validated mechanistic assays for endocrine activity in insects may prevent firm conclusions about whether test chemicals are endocrine dusruptors (EDs) in this taxon, although *in vitro* assays for JH and Ec activity are available in the literature. However, the data from the test may nevertheless be of value for classification and hazard identification/characterisation.

Background to the assay

453. This life cycle assay can be run with one of several chironomid species, including *Chironomus riparius*, *C. dilutus* and *C. yoshimatsui*. It can also be operated in one of two formats, with the test chemical spiked either into the ambient water or into the sediment, thus allowing sparingly soluble or hydrophobic chemicals to be tested. The test with *C. riparius* and *C. yoshimatsui* takes 44 days, while with *C. dilutus* it continues for *ca.* 100 days. The exposure to a range of test concentrations begins with first instar larvae (F0) and continues to fully emerged adulthood of the second generation (F1), so two cycles of reproduction are evaluated.

454. The assay is relatively new (approved by the OECD in 2010), but the available data to date (e.g. Hahn, Liess and Schulz [2001]; Taenzler et al. [2007]; Jungmann et al. [2009]; Tassou and Schulz [2009, 2013]) show that chemicals acting as ant(agonists) of both the JH and Ec hormonal pathways can produce effects on most of the available endpoints.

When/why the assay may be used

455. Although OECD TG 233 could, in principle, be used at any stage in the hazard assessment process, the most likely use scenario will be when there are already some *in vitro* and *in vivo* data available about the possible JH or Ec activity and/or effects of a chemical. Given the significant degree of endocrine system conservation across the arthropods, effects in OECD TG 233 may also indicate the possibility of related activity in other arthropods such as crustaceans (cladocera, copepods and decapods).

456. It is not recommended that OECD TG 233 is deployed as a primary test for JH or Ec activity and effects, but it should be noted that there are no standardised *in vitro* screens for JH or Ec (ant)agonists, although some are described in the scientific literature (e.g. Cherbas, Koehler and Cherbas, 1989).

Existing data to be considered

457. Existing information on endocrine-related effects from other arthropods should also be considered before deployment of OECD TG 233, given the commonality of endocrine mechanisms in these taxa. Existing data available might also include one or more of a range of *in silico* or *in vitro* results which suggest that JH or Ec disruption may occur *in vivo* (but note the limitations of this approach, as indicated above). Such indicators of possible JH or Ec activity might include quantitative structure activity relationship (QSAR) predictions of JH/Ec activity, "read-across" from *in vivo* results obtained with structurally related chemicals or positive results from an *in vitro* screen for JH/Ec (ant)agonist activity. In addition, *in vivo* data should ideally be available from one or more of several assays, possibly including the Short-Term Juvenile Hormone Activity Screening Assay (SJHASA), the Sediment-Water Chironomid Toxicity Test Using Spiked Sediment or Water (OECD TG 218/219), or the *Daphnia magna* Reproduction Test with male neonate option (OECD TG 211). If positive data are available from the *Daphnia* Multigeneration Test (DMGT), these should also be taken into account.

458. In order to provide information relevant for assessing whether or not a chemical may fulfil the WHO/IPCS (2002) definition of an ED, the study design has to be sufficiently robust to demonstrate the presence or absence of effects. In the dose selection, the investigator should also consider and ensure that data generated are adequate to fulfil the regulatory requirement across OECD countries as appropriate (e.g. hazard and risk assessment and labelling, ED assessment, etc.). The top dose or concentration should be sufficiently high to give clear systemic (i.e. non endocrine-specific) toxicity in order to ensure that a wide range of exposures (high to low) is tested. However, endocrine effects observed solely in the presence of clear systemic toxicity should be interpreted with caution and may be disregarded when sufficiently justified to be caused by secondary effects which are unlikely to be due to endocrine activity. The reason for this advice is a concern that some endocrine active substance (EAS) sensitive assays are being run at doses/concentrations of EASs that are too low to trigger direct impacts on the endocrine system. This guidance document is not the place to address this issue directly, but it should be considered when EAS-sensitive test guidelines (TGs) are revised in the future. In addition, the number and spacing of dose/concentration levels should also be adequate to fulfil the objectives of the study (e.g. to demonstrate dose response relationships if this is required).

Scenarios: Positive and negative results combined with existing data

459. The scenarios (A to R) presented in Table C.2.12 represent all the possibilities of positive or negative results in combination with the presence or absence of existing data. The action taken will also depend on the regulatory environment, but the considerations given here are generally science based. Wherever possible, the recommended "next step which could be taken" avoids unnecessary animal testing. However, sometimes conducting an animal test will be indicated and then the relevance of species, strain and exposure route should always be considered. Further considerations specific to each scenario are given in the table.

460. Positive results obtained with OECD TG 233 (Table C.2.12, Scenarios A-I) result in the conclusion that the test chemical has adverse apical effects, at least in insects, but these are not necessarily caused by JH or Ec activity. However, although a positive response of the OECD TG 233 indicates that the chemical has adverse effects in insects, it

should be noted that crustacean species such as *Daphnia* have a parthenogenetic reproductive strategy and so may respond differently to *Chironomus*. Therefore, if countries need further evidence concerning growth and sexual development, etc. in this phylum, a Harpacticoid Copepod Development and Reproduction Test (OECD GD 201) and/or the DMGT would be able to provide information on adverse effects in other arthropod species. In other words, in order to strengthen weight of evidence, a positive result in OECD TG 233 could be followed by GD 201 (Level 4) and/or the DMGT (Level 5). Existing data suggesting endocrine-specific activity (e.g. positive *in vitro* data, or positive *in vivo* data from other species) will strengthen the case for additional testing still further.

461. The situation in which OECD TG 233 gives a negative result (Table C.2.12, Scenarios J-R) needs careful consideration of any existing data. If these data suggest that the chemical is endocrine active both *in vitro* and *in vivo* (Scenario J), then it is possible that OECD TG 233 is simply insufficiently sensitive.

462. If OECD TG 233 and existing *in vivo* data are all negative, but *in vitro* data reveal some JH or Ec activity (Scenario K), the probability is that the test chemical is not sufficiently potent to produce JH/Ec (ant)agonism *in vivo* in arthropods, or it may be rapidly metabolised. In such a situation, further testing is probably not necessary.

463. On the other hand, if OECD TG 233 and the *in vitro* tests are negative (Scenario M), but there are positive existing *in vivo* data, the nature of those existing data should be considered. Unless the existing data are from another insect, the chemical is possibly not a JH or Ec (ant)agonist acting in insects, but it may be more potent in species (e.g. crustaceans) or life stages that have not been tested. In this situation, the existing *in vivo* data should be used to guide decisions about whether to conduct any further testing (e.g. with OECD GD 201).

464. Finally, a negative OECD TG 233, set against a background of negative *in vitro* and *in vivo* data (Scenario N), suggests that the test chemical is probably not a JH or Ec (ant)agonist *in vitro* or *in vivo*, and further action is unnecessary.

465. In each of the above scenarios, it is possible that existing data will be equivocal, or there may be no existing data. This will weaken the conclusions which can be drawn about a negative OECD TG 233, and this is reflected in Table C.2.12. However, a lack of mechanistic data on JH or Ec activity should ideally be addressed before any further *in vivo* testing is finally conducted, although as indicated above, *in vitro* JH/Ec screens have not yet been internationally standardised. On the other hand, if OECD TG 233 is positive, further *in vivo* testing would generally be needed to quantify any adverse effects in crustaceans, even if all existing data are equivocal, or if there are no existing data. Again, however, it may be useful to obtain some mechanistic information before conducting further *in vivo* testing. There is also the possibility that equivocal mechanistic data may be the result of multiple modes of endocrine action. Under some circumstances, two opposite modes of simultaneous action (e.g. JH or Ec agonistic and antagonistic) could, depending on dose, lead to a minimisation or abolition of adverse effects, while in others two different modes of action (MOA) could potentially reinforce effects on the OECD TG 233 endpoint. If multiple MOA are suspected, either from the existing results or based on QSAR/read-across/integrated approaches, this situation should be investigated further if needed for regulatory decision making.

466. The scenario in which the results of OECD TG 233 are themselves equivocal has not been dealt with in Table C.2.12, for reasons of brevity. In this context, an equivocal result might be an inconsistent concentration-response (e.g. no effect at a high concentration but effects at a lower concentration), or a result which borders on statistical significance.

Without knowing the exact circumstances, reliable advice cannot be given, but the opinions of an experienced ecotoxicologist should be sought. Clearly, however, such equivocal results do not necessarily rule out the existence of *in vivo* endocrine activity. If possible reasons for false negatives are suspected, OECD TG 233 could be repeated (e.g. conduct it at lower concentrations which avoid systemic toxicity).

467. In summary, positive results in OECD TG 233 indicate that a chemical has adverse effects in insects which may or may not be via JH or Ec (ant)agonism. This may need to be followed up with further apical testing with crustaceans. Negative results in OECD TG 233 do not necessarily mean that the chemical is not a potential ED – a judgement about the endocrine disruption potential in other arthropods will have to be made based on a weight of evidence evaluation of existing *in vitro* and *in vivo* data.

References

Cherbas, L., M.M.D. Koehler and P. Cherbas (1989), "Effects of juvenile hormone on the ecdysone response of *Drosophila* Kc cells", *Developmental Genetics*, Vol. 10/3, pp. 177-188, https://doi.org/10.1002/dvg.1020100307.

Hahn, T., M. Liess and R. Schulz (2001), "Effects of the hormone mimetic insecticide tebufenozide in *Chironomus riparius* larvae in two different exposure setups", *Ecotoxicology and Environmental Safety*, Vol. 49/2, pp. 171-178, https://doi.org/10.1006/eesa.2001.2055.

Jungmann, D. et al. (2009), "Chronic toxicity of fenoxycarb to the midge *Chironomus riparius* after exposure in sediments of different composition", *Journal of Soils and Sediments*, Vol. 9/2, pp. 94-102, http://dx.doi.org/10.1007/s11368-009-0056-2.

Taenzler, V. et al. (2007), "Chironomids: Suitable test organisms for risk assessment investigations on the potential endocrine disrupting properties of pesticides", *Ecotoxicology*, Vol. 16/1, pp. 221-230, https://doi.org/10.1007/s10646-006-0117-x.

Tassou, K.T. and R. Schulz (2013), "Low field-relevant tebufenozide concentrations affect reproduction in *Chironomus riparius* (Diptera: Chironomidae) in a long-term toxicity test", *Environmental Science and Pollution Research*, Vol. 20/6, pp. 3735-3742, http://dx.doi.org/10.1007/s11356-012-1311-4.

Tassou, K.T. and R. Schulz (2009), "Effects of the insect growth regulator pyriproxyfen in a two-generation test with *Chironomus riparius*", *Ecotoxicology and Environmental Safety*, Vol. 72, pp. 1058-1062, https://doi.org/10.1016/j.ecoenv.2009.02.001.

WHO/IPCS (2002), "Global assessment of the state-of-the-science of endocrine disrupters", Damstra, T. et al. (eds.) WHO/PCS/EDC/02.2, World Health Organization, Geneva, www.who.int/ipcs/publications/new_issues/endocrine_disruptors/en.

Table C.2.12. **Sediment-Water Chironomid Life-Cycle Toxicity Test (OECD TG 233): Guidance for scenarios of combinations of results with existing data**

This table represents possible conclusions to be drawn from assay data, and a next step which could be taken if further evidence is required about possible endocrine disrupting properties and/or effects. The guidance offered is not meant to be prescriptive, but provides science-based considerations. It encourages the use of all available data and expert judgement in a weight of evidence approach. Regional and national interpretation of results and "next steps" may vary.

The conclusions are grouped into a series of scenarios (A-R), each scenario representing a different combination of assay results, existing *in vitro* data and existing *in vivo* data. The symbol "+" indicates that the data in question represent a positive result, «-" indicates a negative result, and "Eq/0" indicates that the data are either equivocal or are not available.

Existing results: * "Mechanism (*in vitro* mechanistic data)" assumes that mechanistic data are available from available from juvenile hormone- (JH) or ecdysteroid (Ec-) based assays. JH or Ec assays concerning mechanisms of disruption may be available, but they are have not yet been internationally standardised. In practice, data from all assays may not be available and therefore this must be taken into account when deciding on the "next step".

Existing results: ** "Effects (*in vivo* effects of concern)" assumes effects have been observed in other *in vivo* screens/tests which give rise to concern that the test chemical may be a JH or Ec disrupter.

Scenarios	Result of OECD TG 233	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
A	+	+	+	Strong evidence for adverse *in vivo* effects in insects, possibly but not necessarily caused by juvenile hormone (JH) or ecdysteroid (Ec) (ant)agonists, plus possible JH or Ec effects in other arthropods.	It would be desirable (if not already conducted) to perform an apical test with crustaceans (e.g. the Harpacticoid Copepod Development and Reproduction Test – OECD GD 201; and/or the *Daphnia* Multigeneration Test [DMGT]).	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is a JH or Ec (ant)agonist. It should be noted that chironomids reproduce sexually and therefore may respond differently to a parthenogenetic species such as *Daphnia*.
B	+	+	–	Strong evidence for adverse *in vivo* effects in insects, possibly but not necessarily caused by JH or Ec (ant)agonists.	It would be desirable (if not already conducted) to perform an apical test with crustaceans (e.g. the Harpacticoid Copepod Development and Reproduction Test – OECD GD 201; and/or the DMGT).	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is a JH or Ec (ant)agonist. It should be noted that chironomids reproduce sexually and therefore may respond differently to a parthenogenetic species such as *Daphnia*.
C	+	+	Eq/0	Strong evidence for adverse *in vivo* effects in insects, possibly but not necessarily caused by JH or Ec (ant)agonists.	It would be desirable (if not already conducted) to perform an apical test with crustaceans (e.g. the Harpacticoid Copepod Development and Reproduction Test – OECD GD 201; and/or the DMGT).	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is a JH or Ec (ant)agonist. It should be noted that chironomids reproduce sexually and therefore may respond differently to a parthenogenetic species such as *Daphnia*. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple modes of action (MOA). If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
D	+	–	+	Strong evidence for adverse *in vivo* effects in insects, possibly but not necessarily caused by JH or Ec (ant)agonists, plus possible JH or Ec effects in other arthropods.	It would be desirable (if not already conducted) to perform an apical test with crustaceans (e.g. the Harpacticoid Copepod Development and Reproduction Test – OECD GD 201; and/or the DMGT).	The lack of *in vitro* JH or Ec activity is not necessarily evidence against any JH/Ec activity, due to the limited nature of current *in vitro* JH and Ec screens. It should be noted that chironomids reproduce sexually and therefore may respond differently to a parthenogenetic species such as *Daphnia*.
E	+	–	–	Some evidence for adverse *in vivo* effects in insects, possibly but not necessarily caused by JH or Ec (ant)agonists.	It would be desirable (if not already conducted) to perform an apical test with crustaceans (e.g. the Harpacticoid Copepod Development and Reproduction Test – OECD GD 201; and/or the DMGT).	The lack of *in vitro* JH or Ec activity is not evidence against any JH/Ec activity, due to the limited nature of current *in vitro* JH and Ec screens. It should be noted that chironomids reproduce sexually and therefore may respond differently to a parthenogenetic species such as *Daphnia*.

Scenarios	Result of OECD TG 233	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
F	+	–	Eq/0	Some evidence for adverse *in vivo* effects in insects, possibly but not necessarily caused by JH or Ec (ant)agonists.	Given the absence or equivocal nature of existing *in vivo* data, it might also be sensible to conduct a JH/Ec-responsive crustacean assay (e.g. the Harpacticoid Copepod Development and Reproduction Test – OECD GD 201) or a JH-responsive DMGT.	The lack of *in vitro* JH/Ec activity is not evidence against any JH/Ec activity, due to the limited nature of current *in vitro* JH/Ec screens. It should be noted that chironomids reproduce sexually and therefore may respond differently to a parthenogenetic species such as *Daphnia*. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
G	+	Eq/0	+	Strong evidence for adverse *in vivo* effects in insects, possibly but not necessarily caused by JH or Ec (ant)agonists, plus possible JH or Ec effects in other arthropods.	Given the absence or equivocal nature of existing *in vitro* data, it would be desirable to obtain further *in vitro* data on JH/Ec activity if possible. It might also be sensible to conduct a JH/Ec-responsive crustacean assay (e.g. the Harpacticoid Copepod Development and Reproduction Test – OECD GD 201) or a JH-responsive DMGT.	If a new *in vitro* mechanistic assay is conducted, note that a negative does not mean that the test material has no JH/Ec activity. It should be noted that chironomids reproduce sexually and therefore may respond differently to a parthenogenetic species such as *Daphnia*. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
H	+	Eq/0	–	Some evidence for adverse *in vivo* effects in insects, possibly but not necessarily caused by JH or Ec (ant)agonists.	Given the absence or equivocal nature of the *in vitro* mechanistic data, it might also be helpful to conduct an *in vitro* screen for JH/Ec activity. It might also be sensible to conduct a JH/Ec-responsive crustacean assay (e.g. the Harpacticoid Copepod Development and Reproduction Test – OECD GD 201) or a JH-responsive DMGT.	If a new *in vitro* mechanistic assay is conducted, note that a negative does not mean that the test material has no JH/Ec activity. It should be noted that chironomids reproduce sexually and therefore may respond differently to a parthenogenetic species such as *Daphnia*. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

Scenarios	Result of OECD TG 233	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
I	+	Eq/0	Eq/0	Some evidence for adverse *in vivo* effects in insects, possibly but not necessarily caused by JH or Ec (ant)agonists.	Given the absence or equivocal nature of the *in vitro* mechanistic data, it might also be helpful to conduct an *in vitro* screen for JH/Ec activity. It might also be sensible to conduct a JH/Ec-responsive crustacean assay (e.g. the Harpacticoid Copepod Development and Reproduction Test – OECD GD 201) or a JH-responsive DMGT.	If a new *in vitro* mechanistic assay is conducted, note that a negative does not mean that the test material has no JH/Ec activity. It should be noted that chironomids reproduce sexually and therefore may respond differently to a parthenogenetic species such as *Daphnia*. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
J	–	+	+	The test chemical is probably a JH or Ec (ant)agonist without adverse effects in insects, although it is possible that *Chironomus* spp. respond atypically in this case.	Some regulatory authorities may conclude that no further evidence is required. However, it might be desirable to obtain data from crustaceans (e.g. the Harpacticoid Copepod Development and Reproduction Test – OECD GD 201; or the DMGT) if these are not already available.	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is a JH/Ec (ant)agonist. It should be noted that chironomids reproduce sexually and therefore may respond differently to a parthenogenetic species such as *Daphnia*.
K	–	+	–	The test chemical is probably a JH or Ec (ant)agonist without adverse effects in insects or other taxa, although it is possible that *Chironomus* responds atypically in this case.	If there is no activity in crustaceans or insects, further evidence is probably not needed.	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is a JH/Ec (ant)agonist.
L	–	+	Eq/0	The test chemical is probably a JH or Ec (ant)agonist without adverse effects in insects, although it is possible that *Chironomus* responds atypically in this case.	Some regulatory authorities may conclude that no further evidence is required, but if crustacean data are absent, it might be desirable to conduct a Harpacticoid Copepod Development and Reproduction Test – OECD GD 201; or a DMGT.	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is a JH/Ec agonist. It should be noted that chironomids reproduce sexually and therefore may respond differently to a parthenogenetic species such as *Daphnia*. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
M	–	–	+	The test chemical is probably without JH or Ec activity in insects, although it is possible that *Chironomus* responds atypically in this case.	Some regulatory authorities may conclude that no further evidence is required. However, it might be desirable to obtain data from crustaceans (e.g. Harpacticoid Copepod Development and Reproduction Test – OECD GD 201; or a DMGT) if these are not already available.	The lack of *in vitro* JH/Ec activity is not evidence against any JH activity, due to the limited nature of current *in vitro* JH screens. It should be noted that chironomids reproduce sexually and therefore may respond differently to a parthenogenetic species such as *Daphnia*.

Scenarios	Result of OECD TG 233	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
N	–	–	–	The test chemical is probably without JH or Ec activity in arthropods.	No further action is necessary.	–
O	–	–	Eq/0	The test chemical is probably without JH or Ec activity in arthropods.	Some regulatory authorities may conclude that no further evidence is required. However, it might be desirable to obtain data from crustaceans (e.g. the Harpacticoid Copepod Development and Reproduction Test – OECD GD 201; or a DMGT) if these are not already available.	The lack of *in vitro* JH/Ec activity is not evidence against any JH activity, due to the limited nature of current *in vitro* JH screens. It should be noted that chironomids reproduce sexually and therefore may respond differently to a parthenogenetic species such as *Daphnia*. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
P	–	Eq/0	+	The test chemical is probably without JH or Ec activity in insects, although it is possible that *Chironomus* responds atypically in this case.	Some regulatory authorities may conclude that no further evidence is required. Also, if clear *in vitro* mechanistic data are missing, it might be desirable to obtain some.	If a new *in vitro* mechanistic assay is conducted, note that a negative does not mean that the test material has no JH/Ec activity. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
Q	–	Eq/0	–	The test chemical is probably without JH or Ec activity in arthropods.	No further action is necessary.	It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
R	–	Eq/0	Eq/0	The test chemical is probably without JH or Ec activity in insects and possibly crustaceans.	Some regulatory authorities may conclude that no further evidence is required. However, it might be desirable to obtain data from crustaceans (e.g. the Harpacticoid Copepod Development and Reproduction Test – OECD GD 201; or a DMGT) if these are not already available.	It should be noted that chironomids reproduce sexually and therefore may respond differently to a parthenogenetic species such as *Daphnia*. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

C.2.13. Medaka Extended One-Generation Reproduction Test (MEOGRT) (OECD TG 240)

Status: Assay validated by the OECD.

468. Modality detected/endpoints: This fish life cycle test was specifically designed to investigate the apical effects of endocrine disrupters, and has several endpoints which can be considered diagnostic of some types of estrogen/androgen/thyroid/steroidogenesis (E,A,T,S) activity. This gives it an advantage over other currently standardised life cycle tests, and its use for evaluating endocrine disruptors (EDs) is to be preferred to the Fish Life Cycle Toxicity Test (see Section C.2.21) which, although sensitive to the apical effects of some EDs, contains no endocrine-sensitive endpoints. In view of the inclusion of certain ED-specific endpoints, the MEOGRT can contribute useful evidence about the probable causality of apical effects, which is a key issue in the definition of EDs.

Background to the assay

469. This assay is a comprehensive test using medaka (*Oryzias latipes*) exposed continuously from the adult stage of the first generation (F0) to the newly hatched stage of the third generation (F2). In other words, it includes two phases of reproductive activity, and two phases of embryonic development and hatching, separated by a full phase of growth and sexual development. It begins with pairs of sexually mature F0 fish (at least 12 weeks post-fertilisation, or wpf) reproducing for 3 weeks, brings their F1 offspring to sexual maturity (15 weeks), then allows the F1 adults to breed, and finally follows their offspring (F2) to hatching (up to 18 days post-fertilisation, or dpf). The main emphasis of the assay concerns population-relevant apical endpoints (e.g. survival, development, growth and reproduction). However, in order to obtain mechanistic information, additional endpoints include measurements of vitellogenin (either as protein – VTG, or as mRNA coding for vitellogenin – *vtg*), secondary sex characteristics, phenotypic sex compared with genetic sex, and gonadal histopathology. Histopathology of liver and kidney may also be measured in order to distinguish between endocrine effects and possible systemic or other toxicity. While the assay is able to distinguish large deviations from the expected 50:50 sex ratio of F1 offspring, it has less power than the Fish Sexual Development Test (FSDT) to distinguish small deviations due to the relatively small number of fish per replicate (12).

470. It should be noted that the MEOGRT is a relatively new test (adopted by the OECD in 2015) which has not yet been widely used (Watanabe et al., 2017). Furthermore, due to the test's cost and complexity, the validation process involved fewer laboratories than for many simpler assays. A recent publication (Flynn et al., 2017) which evaluated nine validation studies of the MEOGRT found that only one complied with all the biological validity criteria, so caution should be used when assessing MEOGRT data. There is a significant risk of test failure because of its length and difficulty. Nevertheless, development of the assay has built on experience with shorter assays involving medaka (e.g. the Fish Short-Term Reproduction Assay [FSTRA] and the FSDT), and earlier versions of it have been used for research purposes. It is possible that for some applications (e.g. when testing

highly bioaccumulative chemicals for trans-generational effects or if epigenetic effects are suspected) it might be feasible to extend the MEOGRT to the reproduction phase of the F2 generation, but at present there is insufficient information to warrant this. Currently, however, few testing laboratories have experience with the MEOGRT, and an extended version has not been standardised or validated.

471. Only medaka is recommended for use in this test design. A related assay using zebrafish (*Danio rerio*), the Zebrafish Extended One-Generation Test (ZEOGRT), is currently being validated by the OECD (see Section C.2.22), but is not expected to be adopted for several years.

When/why the assay may be used

472. Although the MEOGRT could, in principle, be used at any stage in the hazard assessment process, the most likely use scenario will be when there are already some data available to suggest possible endocrine disruption properties. In other words, the MEOGRT will generally be used to investigate whether such potential properties result in adverse apical effects on development, growth or reproduction over an entire life cycle. It is unlikely (and undesirable) that the MEOGRT will be the first ED-responsive test procedure to be applied to a chemical. Furthermore, the conduct of a ZEOGRT in addition to a MEOGRT is not likely to be necessary (for example, to address perceived sensitivity differences). Before either assay is initiated, careful thought should be given to which is more appropriate in the circumstances. For example, if previous data are available with zebrafish and the ZEOGRT is sufficiently powerful for the expected endpoint of concern, then conducting a ZEOGRT may be the correct choice. However, if a genetic sex marker or secondary sexual characters are desired, it may be more beneficial to consider a MEOGRT.

473. This is a comprehensive test which examines a range of potentially adverse apical effects, but also considers several ED-specific endpoints. It is therefore suitable for helping to define whether a test chemical is an ED, and the results could be used in an environmental hazard identification/characterisation for fish. Given the high degree of endocrine system conservation across the vertebrates, adverse endocrine-linked effects in the MEOGRT may also indicate the possibility of related activity in other organisms such as amphibians, reptiles, birds or mammals.

474. In order to provide information relevant for assessing whether or not a chemical may fulfil the WHO/IPCS (2002) definition of an ED, the study design has to be sufficiently robust to demonstrate the presence or absence of effects. In the dose selection, the investigator should also consider and ensure that data generated are adequate to fulfil the regulatory requirement across OECD countries as appropriate (e.g. hazard and risk assessment and labelling, ED assessment, etc.). The top dose or concentration should be sufficiently high to give clear systemic (i.e. non endocrine-specific) toxicity in order to ensure that a wide range of exposures (high to low) is tested. However, endocrine effects observed solely in the presence of clear systemic toxicity should be interpreted with caution and may be disregarded when sufficiently justified to be caused by secondary effects which are unlikely to be due to endocrine activity. The reason for this advice is a concern that some endocrine active substance (EAS) sensitive assays are being run at doses/concentrations of EASs that are too low to trigger direct impacts on the endocrine system. This guidance document is not the place to address this issue directly, but it should be considered when EAS-sensitive test guidelines (TGs) are revised in the future. In addition, the number and spacing of dose/concentration levels should also be adequate to

fulfil the objectives of the study (e.g. to demonstrate dose response relationships if this is required).

Existing data to be considered

475. Existing data available before deployment of the MEOGRT for endocrine disruption hazard assessment are likely to include information on possible modes of action (MOA) from quantitative structure activity relationships (QSARs), adverse outcome pathways (AOP) and/or *in vitro* screens. These may be accompanied by *in vivo* fish assay data from EASZY, the Juvenile Medaka Anti-Androgen Screening Assay, OECD TG 229 and/or OECD TG 230, and may also include data from TG 234 (FSDT). In addition, existing information on endocrine-related effects from other vertebrates (up to and including mammals, e.g. positive findings for endocrine endpoints in mammalian repeat dose toxicity or reproductive studies) should also be considered, given the commonality of endocrine mechanisms in these taxa. It would not be advisable or ethically desirable to conduct a MEOGRT without mechanistic or *in vivo* screening data because it would then be less straightforward to link any apical effects with endocrine disruption. Furthermore, data from OECD TG 229 and/or TG 234 (FSDT), especially if obtained with medaka, could be of use in focusing attention in the MEOGRT on particularly vulnerable parts of the life cycle. Given the high ethical and financial cost of the MEOGRT, it is important to make full use of existing endocrine-related data, both before the test is begun and during data evaluation.

Scenarios: Positive and negative results combined with existing data

476. The scenarios (A to R) presented in Table C.2.13 represent all the possibilities of positive or negative results in combination with the presence or absence of existing data. The action taken will also depend on the regulatory environment, but the considerations given here are generally science-based. Wherever possible, the recommended "next step which could be taken" avoids unnecessary animal testing. However, sometimes conducting an animal test will be indicated and then the relevance of species, strain and exposure route should always be considered. Further considerations specific to each scenario are given in the table.

477. Positive results obtained with one of the MEOGRT apical endpoints result in the conclusion that the test chemical is able to cause adverse effects *in vivo* (Table C.2.13, Scenarios A-I), but not necessarily that it is an ED. Note that if doubt exists about the test performance (e.g. highly unusual results in controls), a comparison with historical control data with respect to overall test performance might be helpful. However, the nature of these effects and any existing data will require careful consideration. If *in vitro* and/or *in vivo* data already exist which reveal possible endocrine disrupting properties (Scenarios A, B and D), a positive apical endpoint in the MEOGRT could lead to a conclusion that the test chemical is an actual ED if adverse population effects are expected as a consequence. This conclusion will, of course, be reinforced if mechanistic endpoints in the MEOGRT itself also respond. The probability that the test chemical is an ED will also be strengthened considerably if the endocrine modality identified in the present or earlier tests is plausibly linked to the responding endpoint. For example, if the chemical has estrogenic properties (such as the induction of vitellogenin in males) and observations indicate reduced fecundity of the F0 or F1 adults in the MEOGRT, this gives added confidence in this conclusion. On the other hand, it may be harder to argue a plausible link between estrogenic properties on the one hand, and an endpoint such as growth or survival on the other, although it is known that some estrogens are able to cause changes in growth rates (Knacker et al., 2010). In this

example, an effect solely on growth or survival, while potentially of concern from the viewpoint of environmental hazard identification/characterisation, would not on its own lead to a conclusion that the chemical is an ED in fish.

478. If a plausible link of a responding MEOGRT apical endpoint with identified endocrine activity can be made, regulatory authorities may conclude that sufficient evidence is available to categorise the chemical as an ED (i.e. interference with the endocrine system has caused adverse effects *in vivo*), and no further information might then be required. It may also be necessary to consider whether or not effects observed are relevant at the population level (e.g. reproduction, growth, development). On the other hand, if data from prior endocrine screens and tests are negative, including negative mechanistic data from the MEOGRT itself (Scenario E), a positive apical response in the MEOGRT would not in general support the hypothesis that the chemical is an ED in fish (although a change in sex ratio may have been caused by an ED). The chemical could, of course, still be subjected to an environmental hazard identification/characterisation.

479. The scenarios in which the MEOGRT gives a negative apical result (Table C.2.13, Scenarios J-R) lead to a tentative conclusion that the test chemical is not an ED in fish, and this conclusion is strengthened considerably if prior screens, or the MEOGRT itself, have failed to reveal endocrine activity (Scenario N). In the latter circumstances, regulatory authorities would be justified in concluding that no further action is needed. On the other hand, if one or more of those screens was positive (Scenarios J-M and P), the bioconcentration factor (BCF) of the chemical should be checked. If the BCF indicates that the chemical is strongly bioaccumulative and reaches equilibrium slowly, it would be worth considering the conduct of an extended MEOGRT (but no TG is available for this), although as indicated above, there is little evidence at present that EDs with a high BCF would be consistently more potent in such a test. If a chemical which screened positive is not bioaccumulative, the probable reasons for lack of effects in the MEOGRT might be metabolism to an inactive chemical, or failure to reach the active site, and no further action would be indicated.

480. In each of the above scenarios, it is possible that existing data will be equivocal (Table C.2.13, Scenarios C, F-I, L and O-R), or there may be no existing data. This will weaken the conclusions which can be drawn about a positive apical endpoint in the MEOGRT, and this is reflected in Table C.2.13. However, as indicated above, it would be undesirable to proceed with a MEOGRT if prior data on endocrine activity are equivocal or absent, and if there are no other effect- or exposure-related reasons for considering such a comprehensive test. On the other hand, if the MEOGRT shows a positive apical endpoint, it would be essential to obtain some reliable mechanistic data before reaching a conclusion about whether or not the chemical is an ED in fish. There is also the possibility that equivocal mechanistic data may be the result of multiple modes of endocrine action. Under some circumstances, two opposite modes of simultaneous action (e.g. estrogenic and anti-estrogenic) could, depending on dose, lead to a minimisation or abolition of adverse effects, while in others two different MOA (e.g. estrogenic and anti-androgenic) could potentially reinforce effects on certain apical endpoints. If multiple MOA are suspected, either from the existing results or based on QSAR/read-across/integrated approaches, this situation should be investigated further if needed for regulatory decision making.

481. The scenario in which the results of the MEOGRT are themselves equivocal has not been dealt with in Table C.2.13, for reasons of brevity. In this context, an equivocal result might be an inconsistent concentration-response (e.g. no effect at a high concentration but effects at a lower concentration), or a result which borders on statistical significance.

Without knowing the exact circumstances, reliable advice cannot be given, but the opinions of an experienced ecotoxicologist should be sought. However, if a comprehensive set of prior screens are all negative, it is doubtful whether further action is needed, because the chemical is unlikely to be an ED. If an endocrine screen is positive, some types of equivocal MEOGRT apical results would have to be taken more seriously. For example, a non-monotonic concentration-response would not necessarily rule out the test chemical as an ED in fish. An example of this would be a chemical like ethinylestradiol, which causes adverse effects (increased fecundity) on fish reproduction at low doses, but reduced reproductive success at very high doses, thus potentially giving a U-shaped response curve (e.g. Jobling et al., 2004). Ideally, concentrations causing systemic toxicity of this type should not be tested in MEOGRT, but such toxicity may have been missed in earlier screens.

482. In summary, positive apical results in the MEOGRT indicate that a chemical is a probable ED if they can be plausibly linked to an endocrine MOA established on the basis of prior mechanistic screening or concurrent observation of mechanistic effects or their biochemical/physiological manifestations. If such screening data are unavailable or negative, it should not be concluded that a positive MEOGRT is the result of endocrine disruption (although it is likely that biased sex ratio will be the result of ED). On the other hand, a negative MEOGRT combined with a sufficiently comprehensive set of negative screening data could lead to a firm conclusion that a chemical is not an ED in fish. A negative MEOGRT set against a background of a positive screen might, however, raise concerns (e.g. if the chemical is strongly bioaccumulative or known to be involved in epigenesist). In this case an extended MEOGRT could be considered, although this is not covered by OECD TG 240, and its effectiveness in this regard is unproven.

References

Flynn, K. et al. (2017), "Summary of the development the US Environmental Protection Agency's Medaka Extended One-Generation Reproduction Test (MEOGRT) using data from 9 multigenerational medaka tests", *Environmental Toxicology and Chemistry*, Vol. 36/12, pp. 3387-3403, https://doi.org/10.1002/etc.3923.

Jobling, S. et al. (2004), "Comparative responses of molluscs and fish to environmental estrogens and an estrogenic effluent", *Aquatic Toxicology*, Vol. 66/2, pp. 207-222, https://doi.org/10.1016/j.aquatox.2004.01.002.

Knacker, T. et al. (2010), "Environmental effect assessment for sexual endocrine-disrupting chemicals: Fish testing strategy", *Integrated Environmental Assessment and Management*, Vol. 6/4, pp. 653-662, https://doi.org/10.1002/ieam.92.

Watanabe, H. et al. (2017), "Medaka Extended One-Generation Reproduction Test (MEOGRT) evaluating 4-nonylphenol", *Environmental Toxicology and Chemistry*, Vol. 36/12, pp. 3254-3266, https://doi.org/10.1002/etc.3895.

WHO/IPCS (2002), "Global assessment of the state-of-the-science of endocrine disrupters", Damstra, T. et al. (eds.) WHO/PCS/EDC/02.2, World Health Organization, Geneva, www.who.int/ipcs/publications/new_issues/endocrine_disruptors/en.

Table C.2.13. **Medaka Extended One-Generation Reproduction Test (MEOGRT) (OECD TG 240): Guidance for scenarios of combinations of results with existing data**

This table represents possible conclusions to be drawn from assay data, and a next step which could be taken if further evidence is required about possible endocrine disrupting properties and/or effects. The guidance offered is not meant to be prescriptive, but provides science-based considerations. It encourages the use of all available data and expert judgement in a weight of evidence approach. Regional and national interpretation of results and "next steps" may vary.

The conclusions are grouped into a series of scenarios (A-R), each scenario representing a different combination of assay results, existing mechanistic data and existing *in vivo* effects data. The symbol "+" indicates that the data in question represent a positive result, «-" indicates a negative result, and "Eq/0" indicates that the data are either equivocal or are not available.

Results of the MEOGRT: * Apical results of the MEOGRT include effects on survival, growth, development, sex ratio and reproduction. The other MEOGRT endpoints, including vitellogenin, secondary sex characteristics, sex ratio (again) and gonadal histopathology, can be indicative of endocrine mechanisms which may have caused the apical effect.

Existing results: ** "Mechanism (*in vitro* and/or *in vivo* mechanistic data)" assumes that mechanistic data are available from estrogen receptor (ER), androgen receptor (AR) and steroidogenesis-based assays (Level 2). Thyroid hormone receptor (TR) and other assays concerning mechanisms of thyroid disruption may also be available. In practice, data from all assays may not be available and therefore this must be taken into account when deciding on the "next step". Quantitative structure activity relationship (QSAR) predictions of estrogen and androgen binding/activation may be made for some substances. There is no evidence at present that equivalent *in vitro* assays with systems derived from fish offer advantages over their mammalian counterparts.

Existing results: *** "Effects (*in vivo* effects of concern)" assumes effects have been observed in other *in vivo* screens/tests which give rise to concern that the test chemical may be an endocrine disrupter.

Scenario	Apical result of MEOGRT*	Existing results		Possible conclusions: 1. Indicators of endocrine activity and apical endpoints positive 2. Indicators of endocrine activity positive and apical endpoints negative 3. Indicators of endocrine activity negative and apical endpoint positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* and/or *in vivo* mechanistic data)**	Effects (*in vivo* effects of concern)***			
A	+	+	+	1) Strong evidence for adverse effects in fish and other organisms by an endocrine mechanism. 2) Strong evidence for endocrine effects, but they do not appear adverse in fish. 3) Strong evidence for adverse effects in fish and other organisms. There is a possibility that the apical endpoint sex ratio is more sensitive to the test chemical than the mechanistic endpoint vitellogenin (VTG), or mechanism may hypothetically not be via direct interaction with estrogen receptor (ER), androgen receptor (AR) or by aromatase inhibition, even though it is noted that currently there is no evidence for sex ratio change in fish caused by other mechanisms than those mentioned here at otherwise non-toxic concentrations of chemicals.	Further evidence is probably not required.	If the affected apical endpoint in the MEOGRT cannot be plausibly linked to the known modality, the test chemical is unlikely to be an endocrine disruptor (ED). The MEOGRT is unlikely to detect epigenetic effects. If these are suspected, extending the test beyond F2 hatching could be considered, although this would depart from OECD TG 240.
B	+	+	–	1) Strong evidence for adverse effects in fish by an endocrine mechanism. 2) Strong evidence for endocrine effects in fish, but they do not appear adverse. 3) Strong evidence for adverse effects in fish and other organisms. There is a possibility that the apical endpoint sex ratio is more sensitive to the test chemical than the mechanistic endpoint VTG, or mechanism may hypothetically not be via direct interaction with ER, AR or by aromatase inhibition, even though it is noted that currently there is no evidence for sex ratio change in fish caused by other mechanisms than those mentioned here at otherwise non-toxic concentrations of chemicals.	Further evidence is probably not required.	If the affected apical endpoint in the MEOGRT cannot be plausibly linked to the known modality, the test chemical is unlikely to be an ED. The MEOGRT is unlikely to detect epigenetic effects. If these are suspected, extending the test beyond F2 hatching could be considered, although this would depart from TG 240.
C	+	+	Eq/0**	1) Strong evidence for adverse effects in fish by an endocrine mechanism. 2) Strong evidence for endocrine effects in fish, but they do not appear adverse. 3) Strong evidence for adverse effects in fish and other organisms. There is a possibility that the apical endpoint sex ratio is more sensitive to the test chemical than the mechanistic endpoint VTG, or mechanism may hypothetically not be via direct interaction with ER, AR or by aromatase inhibition, even though it is noted that currently there is no evidence for sex ratio change in fish caused by other mechanisms than those mentioned here at otherwise non-toxic concentrations of chemicals.	Further evidence is probably not required.	If the affected apical endpoint in the MEOGRT cannot be plausibly linked to the known modality, the test chemical is unlikely to be an ED. The MEOGRT is unlikely to detect epigenetic effects. If these are suspected, extending the test beyond F2 hatching could be considered, although this would depart from OECD TG 240. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple modes of action (MOA). If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

Scenario	Apical result of MEOGRT*	Existing results		Possible conclusions: 1. Indicators of endocrine activity and apical endpoints positive 2. Indicators of endocrine activity positive and apical endpoints negative 3. Indicators of endocrine activity negative and apical endpoint positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* and/or *in vivo* mechanistic data)**	Effects (*in vivo* effects of concern)***			
D	+	–	+	1) Strong evidence for adverse effects in fish and other organisms, possibly by an unknown endocrine mechanism. 2) Medium-strong evidence for endocrine effects in fish, but they do not appear to be adverse. 3) Strong evidence for adverse effects in more than one organism, but mechanism may not be by endocrine disruption.	Further evidence is probably not required.	If the affected apical endpoint in the MEOGRT cannot be plausibly linked to the known modality, the test chemical is unlikely to be an ED. The MEOGRT is unlikely to detect epigenetic effects. If these are suspected, extending the test beyond F2 hatching could be considered, although this would depart from OECD TG 240.
E	+	–	–	1) Strong evidence for adverse effects in fish, possibly by an unknown endocrine mechanism. 2) Medium-strong evidence for endocrine effects in fish, but they do not appear to be adverse. 3) Strong evidence for adverse effects in fish, but mechanism may not be by endocrine disruption.[1]	Further evidence is probably not required.	It is possible that the effects observed in the MEOGRT have been caused by an unknown endocrine mechanism. This would not, however, prevent the chemical being subjected to hazard identification/characterisation. The MEOGRT is unlikely to detect epigenetic effects. If these are suspected, extending the test beyond F2 hatching could be considered, although this would depart from OECD TG 240.
F	+	–	Eq/0	1) Strong evidence for adverse effects in fish, possibly by an unknown endocrine mechanism. 2) Medium-strong evidence for endocrine effects in fish, but they do not appear to be adverse. 3) Strong evidence for adverse effects in fish, but mechanism may not be by endocrine disruption.	Further evidence is probably not required.	It is possible that the effects observed in the MEOGRT have been caused by an unknown endocrine mechanism – equivocal existing *in vivo* data may throw some light on this. The absence of data on a possible endocrine mechanism would, however, not prevent the chemical being subjected to hazard identification/characterisation. The MEOGRT is unlikely to detect epigenetic effects. If these are suspected, extending the test beyond F2 hatching could be considered, although this would depart from OECD TG 240. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

Note: 1. However, note that if biased sex ratio is observed, it is likely to have been caused by an endocrine disrupting chemical.

Scenario	Apical result of MEOGRT*	Existing results		Possible conclusions: 1. Indicators of endocrine activity and apical endpoints positive 2. Indicators of endocrine activity positive and apical endpoints negative 3. Indicators of endocrine activity negative and apical endpoint positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* and/or *in vivo* mechanistic data)**	Effects (*in vivo* effects of concern)***			
G	+	Eq/0	+	1) Strong evidence for adverse effects in more than one organism, possibly by an unknown endocrine mechanism. 2) Medium-strong evidence for endocrine effects, but they do not appear to be adverse in fish. 3) Strong evidence for adverse effects in more than one organism, but mechanism may not be by endocrine disruption.	If reliable mechanistic data are not available, it would be desirable to obtain some.	The test chemical is probably an ED if a modality identified in the newly commissioned mechanistic screens, or in the existing *in vivo* data, can be plausibly linked to the affected endpoint. The MEOGRT is unlikely to detect epigenetic effects. If these are suspected, extending the test beyond F2 hatching could be considered, although this would depart from OECD TG 240. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
H	+	Eq/0	–	1) Strong evidence for adverse effects in fish, possibly by an unknown endocrine mechanism. 2) Medium-strong evidence for endocrine effects in fish, but they do not appear to be adverse. 3) Strong evidence for adverse effects in fish, but mechanism may not be by endocrine disruption.	If reliable mechanistic data are not available, it would be desirable to obtain some.	The test chemical is probably an ED if a modality identified in the newly commissioned mechanistic screens can be plausibly linked to the affected endpoint. The MEOGRT is unlikely to detect epigenetic effects. If these are suspected, extending the test beyond F2 hatching could be considered, although this would depart from OECD TG 240. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
I	+	Eq/0	Eq/0	1) Strong evidence for adverse effects in fish, possibly by an unknown endocrine mechanism. 2) Moderate-strong evidence for endocrine effects in fish, but they do not appear to be adverse. 3) Strong evidence for adverse effects in fish, but mechanism may not be by endocrine disruption.	If reliable mechanistic data are not available, it would be desirable to obtain some.	The test chemical is probably an ED if a modality identified in the newly commissioned mechanistic screens can be plausibly linked to the affected endpoint. The MEOGRT is unlikely to detect epigenetic effects. If these are suspected, extending the test beyond F2 hatching could be considered, although this would depart from OECD TG 240. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

Scenario	Apical result of MEOGRT*	Existing results		Possible conclusions: 1. Indicators of endocrine activity and apical endpoints positive 2. Indicators of endocrine activity positive and apical endpoints negative 3. Indicators of endocrine activity negative and apical endpoint positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* and/or *in vivo* mechanistic data)**	Effects (*in vivo* effects of concern)***			
J	–	+	+	The chemical is probably not an ED in fish, unless this conclusion is contradicted by existing *in vivo* data.	If the chemical is strongly bioaccumulative, or if epigenetic effects are suspected, an extended version of the MEOGRT could be considered, although this would depart from OECD TG 240.	If any effects in an extended MEOGRT can be plausibly linked with mechanistic data, the test chemical is probably an ED.
K	–	+	–	The chemical is probably not an ED in fish.	If the chemical is strongly bioaccumulative, or if epigenetic effects are suspected, an extended version of the MEOGRT could be considered, although this would depart from OECD TG 240.	If any effects in an extended MEOGRT can be plausibly linked with mechanistic data, the test chemical is probably an ED.
L	–	+	Eq/0	The chemical is probably not an ED in fish.	If the chemical is strongly bioaccumulative, or if epigenetic effects are suspected, an extended version of the MEOGRT could be considered, although this would depart from OECD TG 240.	If any effects in an extended MEOGRT can be plausibly linked with mechanistic data, the test chemical is probably an ED. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
M	–	–	+	The chemical is probably not an ED in fish.	If the chemical is strongly bioaccumulative, or if epigenetic effects are suspected, an extended version of the MEOGRT could be considered, although this would depart from OECD TG 240.	If any effects in an extended MEOGRT can be plausibly linked with *in vivo* data which provide information on ED properties, the test chemical is probably an ED, but likely not by a mechanism covered by the existing *in vitro* screens.
N	–	–	–	The chemical is probably not an ED.	Further evidence is probably not required.	–
O	–	–	Eq/0	The chemical is probably not an ED in fish.	Further evidence is probably not required.	It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to any mechanistic information.

Scenario	Apical result of MEOGRT*	Existing results		Possible conclusions: 1. Indicators of endocrine activity and apical endpoints positive 2. Indicators of endocrine activity positive and apical endpoints negative 3. Indicators of endocrine activity negative and apical endpoint positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* and/or *in vivo* mechanistic data)**	Effects (*in vivo* effects of concern)***			
P	–	Eq/0	+	The chemical is probably not an ED in fish.	If reliable mechanistic data are not available, it would be desirable to obtain some.	If the newly commissioned mechanistic data are positive and the chemical is strongly bioaccumulative, or if epigenetic effects are suspected, consider conducting an extended MEOGRT, although this would depart from OECD TG 240. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
Q	–	Eq/0	–	The chemical is probably not an ED, but confidence in this conclusion is reduced by the lack of clear mechanistic data.	Further evidence is probably not required, but confidence in the conclusion would be increased by the provision of reliable negative mechanistic data.	If the newly commissioned mechanistic data are positive and the chemical is strongly bioaccumulative, or if epigenetic effects are suspected, consider conducting an extended MEOGRT, although this would depart from OECD TG 240. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
R	–	Eq/0	Eq/0	The chemical may not be an ED, but confidence in this conclusion is reduced by the lack of clear mechanistic and existing *in vivo* data.	Further evidence is probably not required, but confidence in the conclusion would be increased by the provision of reliable negative mechanistic data.	If the newly commissioned mechanistic data are positive and the chemical is strongly bioaccumulative, or if epigenetic effects are suspected, consider conducting an extended MEOGRT, although this would depart from OECD TG 240. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

Non-OECD non-mammalian screens and tests (Conceptual Framework Levels 3-5)

C.2.14. Short-Term Juvenile Hormone Activity Screening Assay using *Daphnia magna* (SJHASA) (draft OECD TG)

Status: Assay being validated by the OECD.

483. Modality detected/endpoints: This short-term *in vivo* assay with *Daphnia magna* is expected to be responsive to juvenile hormone (JH) agonists which lead to the production of male offspring.

Background to the assay

484. This *in vivo* assay is in undergoing validation by the OECD, and may be approved as a test guideline (TG) in due course. The SJHASA exposes 17-day-old (i.e. adult) female *D. magna* to dilutions of the test chemical for 5-7 days. Their first brood after exposure is discarded, but all individuals of the second brood are sexed by observation of their longer first antenna. Juvenile hormone (JH) and other JH agonists cause the production of males due to exposure during a short critical period (52-53 hours after ovulation). An adverse outcome pathway for this process is under development – significant male production in a population could potentially lead to its decline. However, due to the very short-term nature of SJHASA, the endpoint of male production should not be considered as an adverse apical endpoint without further investigation in longer term tests.

485. OECD TG 211 (the *Daphnia magna* Reproduction Test) already has an option to measure male production as a response to JH agonists, but it is a much more resource-intensive test than the SJHASA and takes three times as long to perform.

When/why the assay may be used

486. Although the SJHASA could, in principle, be used at any stage in the hazard assessment process, the most likely use scenario will be when there are relatively few data available about the possible JH-disrupting properties of a chemical. The results from this assay are most likely to be available after deployment of a battery of *in vitro* screens, or as a supplement to existing data which suggest possible JH-related activity. Given the significant degree of endocrine system conservation across the arthropods, endocrine-linked effects in the SJHASA may also indicate the possibility of related activity in other arthropods such as copepods, decapods and insects.

487. It is possible that no endocrine-relevant data are available before the SJHASA is deployed (i.e. if the SJHASA has been used as a primary screen), but in that case a positive result in the screen should probably be followed up with relevant *in vitro* screening, if available, to investigate the suspected mode of action (MOA) in more detail. However, it should be noted that there are no standardised *in vitro* screens for JH agonists, although some are described in the scientific literature (for example, Cherbas, Koehler and Cherbas [1989]; Miyakawa and Iguchi [2017]).

488. In order to provide information relevant for assessing whether or not a chemical may fulfil the WHO/IPCS (2002) definition of an ED, the study design has to be sufficiently robust to demonstrate the presence or absence of effects. In the dose selection, the investigator should also consider and ensure that data generated are adequate to fulfil the regulatory requirement across OECD countries as appropriate (e.g. hazard and risk assessment and labelling, ED assessment, etc.). The top dose or concentration should be sufficiently high to give clear systemic (i.e. non endocrine-specific) toxicity in order to ensure that a wide range of exposures (high to low) is tested. However, endocrine effects observed solely in the presence of clear systemic toxicity should be interpreted with caution and may be disregarded when sufficiently justified to be caused by secondary effects which are unlikely to be due to endocrine activity. The reason for this advice is a concern that some endocrine active substance (EAS) sensitive assays are being run at doses/concentrations of EASs that are too low to trigger direct impacts on the endocrine system. This guidance document is not the place to address this issue directly, but it should be considered when EAS-sensitive test guidelines (TGs) are revised in the future. In addition, the number and spacing of dose/concentration levels should also be adequate to fulfil the objectives of the study (e.g. to demonstrate dose response relationships if this is required).

Existing data to be considered

489. Existing information on endocrine-related effects from other arthropods should also be considered before deployment of the SJHASA, given the commonality of endocrine mechanisms in these taxa. Existing data available might also include one or more of a range of *in silico* or *in vitro* results which suggest that JH disruption may occur *in vivo* (but note the limitations of this approach, as indicated above). Such indicators of possible JH activity might include quantitative structure activity relationship (QSAR) predictions of JH activity, "read-across" from *in vivo* results obtained with structurally related chemicals or positive results from an *in vitro* screen for JH agonist activity.

Scenarios: Positive and negative results combined with existing data

490. The scenarios (A to R) presented in Table C.2.14 represent all the possibilities of positive or negative results in combination with the presence or absence of existing data. The action taken will also depend on the regulatory environment, but the considerations given here are generally science based. Wherever possible, the recommended "next step which could be taken" avoids unnecessary animal testing. However, sometimes conducting an animal test will be indicated and then the relevance of species, strain and exposure route should always be considered. Further considerations specific to each scenario are given in the table.

491. Positive results obtained with the SJHASA (Table C.2.14, Scenarios A-I) result in the conclusion that the test chemical is a possible JH disrupter *in vivo*, at least in crustaceans. However, as indicated above, although a positive response of the SJHASA indicates that the chemical is a possible JH agonist, a result of this type would generally need to be followed up with a more comprehensive screen. The most appropriate choice for this is the *Daphnia* Multigeneration Test (DMGT – draft OECD TG). However, if countries need further evidence concerning growth and sexual development, etc., a Harpacticoid Copepod Development and Reproduction Test – OECD GD 201 and/or the Sediment-Water Chironomid Life Cycle Toxicity Test – OECD TG 233 would be able to provide a precise no-observed-effect-concentration/x% effect concentration (NOEC/ECx)

for adverse effects. This may be particularly important because *Daphnia* are parthenogenic under certain circumstances, while *Amphiascus* and *Chironomus* reproduce sexually. In other words, in order to strengthen weight of evidence, a positive result in the SJHASA could be followed by the DMGT at Level 3, which if positive in turn might lead to conduct of OECD TG 233 (Level 5). Existing data suggesting endocrine-specific activity (e.g. positive *in vitro* data, or positive *in vivo* data from other species) will strengthen the case for additional testing still further.

492. The situation in which the SJHASA gives a negative result (Table C.2.14, Scenarios J-R) needs careful consideration of any existing data. If these data suggest that the chemical is endocrine active both *in vitro* and *in vivo* (Scenario J), then it is possible that the SJHASA is simply insufficiently sensitive.

493. If the SJHASA and existing *in vivo* data are all negative, but *in vitro* data reveal some JH activity (Scenario K), the probability is that the test chemical is not sufficiently potent to produce JH agonism *in vivo* in arthropods, or it may be rapidly metabolised. In such a situation, further testing is probably not necessary. However, if the chemical is known to bioaccumulate slowly, it may be that exposures in the *in vivo* tests have been insufficiently prolonged, in which case longer term testing with OECD TG 201 or OECD TG 233 might be justified.

494. On the other hand, if the SJHASA and the *in vitro* tests are negative (Scenario M), but there are positive existing *in vivo* data, the nature of those existing data should be considered. Unless the existing data are from another crustacean, the chemical is possibly not a JH agonist acting in crustaceans, but it may be more potent in species (e.g. insects) or life stages that have not been tested. In this situation, the existing *in vivo* data should be used to guide decisions about whether to conduct any further testing.

495. Finally, a negative SJHASA, set against a background of negative *in vitro* and *in vivo* data (Scenario N), suggests that the test chemical is probably not a JH agonist *in vitro* or *in vivo*, and further action is unnecessary.

496. In each of the above scenarios, it is possible that existing data will be equivocal, or there may be no existing data. This will weaken the conclusions which can be drawn about a negative SJHASA, and this is reflected in Table C.2.14. However, a lack of mechanistic data on JH activity should ideally be rectified before any further *in vivo* testing is finally conducted, although as indicated above, *in vitro* JH screens have not yet been internationally standardised. On the other hand, if the SJHASA is positive, further *in vivo* testing would generally be needed to quantify any adverse effects and/or to establish a NOEC or ECx for such effects, even if all existing data are equivocal, or if there are no existing data. Again, however, it may be useful to obtain some mechanistic information before conducting further *in vivo* testing. There is also the possibility that equivocal mechanistic data may be the result of multiple modes of endocrine action. Under some circumstances, two opposite modes of simultaneous action (e.g. JH agonistic and antagonistic) could, depending on dose, lead to a minimisation or abolition of adverse effects, while in others two different MOA could potentially reinforce effects on the SJHASA endpoint. If multiple MOA are suspected, either from the existing results or based on QSAR/read-across/integrated approaches, this situation should be investigated further if needed for regulatory decision making.

497. The scenario in which the results of the SJHASA are themselves equivocal has not been dealt with in Table C.2.14, for reasons of brevity. In this context, an equivocal result might be an inconsistent concentration-response (e.g. no effect at a high concentration but

effects at a lower concentration), or a result which borders on statistical significance. Without knowing the exact circumstances, reliable advice cannot be given, but the opinions of an experienced ecotoxicologist should be sought. Clearly, however, such equivocal results do not necessarily rule out the existence of *in vivo* endocrine activity. If possible reasons for false negatives are suspected, the SJHASA could be repeated (e.g. conduct it at lower concentrations which avoid systemic toxicity). However, note that a repeat screen in the event of systemic toxicity would not be needed providing at least one tested concentration was not subject to such effects. It should also be borne in mind that changing environmental conditions such as shortening photoperiod, temperature and food shortages can also cause the production of male neonates in *D. magna*, so if these have accidentally occurred during the test, the results should be treated as suspect.

498. In summary, positive results in the SJHASA may indicate that a chemical is endocrine active *in vivo* via JH agonism. This suggests that more comprehensive *in vivo* testing would be needed if the intention is to derive a long-term NOEC/ECx and/or to confirm whether or not the chemical is an actual endocrine disrupter in arthropods due to the occurrence of adverse effects. Negative results in the SJHASA do not necessarily mean that the chemical is not a potential ED – a judgement about the endocrine disruption potential in other arthropods (especially sexually reproducing species) and the possible need for additional testing will have to be made based on a weight of evidence evaluation of existing *in vitro* and *in vivo* data.

References

Cherbas, L., M.M.D. Koehler and P. Cherbas (1989), "Effects of juvenile hormone on the ecdysone response of *Drosophila* Kc cells", *Developmental Genetics*, Vol. 10/3, pp. 177-188, https://doi.org/10.1002/dvg.1020100307.

Miyakawa, H. and T. Iguchi (2017), "Comparative luciferase assay for establishing reliable *in vitro* screening system of juvenile hormone agonists", *Journal of Applied Toxicology*, Vol. 37/9, pp. 1082-1090, https://doi.org/10.1002/jat.3459.

WHO/IPCS (2002), "Global assessment of the state-of-the-science of endocrine disrupters", Damstra, T. et al. (eds.) WHO/PCS/EDC/02.2, World Health Organization, Geneva, www.who.int/ipcs/publications/new_issues/endocrine_disruptors/en.

Table C.2.14. **Short-Term Juvenile Hormone Activity Screening Assay using *Daphnia magna* (SJHASA) (draft OECD TG): Guidance for scenarios of combinations of results with existing data**

This table represents possible conclusions to be drawn from assay data, and a next step which could be taken if further evidence is required about possible endocrine disrupting properties and/or effects. The guidance offered is not meant to be prescriptive, but provides science-based considerations. It encourages the use of all available data and expert judgement in a weight of evidence approach. Regional and national interpretation of results and "next steps" may vary.

The conclusions are grouped into a series of scenarios (A-R), each scenario representing a different combination of assay results, existing *in vitro* data and existing *in vivo* data. The symbol "+" indicates that the data in question represent a positive result, «-" indicates a negative result, and "Eq/0" indicates that the data are either equivocal or are not available.

Existing results: * "Mechanism (*in vitro* mechanistic data)" assumes that mechanistic data are available from available from juvenile hormone (JH-) based assays. JH assays concerning mechanisms of JH disruption may be available, but they are have not yet been internationally standardised. In practice, data from all assays may not be available and therefore this must be taken into account when deciding on the "next step".

Existing results: ** "Effects (*in vivo* effects of concern)" assumes effects have been observed in other *in vivo* screens/tests which give rise to concern that the test chemical may be a JH disrupter.

Scenarios	Result of SJHASA	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
A	+	+	+	Strong evidence for *in vivo* juvenile hormone (JH) activity in crustaceans, plus possible JH effects in other arthropods.	Consider performing a *Daphnia* Multigeneration Test (DMGT – draft OECD TG).	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is a JH agonist. The DMGT will show if sex ratio bias towards males carries over into the F2 generation and some regulatory authorities may consider that this provides sufficient information on adverse apical effects in crustaceans. However, as *Daphnia* are parthenogenetic, it would be desirable to perform an additional apical test with sexually reproducing crustaceans and/or insects (e.g. the Harpacticoid Copepod Development and Reproduction Test – OECD GD 201; and/or the Sediment-Water Chironomid Life Cycle Toxicity Test – OECD TG 233).
B	+	+	–	Strong evidence for *in vivo* JH activity in crustaceans.	Consider performing a DMGT (draft OECD TG).	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is a JH agonist. The DMGT will show if sex ratio bias towards males carries over into the F2 generation and some regulatory authorities may consider that this provides sufficient information on adverse apical effects in crustaceans. However, as *Daphnia* are parthenogenetic, it would be desirable to perform an additional apical test with sexually reproducing crustaceans and/or insects (e.g. the Harpacticoid Copepod Development and Reproduction Test – OECD GD 201; and/or the Sediment-Water Chironomid Life Cycle Toxicity Test – OECD TG 233).
C	+	+	Eq/0	Strong evidence for *in vivo* JH activity in crustaceans.	Consider performing a DMGT (draft OECD TG).	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is a JH agonist. The DMGT will show if sex ratio bias towards males carries over into the F2 generation and some regulatory authorities may consider that this provides sufficient information on adverse apical effects in crustaceans. However, as *Daphnia* are parthenogenetic, it would be desirable to perform an additional apical test with sexually reproducing crustaceans and/or insects (e.g. the Harpacticoid Copepod Development and Reproduction Test – OECD GD 201; and/or the Sediment-Water Chironomid Life Cycle Toxicity Test – OECD TG 233). It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple modes of action (MOA). If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
D	+	–	+	Moderate evidence for *in vivo* JH activity in crustaceans, plus possible JH effects in other arthropods.	Consider performing a DMGT (draft OECD TG).	The lack of *in vitro* JH activity is not evidence against any JH activity, due to the limited nature of current *in vitro* JH screens. The DMGT will show if sex ratio bias towards males carries over into the F2 generation and some regulatory authorities may consider that this provides sufficient information on adverse apical effects in crustaceans. However, as *Daphnia* are parthenogenetic, it would be desirable to perform an additional apical test with sexually reproducing crustaceans and/or insects (e.g. the Harpacticoid Copepod Development and Reproduction Test – OECD GD 201; and/or the Sediment-Water Chironomid Life Cycle Toxicity Test – OECD TG 233).

Scenarios	Result of SJHASA	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
E	+	–	–	Possible evidence for *in vivo* JH activity in crustaceans.	Consider performing a (DMGT (draft OECD TG).	The lack of *in vitro* JH activity is not evidence against any JH activity, due to the limited nature of current *in vitro* JH screens. The DMGT will show if sex ratio bias towards males carries over into the F2 generation and some regulatory authorities may consider that this provides sufficient information on adverse apical effects in crustaceans. However, as *Daphnia* are parthenogenetic, it would be desirable to perform an additional apical test with sexually reproducing crustaceans and/or insects (e.g. the Harpacticoid Copepod Development and Reproduction Test – OECD GD 201; and/or the Sediment-Water Chironomid Life Cycle Toxicity Test – OECD TG 233).
F	+	–	Eq/0	Possible evidence for *in vivo* JH activity in crustaceans.	Consider performing a DMGT (draft OECD TG). Given the absence or equivocal nature of existing *in vivo* data, it might also be sensible to conduct a JH-responsive insect assay (e.g. the Sediment-Water Chironomid Life Cycle Toxicity Test – OECD TG 233).	The lack of *in vitro* JH activity is not evidence against any JH activity, due to the limited nature of current *in vitro* JH screens. The DMGT will show if sex ratio bias towards males carries over into the F2 generation and some regulatory authorities may consider that this provides sufficient information on adverse apical effects in crustaceans. However, as *Daphnia* are parthenogenetic, it would be desirable to perform an additional apical test with sexually reproducing crustaceans (e.g. the Harpacticoid Copepod Development and Reproduction Test – OECD GD 201) It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
G	+	Eq/0	+	Moderate evidence for *in vivo* JH activity in crustaceans, plus possible JH effects in other arthropods.	Consider performing a DMGT (draft OECD TG). Given the absence or equivocal nature of the *in vitro* mechanistic data, it might also be helpful to conduct an *in vitro* screen for JH activity.	If a new *in vitro* mechanistic assay is conducted, note that a negative does not mean that the test material has no JH activity. The DMGT will show if sex ratio bias towards males carries over into the F2 generation and some regulatory authorities may consider that this provides sufficient information on adverse apical effects in crustaceans. However, as *Daphnia* are parthenogenetic, it would be desirable to perform an additional apical test with sexually reproducing crustaceans (e.g. the Harpacticoid Copepod Development and Reproduction Test – OECD GD 201 and/or the Sediment-Water Chironomid Life Cycle Toxicity Test – OECD TG 233). It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
H	+	Eq/0	–	Possible evidence for *in vivo* JH activity in crustaceans.	Consider performing a DMGT (draft OECD TG). Given the absence or equivocal nature of the *in vitro* mechanistic data, it might also be helpful to conduct an *in vitro* screen for JH activity.	The DMGT will show if sex ratio bias towards males carries over into the F2 generation and some regulatory authorities may consider that this provides sufficient information on adverse apical effects in crustaceans. However, as *Daphnia* are parthenogenetic, it would be desirable to perform an additional apical test with sexually reproducing crustaceans (e.g. the Harpacticoid Copepod Development and Reproduction Test – OECD GD 201 and/or the Sediment-Water Chironomid Life Cycle Toxicity Test – OECD TG 233). It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

Scenarios	Result of SJHASA	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
I	+	Eq/0	Eq/0	Possible evidence for *in vivo* JH activity in crustaceans.	Consider performing a DMGT (draft OECD TG). Given the absence or equivocal nature of the *in vitro* mechanistic data, it might also be helpful to conduct an *in vitro* screen for JH activity.	If a new *in vitro* mechanistic assay is conducted, note that a negative does not mean that the test material has no JH activity. The DMGT will show if sex ratio bias towards males carries over into the F2 generation and some regulatory authorities may consider that this provides sufficient information on adverse apical effects in crustaceans. However, as *Daphnia* are parthenogenetic, it would be desirable to perform an additional apical test with sexually reproducing crustaceans (e.g. the Harpacticoid Copepod Development and Reproduction Test – OECD GD 201 and/or the Sediment-Water Chironomid Life Cycle Toxicity Test – OECD TG 233). It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
J	–	+	+	The test chemical is probably a JH agonist without activity in crustaceans, although it is possible that *Daphnia magna* responds atypically in this case.	Some regulatory authorities may conclude that no further evidence is required. However, it might be desirable to obtain data from insects (e.g. the Sediment-Water Chironomid Life Cycle Toxicity Test – OECD TG 233) if these are not already available.	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is a JH agonist.
K	–	+	–	The test chemical is likely to have JH activity; however, without demonstrating sufficient activity to disrupt physiological processes *in vivo*.	If there is no activity in crustaceans or insects, further evidence is probably not needed.	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is a JH agonist.
L	–	+	Eq/0	The test chemical is likely to have JH activity; however, without demonstrating sufficient activity to disrupt physiological processes *in vivo*.	Some regulatory authorities may conclude that no further evidence is required, but if insect data are absent, it might be desirable to conduct a Sediment-Water Chironomid Life Cycle Toxicity Test – OECD TG 233.	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is a JH agonist. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
M	–	–	+	The test chemical is probably without JH activity in crustaceans, although it is possible that *Daphnia magna* responds atypically in this case.	Some regulatory authorities may conclude that no further evidence is required. However, it might be desirable to obtain data from insects (e.g. the Sediment-Water Chironomid Life Cycle Toxicity Test – OECD TG 233) if these are not already available.	The lack of *in vitro* JH activity is not evidence against any JH activity, due to the limited nature of current *in vitro* JH screens. However, it is possible that the existing effects may not be due to JH activity.

Scenarios	Result of SJHASA	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
N	–	–	–	The test chemical is probably without JH activity in arthropods.	No further action is necessary.	–
O	–	–	Eq/0	The test chemical is probably without JH activity in arthropods.	Some regulatory authorities may conclude that no further evidence is required. However, it might be desirable to obtain data from insects (e.g. the Sediment-Water Chironomid Life Cycle Toxicity Test – OECD TG 233) if these are not already available.	The lack of *in vitro* JH activity is not evidence against any JH activity, due to the limited nature of current *in vitro* JH screens. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
P	–	Eq/0	+	The test chemical is probably without JH activity in crustaceans, although it is possible that *Daphnia magna* responds atypically in this case.	Some regulatory authorities may conclude that no further evidence is required. Also, if clear *in vitro* mechanistic data are missing, it might be desirable to obtain some.	If a new *in vitro* mechanistic assay is conducted, note that a negative does not mean that the test material has no JH activity. However, it is possible that the existing effects may not be due to JH activity. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
Q	–	Eq/0	–	The test chemical is probably without JH activity in arthropods.	No further action is necessary.	It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
R	–	Eq/0	Eq/0	The test chemical is probably without JH activity in crustaceans and possibly insects.	Some regulatory authorities may conclude that no further evidence is required. However, it might be desirable to obtain data from insects (e.g. the Sediment-Water Chironomid Life Cycle Toxicity Test – OECD TG 233) if these are not already available.	It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

C.2.15. Androgenised Female Stickleback Screen (AFSS) (GD 148) (variant of OECD TG 230)

Status: Partially validated by the OECD.

499. Modality detected/endpoints: androgens (♀ spiggin ↑); anti-androgens (androgenised ♀ spiggin ↓).

Background to the assay

500. This assay is designed primarily as a screen for chemicals with *in vivo* anti-androgenic activity in fish, but it is also able to detect androgens. It has partially completed validation and has been published as an OECD guidance document (GD 148). The endpoints are indicators of hormonal activity and there are no apical measures of adverse effects **diagnostic of a specific estrogen/androgen/thyroid/steroidogenesis (E,A,T,S) modality**. This assay is a variant of the 21-Day Fish Assay (OECD TG 230) with a more limited range of endpoints, but it has more power to identify anti-androgens than OECD TG 229 or TG 230. An alternative *in vivo* assay with the scope for identifying anti-androgens is the Juvenile Medaka Anti-Androgen Screening Assay (JMASA).

When/why the assay may be used

501. Although the AFSS could, in principle, be used at any stage in the hazard assessment process, the most likely use scenario will be when there are relatively few data available about the possible endocrine disrupting properties of a chemical. The assay is most likely to be used either as part of a battery of *in vitro* and *in vivo* screens, or to follow up on existing data which suggest possible endocrine disruption activity at the androgen receptor. It would not be necessary for aquatic exposure to have been predicted (because a positive in the AFSS could potentially be extrapolated to terrestrial vertebrates), but such a prediction would provide additional justification for running the screen. It is also possible that no existing endocrine-relevant data are available (i.e. the AFSS has been used as a primary screen), but in that case a positive result in the screen should ideally be followed up with relevant *in vitro* screening in an attempt to confirm the suspected (anti)androgenic mode of action (MOA). Given the high degree of endocrine system conservation across the vertebrates, endocrine-linked effects in the AFSS may also indicate the possibility of related activity in other organisms such as amphibians, reptiles, birds or mammals.

502. In order to provide information relevant for assessing whether or not a chemical may fulfil the WHO/IPCS (2002) definition of an endocrine disruptor (ED), the study design has to be sufficiently robust to demonstrate the presence or absence of effects. In the dose selection, the investigator should also consider and ensure that data generated are adequate to fulfil the regulatory requirement across OECD countries as appropriate (e.g. hazard and risk assessment and labelling, ED assessment, etc.). The top dose or concentration should be sufficiently high to give clear systemic (i.e. non endocrine-specific) toxicity in order to ensure that a wide range of exposures (high to low) is tested. However, endocrine effects

observed solely in the presence of clear systemic toxicity should be interpreted with caution and may be disregarded when sufficiently justified to be caused by secondary effects which are unlikely to be due to endocrine activity. The reason for this advice is a concern that some endocrine active substance (EAS) sensitive assays are being run at doses/concentrations of EASs that are too low to trigger direct impacts on the endocrine system. This guidance document is not the place to address this issue directly, but it should be considered when EAS-sensitive test guidelines (TGs) are revised in the future. In addition, the number and spacing of dose/concentration levels should also be adequate to fulfil the objectives of the study (e.g. to demonstrate dose response relationships if this is required).

Existing data to be considered

503. Given the commonality of endocrine mechanisms in the vertebrates, relevant existing data available before deployment of the AFSS might include *in vivo* results obtained with other vertebrates (e.g. a positive rodent Hershberger Bioassay – OECD TG 441, positive findings for endocrine endpoints in mammalian repeat dose toxicity or reproductive studies), or one or more of a range of *in silico* or *in vitro* results which suggest that the modalities indicated above may occur *in vivo*. Such indicators of possible *in vivo* activity might include quantitative structure activity relationship (QSAR) predictions of endocrine activity, high throughput screening data, "read-across" from *in vivo* results obtained with structurally related chemicals or positive results from an *in vitro* screen for androgen receptor-mediated activity.

Scenarios: Positive and negative results combined with existing data

504. The scenarios (A to R) presented in Table C.2.15 represent all the possibilities of positive or negative results in combination with the presence or absence of existing data. The action taken will also depend on the regulatory environment, but the considerations given here are generally science based. Wherever possible, the recommended "next step which could be taken" avoids unnecessary animal testing. However, sometimes conducting an animal test will be indicated and then the relevance of species, strain and exposure route should always be considered. Further considerations specific to each scenario are given in the table.

505. Positive results obtained with one of the endpoints (Table C.2.15, Scenarios A-I) result in the conclusion that the test chemical is a possible androgen or anti-androgen *in vivo*. If a regulatory authority required more evidence, positive results in the AFSS should be followed up with more comprehensive testing to show whether adverse apical effects occur at any part of the life cycle (and hence to provide evidence supporting a conclusion that the chemical is an actual ED). In other words, to increase confidence, a positive result in the AFSS would trigger fish life cycle testing at Level 5 (OECD TG 240 – MEOGRT or ZEOGRT), or possibly a Fish Sexual Development Test (FSDT) (OECD TG 234) at Level 4 if it is suspected that the most responsive part of the life cycle is sexual development. Existing data suggesting (anti)androgenic activity will strengthen the case for additional testing still further.

506. The situation in which the AFSS gives a negative result (Table C.2.15, Scenarios J-R) needs careful consideration of any existing data. If these data suggest that the chemical is (anti)androgenic both *in vitro* and *in vivo* (Scenario J), then the probability is that the AFSS is simply insufficiently sensitive. It might in these circumstances be appropriate to

conduct OECD TG 234 (FSDT), or alternatively, a fish life cycle test (OECD TG 240 – MEOGRT or ZEOGRT) to confirm that there is no endocrine activity in fish.

507. If the AFSS and existing *in vivo* data are all negative, but *in vitro* data reveal some (anti)androgenic activity (Scenario K), the probability is that the test chemical is not sufficiently potent to produce endocrine effects *in vivo* in fish or other organisms, or it may be rapidly metabolised or simply does not reach the receptor. In such a situation, further testing is probably not necessary. However, if the chemical is known to bioaccumulate slowly, it may be that exposures in the *in vivo* tests have been insufficiently prolonged, in which case longer term testing might be justified. Equally, if existing data suggest thyroid activity, consideration should be given to conducting the Amphibian Metamorphosis Assay (OECD TG 231).

508. On the other hand, if the AFSS and the *in vitro* tests are negative (Scenario M), but there are positive existing *in vivo* data, the chemical is probably not an ED with (anti)androgenic activity, but it may act via modes of action (MOA) not covered by the *in vitro* screens, or it may be more potent in species or life stages that have not been tested. In this situation, the existing *in vivo* data should be used to guide decisions about whether to conduct any further testing, either for modalities such as thyroid activity, or including life stages represented in TG 234 (FSDT) or in the MEOGRT or ZEOGRT.

509. Finally, a negative AFSS, set against a background of negative *in vitro* and *in vivo* data (Scenario N), suggests that the test chemical is not (anti)androgenic in fish, and no further testing for this modality will generally be necessary. It remains possible that it has thyroid activity, although if any existing tests for this modality are negative, it would suggest that this scenario is unlikely.

510. In each of the above scenarios, it is possible that existing data will be equivocal (Scenarios C, F-I, L and O-R), or there may be no existing data. This will weaken the conclusions which can be drawn about a negative AFSS, and this is reflected in Table C.2.15. However, a lack of mechanistic data on (anti)androgenic activity should ideally be rectified before any further *in vivo* testing is considered. On the other hand, if the AFSS is positive, further *in vivo* testing to obtain more evidence is generally desirable even if all existing data are equivocal, or if there are no existing data. Again, however, it will always be helpful to obtain some mechanistic information before conducting further *in vivo* testing. There is also the possibility that equivocal mechanistic data may be the result of multiple modes of endocrine action. Under some circumstances, two opposite modes of simultaneous action (e.g. androgenic and anti-androgenic) could, depending on dose, lead to a minimisation or abolition of effects, while in others two different MOA (e.g. estrogenic and anti-androgenic) could potentially reinforce effects. If multiple MOA are suspected, either from the existing results or based on QSAR/read-across/integrated approaches, this situation should be investigated further if needed for regulatory decision making.

511. The scenario in which the results of the AFSS are themselves equivocal has not been dealt with in Table C.2.15, for reasons of brevity. In this context, an equivocal result might be an inconsistent concentration-response (e.g. no effect at a high concentration but effects at a lower concentration), or a result which borders on statistical significance. Without knowing the exact circumstances, reliable advice cannot be given, but the opinions of an experienced ecotoxicologist should be sought. Clearly, however, such equivocal results do not necessarily rule out the existence of *in vivo* endocrine activity. For example, spiggin induction in females at a high concentration might be masked by any systemic toxicity (although it would not be sensible to run the assay at such high concentrations), while spiggin depression in androgenised females might just fail to reach a statistically

significant level because spiggin levels were relatively low to begin with. If these or other possible reasons for false negatives are suspected with good reason, the screen could be repeated (e.g. conduct it at lower concentrations which avoid systemic toxicity, assuming systemic toxicity in the original test occurred at all concentrations), or a more appropriate version of it (e.g. ensure androgenised females have high spiggin levels at the start of the test) could be conducted. In particular, it might be appropriate to run the JMASA if anti-androgenic activity is suspected.

512. In summary, positive results in the AFSS indicate that a chemical is a possible (anti)androgen. If a regulatory authority required further evidence, more comprehensive *in vivo* testing would then be necessary to produce a long-term NOEC/ECx for adverse effects and/or to confirm whether or not the chemical is an actual (anti)androgen. Negative results in the AFSS do not necessarily mean that the chemical is not a possible (anti)androgen – a judgement about this will have to be made in the light of existing *in vitro* and *in vivo* data.

Reference

WHO/IPCS (2002), "Global assessment of the state-of-the-science of endocrine disrupters", Damstra, T. et al. (eds.) WHO/PCS/EDC/02.2, World Health Organization, Geneva, www.who.int/ipcs/publications/new_issues/endocrine_disruptors/en.

Table C.2.15. **Androgenised Female Stickleback Screen (AFSS) (OECD GD 148) (variant of OECD TG 230): Guidance for scenarios of combinations of results with existing data**

This table represents possible conclusions to be drawn from assay data, and a next step which could be taken if further evidence is required about possible endocrine disrupting properties and/or effects. The guidance offered is not meant to be prescriptive, but provides science-based considerations. It encourages the use of all available data and expert judgement in a weight of evidence approach. Regional and national interpretation of results and "next steps" may vary.

The conclusions are grouped into a series of scenarios (A-R), each scenario representing a different combination of assay results, existing *in vitro* data and existing *in vivo* data. The symbol "+" indicates that the data in question represent a positive result, «-" indicates a negative result, and "Eq/0" indicates that the data are either equivocal or are not available.

Existing results: * "Mechanism (*in vitro* mechanistic data)" assumes that mechanistic data are available from estrogen receptor (ER-), androgen receptor (AR-) and steroidogenesis-based assays (Level 2). Thyroid hormone receptor (TR) and other assays concerning mechanisms of thyroid disruption may be available, but they are not in common use. In practice, data from all assays may not be available and therefore this must be taken into account when deciding on the "next step". Quantitative structure activity relationship (QSAR) predictions of estrogen and androgen binding/activation may be made for some substances. There is no evidence at present that equivalent *in vitro* assays with systems derived from fish offer advantages over their mammalian counterparts.

Existing results: ** "Effects (*in vivo* effects of concern)" assumes effects have been observed in other *in vivo* screens/tests which give rise to concern that the test chemical may be an endocrine disrupter.

Note that this assay has been successfully validated, but it has not been published as an OECD test guideline.

Scenario	Result of AFSS	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
A	+	+	+	Strong evidence for *in vivo* (anti)androgenic activity in fish and other organisms.	Consider performing fish life cycle test (ZEOGRT or MEOGRT – OECD TG 240).	An alternative approach would be to deploy TG 234 (Fish Sexual Development Test [FSDT]), especially if sexual development is expected to give a response at lower concentrations than reproduction.
B	+	+	–	Strong evidence for *in vivo* (anti)androgenic activity in fish.	Consider performing fish life cycle test (ZEOGRT or MEOGRT – OECD TG 240).	An alternative approach would be to deploy TG 234 (FSDT), especially if sexual development is expected to give a response at lower concentrations than reproduction. If the negative *in vivo* data are from a fish test (e.g. OECD TG 229 or TG 230), consider possible reasons for the disparity (e.g. differences in species sensitivity) before possibly conducting a life cycle test (OECD TG 240 – MEOGRT or ZEOGRT) or OECD TG 234 (FSDT).
C	+	+	Eq/0**	Strong evidence for *in vivo* (anti)androgenic activity in fish.	Consider performing fish life cycle test (MEOGRT – OECD TG 240, or ZEOGRT) or OECD TG 234 (FSDT).	If no existing fish data are available, it may be worth performing OECD TG 234 (FSDT) before a life cycle test in order to obtain information on whether sexual development is a sensitive part of the life cycle. Such information could influence the design of a life cycle test (MEOGRT or ZEOGRT). It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple modes of action (MOA). If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
D	+	–	+	Strong evidence for *in vivo* (anti)androgenic activity in fish and other organisms, but negative *in vitro* data suggest MOA may not be via interaction with the androgen receptor or interference with steroidogenesis, or that the test chemical may be metabolically activated *in vivo*.	Consider performing fish life cycle test (MEOGRT – OECD TG 240 or ZEOGRT).	An alternative approach would be to deploy TG 234 (FSDT), especially if sexual development is expected to give a response at lower concentrations than reproduction.
E	+	–	–	Strong evidence for *in vivo* (anti)androgenic activity in fish, but negative existing data raise doubts about the MOA, or suggest that the test chemical may be metabolically activated *in vivo*.	Consider performing fish life cycle test (MEOGRT – OECD TG 240 or ZEOGRT).	An alternative approach would be to deploy OECD TG 234 (FSDT), especially if sexual development is expected to give a response at lower concentrations than reproduction. If the negative *in vivo* data are from a fish test (e.g. OECD TG 229 or TG 230), consider possible reasons for the disparity (e.g. differences in species sensitivity) before possibly conducting a life cycle test (FLCTT or OECD TG 240 – MEOGRT, or ZEOGRT) or TG 234 (FSDT).
F	+	–	Eq/0	Strong evidence for *in vivo* (anti)androgenic activity in fish, but negative or equivocal existing data raise doubts about the MOA, or suggest that the test chemical may be rapidly degraded in water or metabolically activated *in vivo*.	Consider performing a fish life cycle test (MEOGRT – OECD TG 240 or ZEOGRT) or TG 234 (FSDT).	If no existing fish data are available, it may be worth performing OECD TG 234 (FSDT) before a life cycle test (MEOGRT or ZEOGRT) in order to obtain information on whether sexual development is a sensitive part of the life cycle. Such information could influence the design of a life cycle test. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

Scenario	Result of AFSS	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
G	+	Eq/0	+	Strong evidence for *in vivo* (anti)androgenic activity in fish and other organisms, but mechanism unconfirmed.	Obtain more predictive mechanistic data and then consider performing fish life cycle test (MEOGRT – OECD TG 240 or ZEOGRT).	An alternative approach would be to deploy OECD TG 234 (FSDT). It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
H	+	Eq/0	–	Strong evidence for *in vivo* (anti)androgenic activity in fish, but mechanism unconfirmed.	Obtain more predictive mechanistic data and then consider performing fish life cycle test (MEOGRT – OECD TG 240 or ZEOGRT).	An alternative approach would be to deploy OECD TG 234 (FSDT), especially if sexual development is expected to give a response at lower concentrations than reproduction. If the negative *in vivo* data are from a fish test (e.g. OECD TG 229 or TG 230), consider possible reasons for the disparity (e.g. differences in species sensitivity) before possibly conducting a life cycle test (MEOGRT or ZEOGRT) or TG 234 (FSDT). It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
I	+	Eq/0	Eq/0	Strong evidence for *in vivo* (anti)androgenic activity in fish, but mechanism unconfirmed.	Obtain more predictive mechanistic data and then consider performing fish life cycle test (MEOGRT – OECD TG 240 or ZEOGRT) or TG 234 (FSDT).	If no existing fish data are available, it may be worth performing OECD TG 234 (FSDT) before a life cycle test in order to obtain information on whether sexual development is a sensitive part of the life cycle. Such information could influence the design of a life cycle test (MEOGRT or ZEOGRT). It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
J	–	+	+	No evidence for (anti)androgenic activity *in vivo* in fish. However, the chemical is an (anti)androgen in other species and this mechanism has been confirmed *in vitro*.	Consider performing OECD TG 234 (FSDT).	It is possible that the failure to give a positive result in the AFSS was caused by the relatively short exposure time (three weeks). If this is suspected, it is worth considering whether to perform a fish life cycle test (MEOGRT or ZEOGRT) or OECD TG 234 (FSDT). Test design should be guided by the existing *in vivo* data.
K	–	+	–	There is no evidence that the chemical is an (anti)androgen *in vivo*, probably because it is very weakly acting or rapidly metabolised or degraded in water.	Probably no further action, but see comments in right-hand column.	It is possible that ndocrine disruptors which bioaccumulate slowly may only cause effects *in vivo* after exposure times longer than three weeks. If this is suspected, and depending on which part of the life cycle is suspected of being the most sensitive, consider performing OECD TG 234 (FSDT) or a fish life cycle test (MEOGRT – OECD TG 240, or ZEOGRT). It is also possible that the chemical may be a thyroid-active chemical *in vivo* (consider performing the Amphibian Metamorphosis Assay [AMA] – OECD TG 231), an (anti)estrogen, or an aromatase inhibitor (consider performing OECD TG 229 or TG 230).

Scenario	Result of AFSS	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
L	–	+	Eq/0	The chemical may not be an (anti)androgen *in vivo*, but the confidence in this conclusion is relatively low as there is only one unequivocal *in vivo* test result (a negative).	Consider performing a fish assay (OECD TG 229 or TG 230) with a different species, or consider a longer term test (TG 234 [FSDT] or life cycle [MEOGRT – OECD TG 240, or ZEOGRT]).	It is also possible that the chemical may be a thyroid-active chemical *in vivo* (consider performing the AMA – OECD TG 231, or *Xenopus* Embryonic Thyroid Signalling Assay [XETA]), an (anti)estrogen, or an aromatase inhibitor (consider performing OECD TG 229, TG 230 or EASZY). It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
M	–	–	+	The chemical is probably not an (anti)androgen in fish. However, it may act through MOA not covered by the available *in vitro* assays, or it may be more potent in a species other than that tested, or over a longer exposure period.	Use the existing *in vivo* data to help choose a possible longer term test with an appropriate species.	It is also possible that the chemical may be a thyroid-active chemical *in vivo* (consider performing the AMA – OECD TG 231, or XETA), an (anti)estrogen, or an aromatase inhibitor (consider performing OECD TG 229, TG 230 or EASZY), although lack of *in vitro* binding affinity with the estrogen or androgen receptors suggests the two former possibilities are unlikely. Use the existing *in vivo* data to guide any further testing.
N	–	–	–	The chemical is probably not an (anti)androgen in fish or other organisms.	No further action with respect to (anti)androgenic MOA.	It is still possible that the chemical may be a thyroid-active chemical *in vivo* (consider performing the AMA – OECD TG 231, or XETA), an (anti)estrogen, or an aromatase inhibitor (consider performing OECD TG 229, TG 230 or EASZY), although lack of *in vitro* binding affinity with the estrogen or androgen receptors suggests the two former possibilities are unlikely.
O	–	–	Eq/0	The chemical is probably not an (anti)androgen in fish or other organisms.	Probably no further action. However, see comments in right-hand column.	If the paucity of *in vivo* data are a concern, performance of a screening test (OECD TG 229 or TG 230) with a different species, or a longer term test (i.e. TG 234 [FSDT] or life cycle [MEOGRT – OECD TG 240, or ZEOGRT]) could be considered. It is also possible that the chemical may be a thyroid-active chemical *in vivo* (consider performing the AMA – OECD TG 231, or XETA), an (anti)estrogen, or an aromatase inhibitor (consider performing OECD TG 229, TG 230 or EASZY), although lack of *in vitro* binding affinity with the estrogen or androgen receptors suggests the two former possibilities are unlikely. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

Scenario	Result of AFSS	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
P	–	Eq/0	+	The chemical is probably not an (anti)androgen in fish, but confidence in this conclusion is low given the lack of more predictive *in vitro* data and the availability of positive existing *in vivo* data.	Obtain more predictive mechanistic data, then consider possible further testing.	If the mechanistic data confirm that the chemical has potential (anti)androgenic action, consider conducting a fish assay (OECD TG 229 or TG 230) with another species, or a longer term test (TG 234 [FSDT] or life cycle [MEOGRT – OECD TG 240, or ZEOGRT]). Use the existing *in vivo* data as a guide to test design. If the mechanistic data reveal (anti)estrogenic/aromatase inhibition activity, perform a fish assay (OECD TG 229 or TG 230). If any existing data suggest thyroid activity, consider an AMA (OECD TG 231) or XETA. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
Q	–	Eq/0	–	The chemical is probably not an (anti)androgen in fish or other organisms, but the lack of more predictive mechanistic data are a concern.	Obtain more predictive mechanistic data, then consider possible further testing.	If the mechanistic data confirm that the chemical has potential (anti)androgenic action, consider conducting a fish assay (OECD TG 229 or TG 230) with another species, or a longer term test (TG 234 [FSDT] or life cycle [MEOGRT – OECD TG 240, or ZEOGRT]). Use the existing *in vivo* data as a guide to test design. If the mechanistic data reveal (anti)estrogenic/aromatase inhibition activity, perform a fish assay (OECD TG 229 or TG 230). If any existing data suggest thyroid activity, consider an AMA (OECD TG 231) or XETA. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
R	–	Eq/0	Eq/0	The chemical is probably not an (anti)androgen in fish, but confidence in this conclusion is low given the lack of more predictive *in vitro* and existing *in vivo* data.	Obtain more predictive mechanistic data, then consider possible further testing.	If the mechanistic data confirm that the chemical has potential (anti)androgenic action, consider conducting a fish assay (OECD TG 229 or TG 230) with another species, or a longer term test (TG 234 [FSDT] or life cycle [MEOGRT – OECD TG 240, or ZEOGRT]). Use the existing *in vivo* data as a guide to test design. If the mechanistic data reveal (anti)estrogenic/aromatase inhibition activity, perform a fish assay (OECD TG 229 or TG 230). If any existing data suggest thyroid activity, consider an AMA (OECD TG 231) or XETA. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

C.2.16. EASZY Assay: Detection of Substances Acting through Estrogen Receptors using Transge nic cyp19a1b GFP Zebrafish Embryos (draft OECD TG)

Status: Assay being validated by the OECD.

513. Modality detected/endpoints: This *in vivo* zebrafish assay is sensitive to estrogen receptor (ER) agonists; pro-estrogens that can be metabolised to become ER agonists; androgens that can be aromatised to ER agonists; and some non-aromatisable androgens. It relies on transgenic zebrafish embryos which fluoresce when exposed to an ER agonist.

Background to the assay

514. This assay is currently being validated by the OECD, and one round of validation with up to five participating laboratories for each test chemical was initiated in 2014. The assay is based on a transgenic zebrafish line expressing green fluorescent protein (GFP) under the control of the promoter of the ER-regulated cyp19a1b gene coding for brain aromatase. The newly fertilised embryos are exposed for 96 hours to dilutions of the test chemical, after which they are scanned using a fluorescence imaging microscope, and the intensity of fluorescence recorded. From the concentration-response curve, ECx concentrations are derived and relative estrogenic potency can be calculated.

When/why the assay may be used

515. Although data from EASZY could, in principle, be available at any stage in the hazard assessment process, the most likely scenario will be when there are relatively few data available about the possible endocrine disrupting properties of a chemical. The assay is most likely to be used either as part of a battery of *in vitro* and *in vivo* screens, or to follow up on existing data which suggest possible endocrine disruption activity. Given the high degree of endocrine system conservation across the vertebrates, adverse endocrine-linked effects in the EASZY Assay may also indicate the possibility of related activity in other organisms such as amphibians, reptiles, birds or mammals. It is also possible that no existing endocrine-relevant data are available (i.e. EASZY has been used as a primary screen), but in that case a positive result in the screen should ideally be followed up with relevant *in vitro* screening in an attempt to confirm the precise mode of action (MOA). Furthermore, a positive EASZY result would also need to be followed up with an additional *in vivo* fish test such as the Fish Short-Term Reproduction Assay (FSTRA – OECD TG 229) or Fish Sexual Development Test (FSDT – OECD TG 234), which will give some indication of any adverse apical effects. Possible conclusions to be derived from the results of EASZY, and guidance about potential additional studies to strengthen weight of evidence, are summarised in Table C.2.16.

516. Caution should be used when negative results are obtained with certain types of chemicals because absorption into the embryo via the chorion may have been impeded. Development of the OECD Fish Embryo Acute Toxicity (FET) test (OECD TG 236) with zebrafish showed that this applies in particular to chemicals with a molecular weight ≥3kDa

and a very bulky molecular structure. Absorption of these chemicals will take place at a higher rate after hatching, but delayed hatch may therefore also protect the embryo from estrogenic effects. Although it is known that fish embryos have some metabolic capacity (e.g. Weigt et al. [2011]), and that EASZY is able to detect pro-estrogens such as methoxychlor that require metabolic activation (Brion et al., 2012), metabolism may be less efficient than in juveniles and adults, so use of the test with endocrine disrupting chemicals that require metabolic activation may give some false negatives. Nonetheless, a recent study comparing the metabolism of two estrogenic substances (BPS and BP2) in zebrafish embryos and adults reported that metabolic profiles were qualitatively the same between embryos and adults, with no major differences, although the biotransformation of both molecules was more extensive in adults (Le Fol et al., 2017).

517. In order to provide information relevant for assessing whether or not a chemical may fulfil the WHO/IPCS (2002) definition of an endocrine disruptor (ED), the study design has to be sufficiently robust to demonstrate the presence or absence of effects. In the dose selection, the investigator should also consider and ensure that data generated are adequate to fulfil the regulatory requirement across OECD countries as appropriate (e.g. hazard and risk assessment and labelling, ED assessment, etc.). The top dose or concentration should be sufficiently high to give clear systemic (i.e. non endocrine-specific) toxicity in order to ensure that a wide range of exposures (high to low) is tested. However, endocrine effects observed solely in the presence of clear systemic toxicity should be interpreted with caution and may be disregarded when sufficiently justified to be caused by secondary effects which are unlikely to be due to endocrine activity. The reason for this advice is a concern that some endocrine active substance (EAS) sensitive assays are being run at doses/concentrations of EASs that are too low to trigger direct impacts on the endocrine system. This guidance document is not the place to address this issue directly, but it should be considered when EAS-sensitive test guidelines (TGs) are revised in the future. In addition, the number and spacing of dose/concentration levels should also be adequate to fulfil the objectives of the study (e.g. to demonstrate dose response relationships if this is required).

Existing data to be considered

518. Existing information on endocrine-related effects from other vertebrates (up to and including mammals, e.g. from OECD TG 407) should always be considered, given the commonality of endocrine mechanisms in these taxa. Existing data available before deployment of EASZY might include *in vivo* results obtained with other vertebrates (e.g. a Uterotrophic Bioassay with rodents, positive findings for endocrine endpoints in mammalian repeat dose toxicity or reproductive studies), or one or more of a range of *in silico* or *in vitro* results which suggest that the modalities indicated above may occur *in vivo*. Such indicators of possible *in vivo* activity might include quantitative structure activity relationship (QSAR) predictions of endocrine activity, high throughput screening data, "read-across" from *in vivo* results obtained with structurally related chemicals, or positive results from an *in vitro* screen for estrogen or androgen receptor-mediated activity, or for effects on steroidogenesis (especially aromatase inhibition).

Scenarios: Positive and negative results combined with existing data

519. The scenarios (A to R) presented in Table C.2.16 represent all the possibilities of positive or negative results in combination with the presence or absence of existing data. The action taken will also depend on the regulatory environment, but the considerations given here are generally science based. Wherever possible, the recommended "next step which could be taken" avoids unnecessary animal testing. However, sometimes conducting an animal

test will be indicated and then the relevance of species, strain and exposure route should always be considered. Further considerations specific to each scenario are given in the table.

520. Positive results obtained with one or more of the endpoints (Table C.2.16, Scenarios A-I) result in the conclusion that the test chemical is a potential ED *in vivo*. When a positive response is observed in the EASZY Assay, confirmatory experiments can be conducted by co-exposing embryos to the test chemical and the estrogen receptor (ER) antagonist ICI 182 780. If a significant down-regulation of the GFP is then observed, the involvement of functional ERs is indicated. Positive responses would ideally need to be followed up with more comprehensive testing to show whether adverse apical effects related to endocrine impacts occur at any part of the life cycle (and hence to discover whether the chemical is an ED acting through certain estrogen/androgen/thyroid/steroidogenesis [E,A,T,S] pathways). In other words, a positive result in the EASZY Assay may trigger TG 234 (FSDT) at Level 4 or fish life cycle testing (e.g. MEOGRT – TG 240) at Level 5. Existing data suggesting endocrine activity will strengthen the case for additional testing.

521. The situation in which the EASZY Assay gives a negative result (Table C.2.16, Scenarios J-R) needs careful consideration of any existing data. If the weight of evidence of these data suggests that the chemical is endocrine active both *in vitro* and *in vivo* in other species (Scenario J), then the probability is that the EASZY may simply be insufficiently responsive in that case, or fish in general may be unresponsive. In some of these circumstances, it might be appropriate to conduct a FSDT (OECD TG 234), or alternatively, a fish life cycle test (e.g. MEOGRT, TG 240) to confirm that there is no endocrine activity in fish.

522. If the EASZY Assay and existing *in vivo* data are all negative, but *in vitro* data reveal some endocrine activity (Scenario K), the probability is that the test chemical is not sufficiently potent to produce endocrine effects *in vivo* in fish, or it may be rapidly metabolised. In such a situation, further testing may or may not be necessary. If the chemical is known to bioaccumulate slowly, it may be that exposures in the *in vivo* tests are not of sufficient duration, in which case longer term testing might be justified. If the *in vitro* data reveal anti-androgenic or thyroid activity, consideration should be given to conducting the Androgenised Female Stickleback Screen (AFSS – OECD GD 148) or Juvenile Medaka Anti-Androgen Screening Assay (JMASA), or the Amphibian Metamorphosis Assay (AMA – OECD TG 231), respectively.

523. On the other hand, if the EASZY Assay and the *in vitro* tests are negative, but there are positive existing *in vivo* data (Scenario M), the chemical is probably not a potential ED with the modalities listed above, but it may act via estrogen- or androgen-related MOA not covered by the *in vitro* screens, or it may be more potent in species or life stages that have not been tested. In this situation, the relevant existing *in vitro* and *in vivo* data should be used to guide decisions about whether to conduct any further testing, either for modalities such as anti-androgenicity or including life stages represented in OECD TG 234 (FSDT) or in TG 240 (MEOGRT).

524. Finally, a negative EASZY, set against a background of negative *in vitro* and *in vivo* data (Scenario N) that includes relevant *in vivo* data for fish, suggests that the test chemical is not a potential ED in fish or other vertebrates, and no further testing for estrogenic, anti-estrogenic, androgenic or steroidogenic MOA will generally be necessary. It remains possible that it has anti-androgenic or thyroid activity, although negative *in vitro* tests for these modalities would suggest that this scenario is unlikely.

525. In each of the above scenarios, it is possible that existing data will be equivocal, or there may be no existing data (Scenarios C, F-I, L and O-R). This will weaken the conclusions which can be drawn about a negative EASZY test, and this is reflected in Table C.2.16. However, a lack of mechanistic data on endocrine activity should usually be rectified before any further *in vivo* testing is finally decided on. Indeed, as a general principle, it is desirable to obtain mechanistic data before any *in vivo* testing. On the other hand, if EASZY is positive, further *in vivo* testing is generally indicated, particularly when existing data are equivocal, or if there are no existing data. There is also the possibility that equivocal mechanistic data may be the result of multiple modes of endocrine action. Under some circumstances, two opposite modes of simultaneous action (e.g. estrogenic and anti-estrogenic) could, depending on dose, lead to a minimisation or abolition of adverse effects, while in others two different MOA (e.g. estrogenic and androgenic) could potentially reinforce effects on EASZY. If multiple MOA are suspected, either from the existing results or based on QSAR/read-across/integrated approaches, this situation should be investigated further if needed for regulatory decision making.

526. The scenario in which the results of the EASZY Assay are themselves equivocal has not been dealt with in Table C.2.16, for reasons of brevity. In this context, an equivocal result might be an inconsistent concentration-response (e.g. no effect at a high concentration but effects at a lower concentration), or a result which borders on statistical significance. Without knowing the exact circumstances, reliable advice cannot be given, but the opinions of an experienced ecotoxicologist should be sought. Clearly, however, such equivocal results do not necessarily rule out the existence of *in vivo* endocrine activity. If false negatives (e.g. systemic toxicity) are suspected with good reason, the screen could be repeated if none of the test concentrations have given reliable data (e.g. conduct it at lower concentrations which avoid systemic toxicity). However, note that a repeat test in the event of systemic toxicity would not be needed providing at least one tested concentration was not subject to such effects.

527. In summary, positive results in the EASZY Assay indicate that a chemical is a possible endocrine disrupter. More predictive *in vivo* testing would then be necessary to produce a long-term no-observed-effect-concentration/x% effect concentration (NOEC/ECx) and/or to confirm whether or not the chemical is an actual endocrine disrupter with adverse effects *in vivo*. Negative results in EASZY do not necessarily mean that the chemical is not a potential ED – a judgement about its endocrine disruption potential and the possible need for additional testing will have to be made based on a weight of evidence evaluation of existing *in vitro* and *in vivo* data.

References

Brion, F. et al. (2012), "Screening estrogenic activities of chemicals or mixtures *in vivo* using transgenic (*cyp19a1b-GFP*) zebrafish embryos", *PLoS ONE*, Vol. 7/5, http://dx.doi.org/10.1371/journal.pone.0036069.

Le Fol, V. et al. (2017), "Life-stage dependant biotransformation of BP2 and BPS in zebrafish supports the use of embryo model as an alternative to adult fish", *International Journal of Molecular Science*, Vol. 18/4, http://dx.doi.org/10.3390/ijms18040704.

Weigt, S. et al. (2011), "Zebrafish (*Danio rerio*) embryos as a model for testing proteratogens", *Toxicology*, Vol. 281/1-3, pp. 25-36, https://doi.org/10.1016/j.tox.2011.01.004.

WHO/IPCS (2002), "Global assessment of the state-of-the-science of endocrine disrupters", Damstra, T. et al. (eds.) WHO/PCS/EDC/02.2, World Health Organization, Geneva, www.who.int/ipcs/publications/new_issues/endocrine_disruptors/en.

Table C.2.16. **EASZY Assay: Detection of substances acting through estrogen receptors using transgenic cyp19a1b GFP zebrafish embryos (draft OECD TG): Guidance for scenarios of combinations of results with existing data**

This table represents possible conclusions to be drawn from assay data, and a next step which could be taken if further evidence is required about possible endocrine disrupting properties and/or effects. The guidance offered is not meant to be prescriptive, but provides science-based considerations. It encourages the use of all available data and expert judgement in a weight of evidence approach. Regional and national interpretation of results and "next steps" may vary.

The conclusions are grouped into a series of scenarios (A-R), each scenario representing a different combination of assay results, existing *in vitro* data and existing *in vivo* data. The symbol "+" indicates that the data in question represent a positive result, «-" indicates a negative result, and "Eq/0" indicates that the data are either equivocal or are not available. Note that there are no apical endpoints in this assay considered to be diagnostic of an estrogen/androgen/thyroid/steroidogenesis (E,A,T,S) modality.

Existing results: * "Mechanism (*in vitro* mechanistic data)" assumes that mechanistic data are available from estrogen receptor (ER-), androgen receptor (AR-) and steroidogenesis-based assays (Level 2). Thyroid hormone receptor (TR) and other assays concerning mechanisms of thyroid disruption may be available, but they are not in common use. In practice, data from all assays may not be available and therefore this must be taken into account when deciding on the "next step". Quantitative structure activity relationship (QSAR) predictions of estrogen and androgen binding/activation may be made for some substances. There is little evidence at present that equivalent *in vitro* assays with systems derived from fish offer advantages over their mammalian counterparts, although some differences in response have been observed.

Existing results: ** "Effects (*in vivo* effects of concern)" assumes effects have been observed in other *in vivo* screens/tests which give rise to concern that the test chemical may be an endocrine disrupter.

Scenarios	Result of EASZY Assay	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
A	+	+	+	Strong evidence for *in vivo* endocrine activity in fish and other organisms.	Consider performing a fish life cycle test (e.g. MEOGRT – OECD TG 240), especially if the intention is to obtain precise data on a reproductive or developmental no-observed-effect-concentration/x% effect concentration (NOEC/ECx).	An alternative approach would be to deploy OECD TG 234 (Fish Sexual Development test [FSDT]), especially if sexual development is expected to give a response at lower concentrations than reproduction.
B	+	+	–	Moderate evidence for *in vivo* endocrine activity in fish, despite lack of *in vivo* effects in existing tests.	Consider performing a fish life cycle test (e.g. MEOGRT – OECD TG 240), especially if the intention is to obtain precise data on a reproductive or developmental NOEC/ECx.	An alternative approach would be to deploy OECD TG 234 (FSDT), especially if sexual development is expected to give a response at lower concentrations than reproduction. If the negative *in vivo* data are from a fish test (e.g. OECD TG 229), consider possible reasons for the disparity (e.g. differences in species sensitivity) before conducting a life cycle test.
C	+	+	Eq/0**	Moderate evidence for *in vivo* endocrine activity in fish despite equivocal or absent *in vivo* data in other species.	Consider performing a fish life cycle test (e.g. MEOGRT – OECD TG 240), especially if the intention is to obtain precise data on a reproductive or developmental NOEC/ECx.	If no existing fish data are available, it may be worth performing OECD TG 234 (FSDT) before a possible life cycle test in order to obtain information on whether sexual development is a sensitive part of the life cycle. Such information could influence the design of the life cycle test. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity, lack of sufficient transformation to endocrine-active products or multiple modes of action (MOA). If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
D	+	–	+	Strong evidence for *in vivo* endocrine activity in fish and other species, but confidence about MOA is reduced by negative mechanistic data.	Consider performing a fish life cycle test (e.g. MEOGRT – OECD TG 240), especially if the intention is to obtain precise data on a reproductive or developmental NOEC/ECx.	The negative *in vitro* data suggest that the test chemical may be metabolically activated *in vivo* (such activation is possible in the EASZY Assay), or it may operate via mechanisms not covered by the *in vitro* screens. An alternative approach to a life cycle test would be to deploy OECD TG 234 (FSDT).
E	+	–	–	Some evidence for *in vivo* endocrine activity in fish, but confidence is reduced by negative *in vitro* data and negative *in vivo* activity in other species.	Consider performing a fish life cycle test (e.g. MEOGRT – OECD TG 240), especially if the intention is to obtain precise data on a reproductive or developmental NOEC/ECx.	The negative *in vitro* data suggest that the test chemical may be metabolically activated *in vivo* (such activation is possible in the EASZY Assay), or it may operate via mechanisms not covered by the *in vitro* screens. An alternative approach to a life cycle test would be to deploy OECD TG 234 (FSDT), especially if sexual development is expected to give a response at lower concentrations than reproduction. If the negative *in vivo* data are from a fish test (e.g. OECD TG 229), consider possible reasons for the disparity (e.g. differences in species sensitivity) before conducting a life cycle test.

Scenarios	Result of EASZY Assay	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
F	+	–	Eq/0	Some evidence for *in vivo* endocrine activity in fish, but confidence is reduced by negative *in vitro* data and equivocal or absent *in vivo* activity in other species.	Consider performing a fish life cycle test (e.g. MEOGRT – OECD TG 240), especially if the intention is to obtain precise data on a reproductive or developmental NOEC/ECx.	The negative *in vitro* data suggest that the test chemical may be metabolically activated *in vivo* (such activation is possible in the EASZY Assay), or it may operate via mechanisms not covered by the *in vitro* screens, or may not be metabolically activated *in vitro*. If no existing fish data are available, it may be worth performing OECD TG 234 (FSDT) before a possible life cycle test in order to obtain information on whether sexual development is a sensitive part of the life cycle. Such information could influence the design of the life cycle test. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
G	+	Eq/0	+	Strong evidence for *in vivo* endocrine activity in fish, but mechanism unconfirmed.	Obtain mechanistic data, then consider performing a fish life cycle test (e.g. MEOGRT – OECD TG 240).	An alternative approach to a life cycle test would be to deploy OECD TG 234 (FSDT). It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
H	+	Eq/0	–	Some evidence for *in vivo* endocrine activity in fish, but mechanism unconfirmed.	Obtain mechanistic data, then consider performing a fish life cycle test (e.g. MEOGRT – OECD TG 240).	An alternative approach to a life cycle test would be to deploy OECD TG 234 (FSDT), especially if sexual development is expected to give a response at lower concentrations than reproduction. If the negative *in vivo* data are from a fish test (e.g. OECD TG 229), consider possible reasons for the disparity (e.g. differences in species sensitivity) before possibly conducting a life cycle test. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
I	+	Eq/0	Eq/0	Moderate evidence for *in vivo* endocrine activity in fish, but mechanism unconfirmed.	Obtain mechanistic data, then consider performing a fish life cycle test (e.g. MEOGRT – OECD TG 240).	If no existing fish data are available, it may be worth performing OECD TG 234 (FSDT) before a possible life cycle test in order to obtain information on whether sexual development is a sensitive part of the life cycle. Such information could influence the design of the life cycle test. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

Scenarios	Result of EASZY Assay	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
J	–	+	+	Based on the existing data, the chemical has endocrine activity *in vivo*. The lack of response in the EASZY Assay suggests that fish are not responsive, unless the existing data are from fish.	Consider performing OECD TG 234 (FSDT).	It is possible that the failure to give a positive result in the EASZY Assay was caused by the short exposure time (96 hours). If this is suspected (e.g. if the chorion has impeded uptake and/or the chemical only bioaccumulates slowly), or if the existing *in vivo* data are from a fish, OECD TG 234 (FSDT) or potentially a life cycle test (e.g. TG 240 – MEOGRT) would be able to study the effects of longer exposure and confirm whether there is a hazard to fish. Choice of test should be guided by the existing *in vivo* data.
K	–	+	–	There is no evidence that the chemical is a potential endocrine disruptor (ED) *in vivo*, probably because it is very weakly acting or rapidly metabolised.	Probably no further action, but see comments in right-hand column.	It is known that uptake of some chemicals can be impeded by the chorion, and it is possible that EDs which bioaccumulate slowly may only cause effects *in vivo* after exposure times longer than 96 hours. If this is suspected, and depending on which part of the life cycle is suspected of being the most sensitive, consider performing OECD TG 229/230 (FSTRA or 21-Day Fish Assay) or TG 234 (FSDT). It is also possible that the chemical may be an anti-androgen *in vivo* (consider performing the Androgenised Female Stickleback Screen [AFSS] or Juvenile Medaka Anti-Androgen Screening Assay [JMASA]), or a thyroid-active chemical *in vivo* (consider performing the Amphibian Metamorphosis Assay [AMA] – OECD TG 231).
L	–	+	Eq/0	The chemical may not be an ED *in vivo*, but the confidence in this conclusion is relatively low as there is only one unequivocal *in vivo* test result (a negative).	Probably no further action, but see comments in right-hand column.	If the equivocal existing data are from a fish assay, consider performing a fish assay (e.g. OECD TG 229 or TG 230) with a different species, or a longer term test (e.g. TG 234 [FSDT] or life cycle test [MEOGRT TG 240]) if the chemical is a slow bioaccumulator. It is also possible that the chemical may be an anti-androgen *in vivo* (consider performing an AFSS), or a thyroid-active chemical *in vivo* (consider performing an AMA– OECD TG 231). It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
M	–	–	+	The chemical is apparently not a potential ED in fish but it does have activity in another species.	Use the existing *in vivo* data to help decide whether a longer term test with an appropriate fish species is needed.	It is also possible that the chemical may be an anti-androgen *in vivo* (consider performing an AFSS), or a thyroid-active chemical *in vivo* (consider performing an AMA – OECD TG 231), although lack of *in vitro* binding affinity with the androgen receptor suggests the former is unlikely. Use the existing *in vivo* data to guide any further testing.
N	–	–	–	The chemical is probably not a potential ED *in vivo*.	No further action with respect to estrogenic, anti-estrogenic, androgenic or steroidogenic MOA.	It is still possible that the chemical may be an anti-androgen *in vivo* (consider performing an AFSS), or a thyroid-active chemical *in vivo* (consider performing an AMA – OECD TG 231), although lack of *in vitro* binding affinity with the androgen receptor suggests the former is unlikely.

Scenarios	Result of EASZY Assay	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
O	–	–	Eq/0	The chemical is probably not a potential ED in fish.	Probably no further action. However, see comments in right-hand column.	If the paucity of *in vivo* data is a concern, performance of a screening test (OECD TG 229 or TG 230) with a different species, or a longer term test (i.e. FSDT – TG 234 or life cycle test [MEOGRT]) could be considered. It is also possible that the chemical may be an anti-androgen *in vivo* (consider performing an AFSS), or a thyroid-active chemical *in vivo* (consider performing an AMA – OECD TG 231), although lack of *in vitro* binding affinity with the androgen receptor suggests the former is unlikely. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
P	–	Eq/0	+	The chemical is probably not a potential ED in fish, but confidence in this conclusion is low given the lack of mechanistic *in vitro* data and the availability of positive existing *in vivo* data.	Obtain mechanistic data, then consider whether further testing is desirable.	If the mechanistic data confirm that the chemical has potential endocrine action, consider conducting a fish assay (OECD TG 229 or TG 230) with another species, or a longer term test (TG 234 – FSDT or life cycle test [MEOGRT]). Use the existing *in vivo* data as a guide to test choice. If the mechanistic data reveal anti-androgenic or thyroid activity, consider performing an AFSS (OECD GD 148)/JMASA, or AMA (OECD TG 231), respectively. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
Q	–	Eq/0	–	The chemical is probably not a potential ED in fish, but the lack of mechanistic *in vitro* data are a concern, even though the existing *in vivo* data are negative.	Obtain mechanistic data, then consider whether further testing is desirable.	If the mechanistic data confirm that the chemical has potential endocrine action, consider conducting a fish assay (OECD TG 229 or TG 230) with another species, or a longer term test (TG 234 – FSDT or life cycle test [MEOGRT]). Use the existing *in vivo* data as a guide to test choice. If the mechanistic data reveal anti-androgenic or thyroid activity, consider performing an AFSS (OECD GD 148)/JMASA, or AMA (OECD TG 231), respectively. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

Scenarios	Result of EASZY Assay	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
R	-	Eq/0	Eq/0	The chemical is probably not a potential ED in fish, but confidence in this conclusion is low given the lack of mechanistic *in vitro* and existing *in vivo* data.	Obtain mechanistic data, then consider whether further testing is desirable.	If the mechanistic data confirm that the chemical has potential endocrine action, consider conducting a fish assay (OECD TG 229 or TG 230) with another species, or a longer term test (TG 234 [FSDT] or life cycle test [MEOGRT]). Use the existing *in vivo* data as a guide to test choice. If the mechanistic data reveal anti-androgenic or thyroid activity, consider performing an AFSS (OECD GD 148)/JMASA, or AMA (OECD TG 231), respectively. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

C.2.17. JMASA: Juvenile Medaka Anti-Androgen Screening Assay (draft OECD GD)

Status: Assay being validated by the OECD.

528. Modality detected/endpoints: This draft *in vivo* medaka assay is sensitive to androgen antagonists and to chemicals interfering with androgen biosynthesis. In principle, it can also be used to identify estrogen agonists and antagonists, as well as aromatase inhibitors. However, this guidance will restrict itself to the detection of anti-androgenicity alone, as data on responses to the other modalities have not yet been published.

Background to the assay

529. This assay is currently starting validation by the OECD for possible approval as a guidance document (GD). No validation data have yet been produced, but some developmental data are available. It is planned to have a GD ready by 2019 at the earliest. The assay is based on juvenile medaka (*Oryzias latipes*), the males of which develop secondary sexual characteristics known as papillary processes (PP) on the anal fin between 42 and 49 days post-fertilisation (dpf). The PP grow under androgenic control, and anti-androgens or chemicals which interfere with androgen biosynthesis can prevent their appearance or limit their number. Juvenile medaka (both sexes at 42 dpf) are exposed to test chemical for 28 days, to 70 dpf, after which their genotypic sex is determined using the *dmy* gene. This enables the males alone to be evaluated for the presence, reduction or absence of PP. It is optionally possible to measure vitellogenin, so the assay can in principle also be used to detect estrogen agonists and antagonists, and aromatase inhibitors. However, this option will only be considered further in future editions of this document when supporting data have been evaluated.

When/why the assay may be used

530. Although data from the JMASA could, in principle, be available at any stage in the hazard assessment process, the most likely scenario will be when there are relatively few data available about the possible endocrine disrupting properties of a chemical. The assay is most likely to be used either as part of a battery of *in vitro* and *in vivo* screens, or to follow up on existing data which suggest possible endocrine disruption activity. Given the high degree of endocrine system conservation across the vertebrates, adverse endocrine-linked effects in the JMASA assay may also indicate the possibility of related activity in other organisms such as amphibians, reptiles, birds or mammals. It is also possible that no existing endocrine-relevant data are available (i.e. the JMASA has been used as a primary screen), but in that case a positive result in the screen should ideally be followed up with relevant *in vitro* screening in an attempt to confirm the precise mode of action (MOA). Furthermore, a positive JMASA result would also need to be followed up with an additional *in vivo* fish test such as the Fish Short-Term Reproduction Assay (FSTRA – OECD TG 229) or Fish Sexual Development Test (FSDT – OECD TG 234), which will give some

indication of any adverse apical effects. Possible conclusions to be derived from the results of the JMASA, and guidance about potential additional studies to strengthen weight of evidence, are summarised in Table C.2.17.

531. In order to provide information relevant for assessing whether or not a chemical may fulfil the WHO/IPCS (2002) definition of an endocrine disruptor (ED), the study design has to be sufficiently robust to demonstrate the presence or absence of effects. In the dose selection, the investigator should also consider and ensure that data generated are adequate to fulfil the regulatory requirement across OECD countries as appropriate (e.g. hazard and risk assessment and labelling, ED assessment, etc.). The top dose or concentration should be sufficiently high to give clear systemic (i.e. non endocrine-specific) toxicity in order to ensure that a wide range of exposures (high to low) is tested. However, endocrine effects observed solely in the presence of clear systemic toxicity should be interpreted with caution and may be disregarded when sufficiently justified to be caused by secondary effects which are unlikely to be due to endocrine activity. The reason for this advice is a concern that some endocrine active substance (EAS) sensitive assays are being run at doses/concentrations of EASs that are too low to trigger direct impacts on the endocrine system. This GD is not the place to address this issue directly, but it should be considered when EAS-sensitive test guidelines (TGs) are revised in the future. In addition, the number and spacing of dose/concentration levels should also be adequate to fulfil the objectives of the study (e.g. to demonstrate dose response relationships if this is required).

Existing data to be considered

532. Existing information on endocrine-related effects from other vertebrates (up to and including mammals, e.g. positive findings for endocrine endpoints in mammalian repeat dose toxicity or reproductive studies) should always be considered, given the commonality of endocrine mechanisms in these taxa. Existing data available before deployment of the JMASA might include *in vivo* results obtained with other vertebrates (e.g. a Hershberger Bioassay in Rats – OECD TG 441), or one or more of a range of *in silico* or *in vitro* results which suggest that anti-androgenicity may occur *in vivo*. Such indicators of possible *in vivo* activity might include quantitative structure activity relationship (QSAR) predictions of endocrine activity, high throughput screening data, "read-across" from *in vivo* results obtained with structurally related chemicals or positive results from an *in vitro* screen for androgen receptor-mediated activity, or for effects on androgen biosynthesis.

533. It should be noted that a sensitive *in vivo* assay for anti-androgenicity is already available: the Androgenised Female Stickleback Screen (AFSS – OECD GD 148). This is slightly shorter than the JMASA (21 days), but relies on the pre-treatment of female sticklebacks (*Gasterosteus aculeatus*) with an androgen before measuring anti-androgenic effects of the test chemical (reduction in induced spiggin glue protein).

Scenarios: Positive and negative results combined with existing data

534. The scenarios (A to R) presented in Table C.2.17 represent all the possibilities of positive or negative results in combination with the presence or absence of existing data. The action taken will also depend on the regulatory environment, but the considerations given here are generally science based. Wherever possible, the recommended "next step which could be taken" avoids unnecessary animal testing. However, sometimes conducting an animal test will be indicated and then the relevance of species, strain and exposure route should

always be considered. Further considerations specific to each scenario are given in the table.

535. Positive results obtained with the secondary sexual characteristics endpoint (Table C.2.17, Scenarios A-I) result in the conclusion that the test chemical is a possible anti-androgen *in vivo*. This would ideally need to be followed up with more comprehensive testing to show whether adverse apical effects related to endocrine impacts occur at any part of the life cycle (and hence to discover whether the chemical is an ED acting through certain estrogen/androgen/thyroid/steroidogenesis [E,A,T,S] pathways). In other words, a positive result in the JMASA may trigger OECD TG 234 (FSDT) at Level 4 or fish life cycle testing (e.g. MEOGRT – TG 240) at Level 5. Existing data suggesting anti-androgenic activity will strengthen the case for additional testing.

536. The situation in which the JMASA gives a negative result (Table C.2.17, Scenarios J-R) needs careful consideration of any existing data. If the weight of evidence of these data suggests that the chemical is endocrine active both *in vitro* and *in vivo* in other species (Scenario J), then the probability is that the JMASA may simply be insufficiently responsive in that case, or fish in general may be unresponsive. In some of these circumstances, it might be appropriate to conduct an FSDT (OECD TG 234), or alternatively, a fish life cycle test (e.g. MEOGRT, TG 240) to confirm that there is no endocrine activity in fish.

537. If the JMASA and existing *in vivo* data are all negative, but *in vitro* data reveal some endocrine activity (Scenario K), the probability is that the test chemical is not sufficiently potent to produce endocrine effects *in vivo* in fish, or it may be rapidly metabolised. In such a situation, further testing may or may not be necessary. If the chemical is known to bioaccumulate slowly, it may be that exposures in the *in vivo* tests are not of sufficient duration, in which case longer term testing might be justified.

538. On the other hand, if the JMASA and the *in vitro* tests are negative, but there are positive existing *in vivo* data (Scenario M), the chemical is probably not a potential ED with anti-androgenic activity, but it may act via estrogen- or androgen-related MOA not covered by the *in vitro* screens, or it may be more potent in species or life stages that have not been tested. In this situation, the relevant existing *in vitro* and *in vivo* data should be used to guide decisions about whether to conduct any further testing, including life stages represented in OECD TG 234 (FSDT) or in TG 240 (MEOGRT).

539. Finally, a negative JMASA, set against a background of negative *in vitro* and *in vivo* data (Scenario N) that includes relevant *in vivo* data for fish, suggests that the test chemical is not a potential ED in fish or other vertebrates, and no further testing for anti-androgenic or anti-steroidogenic MOA will generally be necessary.

540. In each of the above scenarios, it is possible that existing data will be equivocal, or there may be no existing data (Scenarios C, F-I, L and O-R). This will weaken the conclusions which can be drawn about a negative JMASA, and this is reflected in Table C.2.17. However, a lack of mechanistic data on endocrine activity should usually be rectified before any further *in vivo* testing is finally decided on. Indeed, as a general principle, it is desirable to obtain mechanistic data before any *in vivo* testing. On the other hand, if the JMASA is positive, further *in vivo* testing is generally indicated, particularly when existing data are equivocal, or if there are no existing data. There is also the possibility that equivocal mechanistic data may be the result of multiple modes of endocrine action. Under some circumstances, two opposite modes of simultaneous action (e.g. androgenic and anti-androgenic) could, depending on dose, lead to a minimisation or

abolition of adverse effects, while in others two different MOA (e.g. anti-steroidogenic and androgenic) could potentially reinforce effects on the JMASA. If multiple MOA are suspected, either from the existing results or based on QSAR/read-across/integrated approaches, this situation should be investigated further if needed for regulatory decision making.

541. The scenario in which the results of the JMASA are themselves equivocal has not been dealt with in Table C.2.17, for reasons of brevity. In this context, an equivocal result might be an inconsistent concentration-response (e.g. no effect at a high concentration but effects at a lower concentration), or a result which borders on statistical significance. Without knowing the exact circumstances, reliable advice cannot be given, but the opinions of an experienced ecotoxicologist should be sought. Clearly, however, such equivocal results do not necessarily rule out the existence of *in vivo* endocrine activity. If false negatives (e.g. systemic toxicity) are suspected with good reason, the screen could be repeated if none of the test concentrations have given reliable data (e.g. conduct it at lower concentrations which avoid systemic toxicity). However, note that a repeat test in the event of systemic toxicity would not be needed providing at least one tested concentration was not subject to such effects.

542. In summary, positive results in the JMASA indicate that a chemical is a possible endocrine disrupter. More predictive *in vivo* testing would then be necessary to produce a long-term no-observed-effect-concentration/x% effect concentration (NOEC/ECx) and/or to confirm whether or not the chemical is an actual endocrine disrupter with adverse effects *in vivo*. Negative results in the JMASA do not necessarily mean that the chemical is not a potential ED – a judgement about its endocrine disruption potential and the possible need for additional testing will have to be made based on a weight of evidence evaluation of existing *in vitro* and *in vivo* data.

Reference

WHO/IPCS (2002), "Global assessment of the state-of-the-science of endocrine disrupters", Damstra, T. et al. (eds.) WHO/PCS/EDC/02.2, World Health Organization, Geneva, www.who.int/ipcs/publications/new_issues/endocrine_disruptors/en.

Table C.2.17. **JMASA: Juvenile Medaka Anti-Androgen Screening Assay (draft OECD GD): Guidance for scenarios of combinations of results with existing data**

This table represents possible conclusions to be drawn from assay data, and a next step which could be taken if further evidence is required about possible endocrine disrupting properties and/or effects. The guidance offered is not meant to be prescriptive, but provides science-based considerations. It encourages the use of all available data and expert judgement in a weight of evidence approach. Regional and national interpretation of results and "next steps" may vary.

The conclusions are grouped into a series of scenarios (A-R), each scenario representing a different combination of assay results, existing *in vitro* data and existing *in vivo* data. The symbol "+" indicates that the data in question represent a positive result, «-" indicates a negative result, and "Eq/0" indicates that the data are either equivocal or are not available. Note that there are no apical endpoints in this assay considered to be diagnostic of an estrogen/androgen/thyroid/steroidogenesis (E,A,T,S) modality.

Existing results: * "Mechanism (*in vitro* mechanistic data)" assumes that mechanistic data are available from estrogen receptor (ER-), androgen receptor (AR-) and steroidogenesis-based assays (Level 2). Thyroid hormone receptor (TR) and other assays concerning mechanisms of thyroid disruption may be available, but they are not in common use. In practice, data from all assays may not be available and therefore this must be taken into account when deciding on the "next step". Quantitative structure activity relationship (QSAR) predictions of estrogen and androgen binding/activation may be made for some substances. There is no evidence at present that equivalent *in vitro* assays with systems derived from fish offer advantages over their mammalian counterparts.

Existing results: ** "Effects (*in vivo* effects of concern)" assumes effects have been observed in other *in vivo* screens/tests which give rise to concern that the test chemical may be an endocrine disrupter.

Scenarios	Result of JMASA	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
A	+	+	+	Strong evidence for *in vivo* anti-androgenic activity in fish and other organisms.	Consider performing a fish life cycle test (e.g. MEOGRT – OECD TG 240), especially if the intention is to obtain precise data on a reproductive or developmental no-observed-effect-concentration/x% effect concentration (NOEC/ECx).	An alternative approach would be to deploy OECD TG 234 (Fish Sexual Development Test [FSDT]), especially if sexual development is expected to give a response at lower concentrations than reproduction.
B	+	+	–	Strong evidence for *in vivo* anti-androgenic activity in fish, despite lack of *in vivo* effects in existing tests.	Consider performing a fish life cycle test (e.g. MEOGRT – OECD TG 240), especially if the intention is to obtain precise data on a reproductive or developmental NOEC/ECx.	An alternative approach would be to deploy OECD TG 234 (FSDT), especially if sexual development is expected to give a response at lower concentrations than reproduction. If the negative *in vivo* data are from a fish test (e.g. OECD TG 229), consider possible reasons for the disparity (e.g. differences in species sensitivity) before conducting a life cycle test.
C	+	+	Eq/0**	Strong evidence for *in vivo* anti-androgenic activity in fish despite equivocal or absent *in vivo* data in other species.	Consider performing a fish life cycle test (e.g. MEOGRT – OECD TG 240), especially if the intention is to obtain precise data on a reproductive or developmental NOEC/ECx.	If no existing fish data are available, it may be worth performing OECD TG 234 (FSDT) before a possible life cycle test in order to obtain information on whether sexual development is a sensitive part of the life cycle. Such information could influence the design of the life cycle test. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple modes of action (MOA). If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
D	+	–	+	Strong evidence for *in vivo* anti-androgenic activity in fish and other species, but confidence about MOA is reduced by negative mechanistic data.	Consider performing a fish life cycle test (e.g. MEOGRT – OECD TG 240), especially if the intention is to obtain precise data on a reproductive or developmental NOEC/ECx.	The negative *in vitro* data suggest that the test chemical may be metabolically activated *in vivo*, or it may operate via mechanisms not covered by the *in vitro* screens. An alternative approach to a life cycle test would be to deploy OECD TG 234 (FSDT), especially if sexual development is expected to give a response at lower concentrations than reproduction.
E	+	–	–	Moderate-strong evidence for *in vivo* anti-androgenic activity in fish, but confidence is reduced by negative *in vitro* data and negative *in vivo* activity in other species.	Consider performing a fish life cycle test (e.g. MEOGRT – OECD TG 240), especially if the intention is to obtain precise data on a reproductive or developmental NOEC/ECx.	The negative *in vitro* data suggest that the test chemical may be metabolically activated *in vivo*, or it may operate via mechanisms not covered by the *in vitro* screens. An alternative approach to a life cycle test would be to deploy OECD TG 234 (FSDT), especially if sexual development is expected to give a response at lower concentrations than reproduction. If the negative *in vivo* data are from a fish test (e.g. OECD TG 229), consider possible reasons for the disparity (e.g. differences in species sensitivity) before conducting a life cycle test.

Scenarios	Result of JMASA	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
F	+	–	Eq/0	Moderate-strong evidence for *in vivo* anti-androgenic activity in fish, but confidence is reduced by negative *in vitro* data and equivocal or absent *in vivo* activity in other species.	Consider performing a fish life cycle test (e.g. MEOGRT – OECD TG 240), especially if the intention is to obtain precise data on a reproductive or developmental NOEC/ECx.	The negative *in vitro* data suggest that the test chemical may be metabolically activated *in vivo*, or it may operate via mechanisms not covered by the *in vitro* screens. If no existing fish data are available, it may be worth performing OECD TG 234 (FSDT) before a possible life cycle test in order to obtain information on whether sexual development is a sensitive part of the life cycle. Such information could influence the design of the life cycle test. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
G	+	Eq/0	+	Strong evidence for *in vivo* anti-androgenic activity in fish, but mechanism unconfirmed.	Obtain mechanistic data, then consider performing a fish life cycle test (e.g. MEOGRT – OECD TG 240).	An alternative approach to a life cycle test would be to deploy OECD TG 234 (FSDT), especially if sexual development is expected to give a response at lower concentrations than reproduction. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
H	+	Eq/0	–	Strong-moderate evidence for *in vivo* anti-androgenic activity in fish, but mechanism unconfirmed.	Obtain mechanistic data, then consider performing a fish life cycle test (e.g. MEOGRT – OECD TG 240).	An alternative approach to a life cycle test would be to deploy OECD TG 234 (FSDT), especially if sexual development is expected to give a response at lower concentrations than reproduction. If the negative *in vivo* data are from a fish test (e.g. OECD TG 229), consider possible reasons for the disparity (e.g. differences in species sensitivity) before possibly conducting a life cycle test. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
I	+	Eq/0	Eq/0	Moderate evidence for *in vivo* endocrine activity in fish, but mechanism unconfirmed.	Obtain mechanistic data, then consider performing a fish life cycle test (e.g. MEOGRT – OECD TG 240).	If no existing fish data are available, it may be worth performing OECD TG 234 (FSDT) before a possible life cycle test in order to obtain information on whether sexual development is a sensitive part of the life cycle. Such information could influence the design of the life cycle test. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

Scenarios	Result of JMASA	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
J	–	+	+	Based on the existing data, the chemical has anti-androgenic activity *in vivo*. The lack of response in the JMASA suggests that fish are not responsive, unless the existing data are from fish.	Consider performing OECD TG 234 (FSDT).	It is possible that the failure to give a positive result in the JMASA was caused by the relatively short exposure time (28 days). If this is suspected (e.g. the chemical only bioaccumulates slowly), or if the existing *in vivo* data are from a fish, OECD TG 234 (FSDT) or potentially a life cycle test (e.g. TG 240 – MEOGRT) would be able to study the effects of longer exposure and confirm whether there is a hazard to fish. Choice of test should be guided by the existing *in vivo* data.
K	–	+	–	There is no evidence that the chemical is a possible anti-androgenic endocrine disruptor (ED) *in vivo*, probably because it is very weakly acting or rapidly metabolised.	Probably no further action, but see comments in right-hand column.	It is possible that EDs which bioaccumulate slowly may only cause effects *in vivo* after exposure times longer than 28 days. If this is suspected, and depending on which part of the life cycle is suspected of being the most sensitive, consider performing OECD TG 234 (FSDT).
L	–	+	Eq/0	The chemical may not be an anti-androgenic ED *in vivo*, but the confidence in this conclusion is relatively low as there is only one unequivocal *in vivo* test result (a negative).	Probably no further action, but see comments in right-hand column.	If the equivocal existing data are from a fish assay, consider performing a fish assay (e.g. OECD TG 229 or TG 230) with a different species, or a longer term test (e.g. TG 234 – FSDT or life cycle test [MEOGRT]) if the chemical is a slow bioaccumulator. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
M	–	–	+	The chemical is apparently not a possible anti-androgenic ED in fish but it does have activity in another species.	Use the existing *in vivo* data to help decide whether a longer term test with an appropriate fish species is indicated.	
N	–	–	–	The chemical is probably not a possible anti-androgenic ED *in vivo*.	No further action with respect to anti-androgenic or anti-steroidogenic MOA.	
O	–	–	Eq/0	The chemical is probably not a possible anti-androgenic ED in fish.	Probably no further action. However, see comments in right-hand column.	If the paucity of *in vivo* data is a concern, performance of a screening test (OECD TG 229 or TG 230) with a different species, or a longer term test (i.e. TG 234 – FSDT or life cycle test [MEOGRT]) could be considered. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
P	–	Eq/0	+	The chemical is probably not a possible anti-androgenic ED in fish, but confidence in this conclusion is low given the lack of mechanistic *in vitro* data and the availability of positive existing *in vivo* data.	Obtain mechanistic data, then consider whether further testing is desirable.	If the mechanistic data confirm that the chemical has potential anti-androgenic action, consider conducting a fish assay (OECD TG 229 or TG 230) with another species, or a longer term test (TG 234 – FSDT or life cycle test [MEOGRT]). Use the existing *in vivo* data as a guide to test choice. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

Scenarios	Result of JMASA	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
Q	–	Eq/0	–	The chemical is probably not a possible anti-androgenic ED in fish, but the lack of mechanistic *in vitro* data are a concern, even though the existing *in vivo* data are negative.	Obtain mechanistic data, then consider whether further testing is desirable.	If the mechanistic data confirm that the chemical has potential endocrine action, consider conducting a fish assay (OECD TG 229 or TG 230) with another species, or a longer term test (TG 234 – FSDT or life cycle test [MEOGRT]). Use the existing *in vivo* data as a guide to test choice. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
R	–	Eq/0	Eq/0	The chemical is probably not a possible anti-androgenic ED in fish, but confidence in this conclusion is low given the lack of mechanistic *in vitro* and existing *in vivo* data.	Obtain mechanistic data, then consider whether further testing is desirable.	If the mechanistic data confirm that the chemical has potential anti-androgenic action, consider conducting a fish assay (OECD TG 229 or TG 230) with another species, or a longer term test (TG 234 – FSDT or life cycle test [MEOGRT]). Use the existing *in vivo* data as a guide to test choice. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

C.2.18. *Xenopus* Embryonic Thyroid Signalling Assay (XETA) (draft OECD TG)

Status: Assay being validated by the OECD.

543. Modality detected/endpoints: This *in vivo* assay with transgenic THbZIP *Xenopus laevis* tadpoles is responsive to pro-thyroid chemicals and anti-thyroid chemicals, but not to estrogens or (probably) other steroid hormone (ant)agonists.

Background to the assay

544. This *in vivo* assay is currently being validated by the OECD, and may be approved as an OECD test guideline (TG) in due course. It is designed as a screen for thyroid activity in amphibians, not to provide information on endocrine activity for use in assessing the environmental risks of an individual chemical based on a predicted environmental concentration/predicted no-effect concentration (PEC/PNEC) approach. It is important to note that there are several types of thyroid disruption, not all of which involve direct interactions with the thyroid receptor. Although it is to be expected that the assay will be responsive to all chemicals that interact with thyroid hormone (TH) receptors, or that lead to either an increase or a decrease in TH levels (thyroxine – T4; or the active form triiodothyronine – T3), it cannot be used to provide an unequivocal identification of the precise mode of action (MOA) of a chemical.

545. The XETA uses transgenic *Xenopus laevis* tadpoles into which a THbZIP promoter has been inserted that is regulated by TH and other TH agonists (Morvan-Dubois, Demeneix and Sachs, 2008). The promoter is linked to the gene coding for green fluorescent protein, and thus the degree of receptor response can be measured by fluorescence in the transparent tadpoles. Transgenic tadpoles are produced by mating wild-type females with THbZIP males, and these tadpoles are then exposed to dilutions of the test chemical from six-day post-fertilisation (N&F stage 45) for three days (to stage 47), after which their degree of fluorescence is recorded. The XETA has to be run both with and without a T3 internal spike because *Xenopus* tadpoles produce a very low level of TH at these developmental stages. Without a T3 internal spike, the XETA will respond only to pro-thyroid chemicals with an increase in fluorescence. If run with a T3 spike, it will respond to anti-thyroid chemicals and to pro-thyroid chemicals. The latter will be detected with a higher sensitivity in unspiked mode. Anti-thyroid chemicals will produce either an increase or a decrease in fluorescence, depending on the MOA. The tadpoles are metabolically competent, so the XETA is expected to be responsive to active metabolites.

When/why the assay may be used

546. Although the XETA could, in principle, be used at any stage in the hazard assessment process, the most likely use scenario will be when there are relatively few data available about the possible thyroid disrupting properties of a chemical. The results from this assay are most likely to be available after deployment of a battery of *in vitro* and *in vivo* screens,

or as a supplement to existing data which suggest potential endocrine disruptor (ED) activity. Given the high degree of endocrine system conservation across the vertebrates, endocrine-linked effects in the XETA may also indicate the possibility of related activity in other organisms such as fish, reptiles, birds or mammals. A number of mammalian (rat) assays are sensitive to thyroid disruption, particularly thyroid antagonists, including the pubertal assay (male or female), the enhanced repeat dose assay (OECD TG 407), and the intact male screening assay. Note that these assays utilise different routes of exposure than the XETA and therefore, depending on the properties of the chemical, have differing potentials for the test substance to be metabolised. It should also be noted that, at present, the only validated screening assay for thyroid-active chemicals is the Amphibian Metamorphosis Assay (AMA – OECD TG 231). The AMA uses more tadpoles than the XETA and is more time-consuming (21 days compared with 3), so the XETA may be more appropriate for rapid screening of large numbers of chemicals. On the other hand, the AMA measures several apical endpoints including speed of development and metamorphosis, whereas conclusions about adverse responses cannot be drawn from the XETA.

547. It is possible that no endocrine-relevant data are available before the XETA is deployed (i.e. if the XETA has been used as a primary screen), but in that case a positive result in the screen should probably be followed up with relevant *in vitro* screening, if available, to investigate the suspected MOA in more detail. It should be noted that *in vitro* screens exist for thyroid agonists and antagonists (e.g. GH_3 rat pituitary somatotroph cell proliferation; solid state thyroid receptor binding assays; transfected reporter gene assays in yeast or mammalian cell lines), but also for thyroid disruption occurring at other points in the endocrine system (e.g. porcine thyroperoxidase assay, TBG/TTR binding assays, FRTL-5 rat cell lines sensitive to iodide uptake inhibitors (see Table A.1). However, most of these screens are still at the research stage and none have yet been validated and standardised at the international level.

548. In order to provide information relevant for assessing whether or not a chemical may fulfil the WHO/IPCS (2002) definition of an ED, the study design has to be sufficiently robust to demonstrate the presence or absence of effects. In the dose selection, the investigator should also consider and ensure that data generated are adequate to fulfil the regulatory requirement across OECD countries as appropriate (e.g. hazard and risk assessment and labelling, ED assessment, etc.). The top dose or concentration should be sufficiently high to give clear systemic (i.e. non endocrine-specific) toxicity in order to ensure that a wide range of exposures (high to low) is tested. However, endocrine effects observed solely in the presence of clear systemic toxicity should be interpreted with caution and may be disregarded when sufficiently justified to be caused by secondary effects which are unlikely to be due to endocrine activity. The reason for this advice is a concern that some endocrine active substance (EAS) sensitive assays are being run at doses/concentrations of EASs that are too low to trigger direct impacts on the endocrine system. This guidance document is not the place to address this issue directly, but it should be considered when EAS-sensitive test guidelines (TGs) are revised in the future. In addition, the number and spacing of dose/concentration levels should also be adequate to fulfil the objectives of the study (e.g. to demonstrate dose response relationships if this is required).

Existing data to be considered

549. Existing information on endocrine-related effects from other vertebrates (up to and including mammals, e.g. positive findings for endocrine endpoints in mammalian repeat dose toxicity or reproductive studies) should also be considered before deployment of the

XETA, given the commonality of endocrine mechanisms in these taxa. Existing data available might also include one or more of a range of *in silico* or *in vitro* results which suggest that thyroid disruption may occur *in vivo* (but note the limitations of this approach, as indicated above). Such indicators of possible thyroid activity might include quantitative structure activity relationship (QSAR) predictions of thyroid activity, "read-across" from *in vivo* results obtained with structurally related chemicals or positive results from an *in vitro* screen for thyroid agonist/antagonist activity.

Scenarios: Positive and negative results combined with existing data

550. The scenarios (A to R) presented in Table C.2.18 represent all the possibilities of positive or negative results in combination with the presence or absence of existing data. The action taken will also depend on the regulatory environment, but the considerations given here are generally science based. Wherever possible, the recommended "next step which could be taken" avoids unnecessary animal testing. However, sometimes conducting an animal test will be indicated and then the relevance of species, strain and exposure route should always be considered. Further considerations, specific to each scenario are given in the table.

551. Positive results obtained with the XETA (Table C.2.18, Scenarios A-I) result in the conclusion that the test chemical is a possible thyroid disrupter *in vivo*, at least in amphibians. However, as indicated above, although a positive response of the XETA indicates that the chemical is a possible thyroid disrupter, a result of this type would generally need to be followed up with a more comprehensive screen. The most appropriate choice for this is the Amphibian Metamorphosis Assay (AMA – OECD TG 231). However, if countries need further evidence concerning growth and sexual development, a Larval Amphibian Growth and Development Assay (LAGDA – OECD TG 241) would be able to provide a precise no-observed-effect-concentration/x% effect concentration (NOEC/ECx) for adverse effects. In other words, in order to strengthen weight of evidence, a positive result in the XETA could be followed by an AMA at Level 3, which if positive in turn might lead to conduct of a LAGDA (Level 4). Existing data suggesting endocrine-specific activity (e.g. positive *in vitro* data, or positive *in vivo* data from other species) will strengthen the case for additional testing.

552. The situation in which the XETA gives a negative result (Table C.2.18, Scenarios J-R) needs careful consideration of any existing data. If these data suggest that the chemical is endocrine active both *in vitro* and *in vivo* (Scenario J), then it is possible that the XETA is simply insufficiently sensitive.

553. If the XETA and existing *in vivo* data are all negative, but *in vitro* data reveal some endocrine activity (Scenario K), the probability is that the test chemical is not sufficiently potent to produce thyroid effects *in vivo* in amphibians or other organisms, or it may be rapidly metabolised. In such a situation, further testing is probably not necessary. However, if the chemical is known to bioaccumulate slowly, it may be that exposures in the *in vivo* tests have been insufficiently prolonged, in which case longer term testing with the AMA or LAGDA might be justified.

554. On the other hand, if the XETA and the *in vitro* tests are negative (Scenario M), but there are positive existing *in vivo* data, the nature of those existing data should be considered. Unless the existing data are from another amphibian, the chemical is possibly not an ED acting in amphibians, but it may act via MOA not covered by the *in vitro* screens, or it may be more potent in species or life stages that have not been tested. In this situation,

the existing *in vivo* data should be used to guide decisions about whether to conduct any further testing.

555. Finally, a negative XETA, set against a background of negative *in vitro* and *in vivo* data (Scenario N), suggests that the test chemical is probably not a thyroid-active ED, and further action is unnecessary.

556. In each of the above scenarios, it is possible that existing data will be equivocal, or there may be no existing data. This will weaken the conclusions which can be drawn about a negative XETA, and this is reflected in Table C.2.18. However, a lack of mechanistic data on thyroid activity should ideally be rectified before any further *in vivo* testing is finally conducted, although as indicated above, many thyroid modalities are not detectable in *in vitro* screens. On the other hand, if the XETA is positive, further *in vivo* testing would generally be needed to quantify any adverse effects and/or to establish a NOEC or ECx for such effects, even if all existing data are equivocal, or if there are no existing data. Again, however, it may be useful to obtain some mechanistic information before conducting further *in vivo* testing. There is also the possibility that equivocal mechanistic data may be the result of multiple modes of endocrine action. Under some circumstances, two opposite modes of simultaneous action (e.g. thyroidogenic and anti-thyroidogenic) could, depending on dose, lead to a minimisation or abolition of adverse effects, while in others two different MOA could potentially reinforce effects the XETA endpoint. If multiple MOA are suspected, either from the existing results or based on QSAR/read-across/integrated approaches, this situation should be investigated further if needed for regulatory decision making.

557. The scenario in which the results of the XETA are themselves equivocal has not been dealt with in Table C.2.18, for reasons of brevity. In this context, an equivocal result might be an inconsistent concentration-response (e.g. no effect at a high concentration but effects at a lower concentration), or a result which borders on statistical significance. Without knowing the exact circumstances, reliable advice cannot be given, but the opinions of an experienced ecotoxicologist should be sought. Clearly, however, such equivocal results do not necessarily rule out the existence of *in vivo* endocrine activity. If possible reasons for false negatives are suspected, the XETA could be repeated (e.g. conduct it at lower concentrations which avoid systemic toxicity). However, note that a repeat screen in the event of systemic toxicity would not be needed providing at least one tested concentration was not subject to such effects.

558. In summary, positive results in the XETA may indicate that a chemical is a possible endocrine disrupter via one or more of several types of thyroid activity. This suggests that more comprehensive *in vivo* testing would be needed if the intention is to derive a long-term NOEC/ECx and/or to confirm whether or not the chemical is an actual endocrine disrupter due to the occurrence of adverse effects. Negative results in XETA do not necessarily mean that the chemical is not a potential ED – a judgement about the endocrine disruption potential and the possible need for additional testing will have to be made based on a weight of evidence evaluation of existing *in vitro* and *in vivo* data.

References

Morvan-Dubois, G., B.A. Demeneix and L.M. Sachs (2008), "*Xenopus laevis* as a model for studying thyroid hormone signalling: From development to metamorphosis", *Molecular and Cellular Endocrinology*, Vol. 293/1-2, pp. 71-79, https://doi.org/10.1016/j.mce.2008.06.012.

WHO/IPCS (2002), "Global assessment of the state-of-the-science of endocrine disrupters", Damstra, T. et al. (eds.) WHO/PCS/EDC/02.2, World Health Organization, Geneva, www.who.int/ipcs/publications/new_issues/endocrine_disruptors/en.

Table C.2.18. ***Xenopus*** **Embryonic Thyroid Signalling Assay (XETA) (draft OECD TG): Guidance for scenarios of combinations of results with existing data**

This table represents possible conclusions to be drawn from assay data, and a next step which could be taken if further evidence is required about possible endocrine disrupting properties and/or effects. The guidance offered is not meant to be prescriptive, but provides science-based considerations. It encourages the use of all available data and expert judgement in a weight of evidence approach. Regional and national interpretation of results and "next steps" may vary.

The conclusions are grouped into a series of scenarios (A-R), each scenario representing a different combination of assay results, existing *in vitro* data and existing *in vivo* data. The symbol "+" indicates that the data in question represent a positive result, «-" indicates a negative result, and "Eq/0" indicates that the data are either equivocal or are not available.

Existing results: * "Mechanism (*in vitro* mechanistic data)" assumes that mechanistic data are available from thyroid hormone receptor (TR) and other assays concerning mechanisms of thyroid disruption although these are not yet in common use. In practice, data from all assays may not be available and therefore this must be taken into account when deciding on the "next step". Quantitative structure activity relationship (QSAR) predictions of TR binding/activation may be made for some substances.

Existing results: ** "Effects (*in vivo* effects of concern)" assumes effects have been observed in other *in vivo* screens/tests which give rise to concern that the test chemical may be an thyroid disrupter.

Scenarios	Result of XETA	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
A	+	+	+	Strong evidence for *in vivo* thyroid activity in amphibians, plus thyroid effects in other species.	Consider performing an Amphibian Metamorphosis Assay (AMA – OECD TG 231).	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is a thyroid (ant)agonist. Note that although the AMA will provide some data on apical effects, more definitive information on adverse outcomes would be provided by the Larval Amphibian Growth and Development Assay (LAGDA – OECD TG 241).
B	+	+	–	Strong evidence for *in vivo* thyroid activity in amphibians.	Consider performing an AMA (OECD TG 231).	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is a thyroid (ant)agonist. Note that although the AMA will provide some data on apical effects, more definitive information on adverse outcomes would be provided by the LAGDA (OECD TG 241).
C	+	+	Eq/0	Strong evidence for *in vivo* thyroid activity in amphibians.	Consider performing an AMA (OECD TG 231).	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is a thyroid (ant)agonist. Note that although the AMA will provide some data on apical effects, more definitive information on adverse outcomes would be provided by the LAGDA (OECD TG 241). It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple modes of action (MOA). If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
D	+	–	+	Strong evidence for *in vivo* thyroid activity in amphibians, plus thyroid effects in other species.	Consider performing an AMA (OECD TG 231).	The lack of *in vitro* thyroid activity is not evidence against any thyroid activity, due to the limited nature of current *in vitro* thyroid screens. Note that although the AMA will provide some data on apical effects, more definitive information on adverse outcomes would be provided by the LAGDA (OECD TG 241).
E	+	–	–	Strong evidence for *in vivo* thyroid activity in amphibians.	Consider performing an AMA (OECD TG 231).	The lack of *in vitro* thyroid activity is not evidence against any thyroid activity, due to the limited nature of current *in vitro* thyroid screens. Note that although the AMA will provide some data on apical effects, more definitive information on adverse outcomes would be provided by the LAGDA (OECD TG 241).
F	+	–	Eq/0	Strong evidence for *in vivo* thyroid activity in amphibians.	Consider performing an AMA (OECD TG 231). Given the absence or equivocal nature of existing *in vivo* data, it might also be sensible to conduct a thyroid-responsive mammalian assay (e.g. rat pubertal).	The lack of *in vitro* thyroid activity is not evidence against any thyroid activity, due to the limited nature of current *in vitro* thyroid screens. Note that although the AMA will provide some data on apical effects, more definitive information on adverse outcomes would be provided by the LAGDA (OECD TG 241). It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

Scenarios	Result of JMASA	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
G	+	Eq/0	+	Strong evidence for *in vivo* thyroid activity in amphibians, plus thyroid effects in other species.	Consider performing an AMA (OECD TG 231). Given the absence or equivocal nature of the *in vitro* mechanistic data, it might also be helpful to conduct an *in vitro* screen for thyroid (ant)agonistic activity.	If a new *in vitro* mechanistic assay is conducted, note that a negative does not mean that the test material has no thyroid activity. Note that although the AMA will provide some data on apical effects, more definitive information on adverse outcomes would be provided by the LAGDA (OECD TG 241). It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
H	+	Eq/0	–	Strong evidence for *in vivo* thyroid activity in amphibians.	Consider performing an AMA (OECD TG 231). Given the absence or equivocal nature of the *in vitro* mechanistic data, it might also be helpful to conduct an *in vitro* screen for thyroid (ant)agonistic activity.	Note that although the AMA will provide some data on apical effects, more definitive information on adverse outcomes would be provided by the LAGDA (OECD TG 241). It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
I	+	Eq/0	Eq/0	Strong evidence for *in vivo* thyroid activity in amphibians.	Consider performing an AMA (OECD TG 231). Given the absence or equivocal nature of the *in vitro* mechanistic data, it might also be helpful to conduct an *in vitro* screen for thyroid (ant)agonistic activity. Given the absence or equivocal nature of existing *in vivo* data, it might also be sensible to conduct a thyroid-responsive mammalian assay (e.g. rat pubertal).	If a new *in vitro* mechanistic assay is conducted, note that a negative does not mean that the test material has no thyroid activity. Note that although the AMA will provide some data on apical effects, more definitive information on adverse outcomes would be provided by the LAGDA (OECD TG 241). It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
J	–	+	+	The test chemical is probably a thyroid (ant)agonist without activity in amphibians, although it is possible that *Xenopus laevis* responds atypically in this case.	Some regulatory authorities may conclude that no further evidence is required.	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is a thyroid (ant)agonist.
K	–	+	–	The test chemical is probably a thyroid (ant)agonist without activity in amphibians or other taxa, although it is possible that *Xenopus laevis* responds atypically in this case.	If there is no activity in amphibians or mammals, further evidence is probably not needed.	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is a thyroid (ant)agonist.

Scenarios	Result of JMASA	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
L	–	+	Eq/0	The test chemical is probably a thyroid (ant)agonist without activity in amphibians, although it is possible that *Xenopus laevis* responds atypically in this case.	Some regulatory authorities may conclude that no further evidence is required, but if mammalian data are absent, it might be desirable to conduct a thyroid-responsive rodent screen (e.g. rat pubertal).	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is a thyroid (ant)agonist. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
M	–	–	+	The test chemical is probably without thyroid activity in amphibians, although it is possible that *Xenopus laevis* responds atypically in this case.	Some regulatory authorities may conclude that no further evidence is required.	The lack of *in vitro* thyroid activity is not evidence against any thyroid activity, due to the limited nature of current *in vitro* thyroid screens.
N	–	–	–	The test chemical is probably without thyroid activity in amphibians or other taxa.	No further action is necessary.	–
O	–	–	Eq/0	The test chemical is probably without thyroid activity in amphibians.	Some regulatory authorities may conclude that no further evidence is required, but if mammalian data are absent, it might be desirable to conduct a thyroid-responsive rodent screen (e.g. rat pubertal).	The lack of *in vitro* thyroid activity is not evidence against any thyroid activity, due to the limited nature of current *in vitro* thyroid screens. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
P	–	Eq/0	+	The test chemical is probably without thyroid activity in amphibians, although it is possible that *Xenopus laevis* responds atypically in this case.	Some regulatory authorities may conclude that no further evidence is required. Also, if clear *in vitro* mechanistic data are missing, it might be desirable to obtain some.	If a new *in vitro* mechanistic assay is conducted, note that a negative does not mean that the test material has no thyroid activity. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
Q	–	Eq/0	–	The test chemical is probably without thyroid activity in amphibians or other taxa.	No further action is necessary.	It should be borne in mind that equivocal data may be due to a variety of causes including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
R	–	Eq/0	Eq/0	The test chemical is probably without thyroid activity in amphibians.	Some regulatory authorities may conclude that no further evidence is required, but if mammalian data are absent, it might be desirable to conduct a thyroid-responsive rodent screen (e.g. rat pubertal).	It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

C.2.19. Harpacticoid Copepod Development and Reproduction Test with *Amphiascus* (OECD GD 201)

Status: Partially validated by the OECD.

559. Modality detected/endpoints: This long-term *in vivo* assay with the marine copepod crustacean *Amphiascus tenuiramis* could not be fully validated as an OECD test guideline (TG) due to the complexity of the methodology which requires considerable operator training. Although the assay has not been conducted with hormonally based insect growth regulators, it is expected (by extension from experience with other copepods) to be responsive to juvenile hormone (JH) (ant)agonists and ecdysteroid (Ec) (ant)agonists which can interfere with such processes as metamorphosis, moulting, growth and reproduction. *A. tenuiramis* is known to contain the ecdysone receptor (Gaertner et al., 2012) and the moulting hormone 20-hydroxyecdysone (Block, Bejarano and Chandler, 2003). The assay exposes the test organisms over at least one generation. It is important to note, however, that none of the endpoints in this apical test are specifically responsive to JH- or Ec-active chemicals, and the assay will give positive results with many other substances. The lack of internationally validated mechanistic assays for endocrine activity in crustaceans may prevent firm conclusions about whether test chemicals are endocrine disruptors (EDs) in this taxon, although *in vitro* assays for JH and Ec activity are available in the literature. However, the data from the test may nevertheless be of value for classification and hazard identification/characterisation.

Background to the assay

560. This is a one-generation growth and reproduction assay using the marine benthic copepod crustacean *Amphiascus tenuiramis*. It is not specifically sensitive to JH or Ec (ant)agonists, but is expected to be apically responsive to some of them. However, many non-endocrine toxicants will also produce a response. The assay is technically demanding and operators require significant training before it can be conducted repeatably. Newly hatched larvae (F0) <24-hours-old are exposed individually until adulthood, at which time they are paired and allowed to breed. The test can be terminated after two clutches of offspring (F1) have been produced, and the whole test from F0 to F1 larvae takes 36 days. The test can optionally be extended to study survival and reproduction of the F1 animals. Endpoints include survival of the F0 generation, developmental rate, time to production of the first and second clutches, reproductive success, clutch size, fertility, and number of viable hatched F1 offspring (nauplii).

When/why the assay may be used

561. Although OECD GD 201 could, in principle, be used at any stage in the hazard assessment process, the most likely use scenario will be when there are already some *in vitro* and *in vivo* data available about the possible JH or Ec activity and/or effects of a chemical. Given the significant degree of endocrine system conservation across the arthropods, effects in OECD GD 201 may also indicate the possibility of related activity in other crustaceans (e.g. cladocera and decapods) and insects.

562. It is not recommended that OECD GD 201 be deployed as a primary test for JH or Ec activity and effects, but it should be noted that there are no standardised *in vitro* screens for JH or Ec (ant)agonists, although some are described in the scientific literature (e.g. Cherbas, Koehler and Cherbas [1989]; Dinan et al. [2001]; Smagghe et al. [2003]; Swevers et al. [2003]).

563. In order to provide information relevant for assessing whether or not a chemical may fulfil the WHO/IPCS (2002) definition of an ED, the study design has to be sufficiently robust to demonstrate the presence or absence of effects. In the dose selection, the investigator should also consider and ensure that data generated are adequate to fulfil the regulatory requirement across OECD countries as appropriate (e.g. hazard and risk assessment and labelling, ED assessment, etc.). The top dose or concentration should be sufficiently high to give clear systemic (i.e. non endocrine-specific) toxicity in order to ensure that a wide range of exposures (high to low) is tested. However, endocrine effects observed solely in the presence of clear systemic toxicity should be interpreted with caution and may be disregarded when sufficiently justified to be caused by secondary effects which are unlikely to be due to endocrine activity. The reason for this advice is a concern that some endocrine active substance (EAS) sensitive assays are being run at doses/concentrations of EASs that are too low to trigger direct impacts on the endocrine system. This guidance document is not the place to address this issue directly, but it should be considered when EAS-sensitive TGs are revised in the future. In addition, the number and spacing of dose/concentration levels should also be adequate to fulfil the objectives of the study (e.g. to demonstrate dose response relationships if this is required).

Existing data to be considered

564. Existing information on endocrine-related effects from other arthropods should also be considered before deployment of OECD GD 201, given the commonality of endocrine mechanisms in these taxa. Existing data available might also include one or more of a range of *in silico* or *in vitro* results which suggest that JH or Ec disruption may occur *in vivo* (but note the limitations of this approach, as indicated above). Such indicators of possible JH or Ec activity might include quantitative structure activity relationship (QSAR) predictions of JH/Ec activity, "read-across" from *in vivo* results obtained with structurally related chemicals or positive results from an *in vitro* screen for JH/Ec (ant)agonist activity. In addition, *in vivo* data should ideally be available from one or more of several assays, possibly including the Short-Term Juvenile Hormone Activity Screening Assay (SJHASA), the Sediment-Water Chironomid Toxicity Test Using Spiked Sediment or Water (OECD TG 218/219), the Sediment Water Chironomid Life Cycle Toxicity Test (OECD TG 233), the *Daphnia magna* Reproduction Test with male neonate option (OECD TG 211) or the *Daphnia* Multigeneration Test (DMGT).

Scenarios: Positive and negative results combined with existing data

565. The scenarios (A to R) presented in Table C.2.19 represent all the possibilities of positive or negative results in combination with the presence or absence of existing data. The action taken will also depend on the regulatory environment, but the considerations given here are generally science based. Wherever possible, the recommended "next step which could be taken" avoids unnecessary animal testing. However, sometimes conducting an animal test will be indicated and then the relevance of species, strain and exposure route should always be considered. Further considerations, specific to each scenario are given in the table.

566. Positive results obtained with OECD GD 201 (Table C.2.19, Scenarios A-I) result in the conclusion that the test chemical has adverse apical effects, at least in crustaceans, but these are not necessarily caused by JH or Ec activity. If countries need further evidence concerning growth and sexual development, etc. in this phylum, a Chironomid Life Cycle Toxicity Test (OECD TG 233) would be able to provide information on adverse effects in insects. In other words, in order to strengthen weight of evidence, a positive result in OECD GD 201 could be followed by the OECD TG 233 (Level 5). Existing data suggesting endocrine-specific activity (e.g. positive *in vitro* data, or positive *in vivo* data from other species) will strengthen the case for additional testing still further.

567. The situation in which OECD GD 201 gives a negative result (Table C.2.19, Scenarios J-R) needs careful consideration of any existing data. If these data suggest that the chemical is endocrine active both *in vitro* and *in vivo* (Scenario J), then it is possible that OECD GD 201 is simply insufficiently sensitive.

568. If OECD GD 201 and existing *in vivo* data are all negative, but *in vitro* data reveal some JH or Ec activity (Scenario K), the probability is that the test chemical is not sufficiently potent to produce JH/Ec (ant)agonism *in vivo* in arthropods, or it may be rapidly metabolised. In such a situation, further testing is probably not necessary.

569. On the other hand, if OECD GD 201 and the *in vitro* tests are negative (Scenario M), but there are positive existing *in vivo* data, the nature of those existing data should be considered. Unless the existing data are from another crustaceans, the chemical is possibly not a JH or Ec (ant)agonist acting in crustaceans, but it may be more potent in species (e.g. insects) or life stages that have not been tested. In this situation, the existing *in vivo* data should be used to guide decisions about whether to conduct any further testing (e.g. with OECD TG 233).

570. Finally, a negative OECD GD 201, set against a background of negative *in vitro* and *in vivo* data (Scenario N), suggests that the test chemical is probably not a JH or Ec (ant)agonist *in vitro* or *in vivo*, and further action is unnecessary.

571. In each of the above scenarios, it is possible that existing data will be equivocal, or there may be no existing data. This will weaken the conclusions which can be drawn about a negative OECD GD 201, and this is reflected in Table C.2.19. However, a lack of mechanistic data on JH or Ec activity should ideally be rectified before any further *in vivo* testing is finally conducted, although as indicated above, *in vitro* JH/Ec screens have not yet been internationally standardised. On the other hand, if OECD GD 201 is positive, further *in vivo* testing would generally be needed to quantify any adverse effects in insects, even if all existing data are equivocal, or if there are no existing data. Again, however, it may be useful to obtain some mechanistic information before conducting further *in vivo* testing. There is also the possibility that equivocal mechanistic data may be the result of multiple modes of endocrine action. Under some circumstances, two opposite modes of simultaneous action (e.g. JH or Ec agonistic and antagonistic) could, depending on dose, lead to a minimisation or abolition of adverse effects, while in others two different MOA could potentially reinforce effects on the OECD GD 201 endpoint. If multiple MOA are suspected, either from the existing results or based on QSAR/read-across/integrated approaches, this situation should be investigated further if needed for regulatory decision making.

572. The scenario in which the results of OECD GD 201 are themselves equivocal has not been dealt with in Table C.2.19, for reasons of brevity. In this context, an equivocal result might be an inconsistent concentration-response (e.g. no effect at a high concentration but effects at a lower concentration), or a result which borders on statistical

significance. Without knowing the exact circumstances, reliable advice cannot be given, but the opinions of an experienced ecotoxicologist should be sought. Clearly, however, such equivocal results do not necessarily rule out the existence of *in vivo* endocrine activity. If possible reasons for false negatives are suspected, OECD GD 201 could be repeated (e.g. conduct it at lower concentrations which avoid systemic toxicity).

573. In summary, positive results in OECD GD 201 indicate that a chemical has adverse effects in crustaceans which may or may not be via JH or Ec (ant)agonism. This may need to be followed up with further apical testing with insects. Negative results in OECD GD 201 do not necessarily mean that the chemical is not a potential ED – a judgement about the endocrine disruption potential in other arthropods will have to be made based on a weight of evidence evaluation of existing *in vitro* and *in vivo* data.

References

Block, D.S., A.C. Bejarano and G.T. Chandler (2003), "Ecdysteroid concentrations through various life-stages of the meiobenthic harpacticoid copepod, *Amphiascus tenuiramis* and the benthic estuarine amphipod, *Leptocheirus plumulosus*", *General and Comparative Endocrinology*, Vol. 132/1, pp. 151-160, https://doi.org/10.1016/S0016-6480(03)00062-5.

Cherbas, L., M.M.D. Koehler and P. Cherbas (1989), "Effects of juvenile hormone on the ecdysone response of *Drosophila* Kc cells", *Developmental Genetics*, Vol. 10/3, pp. 177-188, https://doi.org/10.1002/dvg.1020100307.

Dinan, L. et al. (2001), "Screening of environmental contaminants for ecdysteroid agonistand antagonist activity using the *Drosophila melanogaster* B-II cell *in vitro* assay", *Environmental Toxicology and Chemistry*, Vol. 20/9, pp. 2038-2046, https://doi.org/10.1002/etc.5620200924.

Gaertner, K. et al. (2012), "Identification and expression of the ecdysone receptor in the harpacticoid copepod, *Amphiascus tenuiramis*, in response to fipronil", *Ecotoxicology and Environmental Safety*, Vol. 76, pp. 39-45, https://doi.org/10.1016/j.ecoenv.2011.09.008.

Smagghe, G. et al. (2003), "Cultured mosquito cells *Aedes albopictus* C6/36 (Dip, Culicidae) responsive to 20-hydroxyecdysone and non-steroidal ecdysone antagonists", *Journal of Applied Entomology*, Vol. 127/3, pp. 167-173, https://doi.org/10.1046/j.1439-0418.2003.00727.x.

Swevers, L. et al. (2003), "A high-throughput screening system for fast detection of ecdysteroid mimetic and antagonistic substances using transformed *Bombyx mori*-derived cell lines", *The FASEB Journal*, Vol. 18/1, pp. 134-136, http://dx.doi.org/10.1096/fj.03-0627fje.

WHO/IPCS (2002), "Global assessment of the state-of-the-science of endocrine disrupters", Damstra, T. et al. (eds.) WHO/PCS/EDC/02.2, World Health Organization, Geneva, www.who.int/ipcs/publications/new_issues/endocrine_disruptors/en.

Table C.2.19. **Harpacticoid Copepod Development and Reproduction Test with *Amphiascus* (OECD GD 201): Guidance for scenarios of combinations of results with existing data**

This table represents possible conclusions to be drawn from assay data, and a next step which could be taken if further evidence is required about possible endocrine disrupting properties and/or effects. The guidance offered is not meant to be prescriptive, but provides science-based considerations. It encourages the use of all available data and expert judgement in a weight of evidence approach. Regional and national interpretation of results and "next steps" may vary.

The conclusions are grouped into a series of scenarios (A-R), each scenario representing a different combination of assay results, existing *in vitro* data and existing *in vivo* data. The symbol "+" indicates that the data in question represent a positive result, «-" indicates a negative result, and "Eq/0" indicates that the data are either equivocal or are not available.

Existing results: * "Mechanism (*in vitro* mechanistic data)" assumes that mechanistic data are available from available from juvenile hormone (JH-) or ecdysteroid (Ec-) based assays. JH or Ec assays concerning mechanisms of disruption may be available, but they are have not yet been internationally standardised. In practice, data from all assays may not be available and therefore this must be taken into account when deciding on the "next step".

Existing results: ** "Effects (*in vivo* effects of concern)" assumes effects have been observed in other *in vivo* screens/tests which give rise to concern that the test chemical may be a JH or Ec disrupter.

Scenarios	Result of OECD GD 201	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
A	+	+	+	Strong evidence for adverse *in vivo* effects in crustaceans, possibly but not necessarily caused by juvenile hormone (JH) or ecdysteroid (Ec) (ant)agonists, plus possible JH or Ec effects in other arthropods.	It would be desirable (if not already conducted) to perform an apical test with insects (e.g. the Chironomid Life Cycle Toxicity Test – OECD TG 233).	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is a JH or Ec (ant)agonist.
B	+	+	–	Moderate evidence for adverse *in vivo* effects in crustaceans, possibly but not necessarily caused by JH or Ec (ant)agonists.	It would be desirable (if not already conducted) to perform an apical test with insects (e.g. the Chironomid Life Cycle Toxicity Test – OECD TG 233).	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is a JH or Ec (ant)agonist.
C	+	+	Eq/0	Moderate evidence for adverse *in vivo* effects in crustaceans, possibly but not necessarily caused by JH or Ec (ant)agonists.	It would be desirable (if not already conducted) to perform an apical test with insects (e.g. the Chironomid Life Cycle Toxicity Test – OECD TG 233).	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is a JH or Ec (ant)agonist. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple modes of action (MOA). If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
D	+	–	+	Strong evidence for adverse *in vivo* effects in crustaceans, possibly but not necessarily caused by JH or Ec (ant)agonists, plus possible JH or Ec effects in other arthropods.	It would be desirable (if not already conducted) to perform an apical test with insects (e.g. the Chironomid Life Cycle Toxicity Test – OECD TG 233).	The lack of *in vitro* JH or Ec activity is not evidence against any JH/Ec activity, due to the limited nature of current *in vitro* JH and Ec screens.
E	+	–	–	Some evidence for adverse *in vivo* effects in crustaceans, possibly but not necessarily caused by JH or Ec (ant)agonists.	It would be desirable (if not already conducted) to perform an apical test with insects (e.g. the Chironomid Life Cycle Toxicity Test – OECD TG 233).	The lack of *in vitro* JH or Ec activity is not evidence against any JH/Ec activity, due to the limited nature of current *in vitro* JH and Ec screens.
F	+	–	Eq/0	Some evidence for adverse *in vivo* effects in crustaceans, possibly but not necessarily caused by JH or Ec (ant)agonists.	Given the absence or equivocal nature of existing *in vivo* data, it might also be sensible to conduct a JH-responsive crustacean assay (e.g. the *Daphnia* Multigeneration Test [DMGT] and/or a JH/E-responsive insect assay [e.g. the Chironomid Life Cycle Toxicity Test – OECD TG 233]).	The lack of *in vitro* JH/Ec activity is not evidence against any JH/Ec activity, due to the limited nature of current *in vitro* JH/Ec screens. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

Scenarios	Result of OECD GD 201	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
G	+	Eq/0	+	Strong evidence for adverse *in vivo* effects in crustaceans, possibly but not necessarily caused by JH or Ec (ant)agonists, plus possible JH or Ec effects in other arthropods.	Given the absence or equivocal nature of existing *in vitro* data, it would be desirable to obtain further *in vitro* data on JH/Ec activity if possible. It might also be sensible to conduct a JH/Ec-responsive insect assay if not already performed (e.g. the Chironomid Life Cycle Toxicity Test – OECD TG 233).	If a new *in vitro* mechanistic assay is conducted, note that a negative does not mean that the test material has no JH/Ec activity. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
H	+	Eq/0	–	Some evidence for adverse *in vivo* effects in crustaceans, possibly but not necessarily caused by JH or Ec (ant)agonists.	Given the absence or equivocal nature of the *in vitro* mechanistic data, it might also be helpful to conduct an *in vitro* screen for JH/Ec activity. It might also be sensible to conduct a JH/Ec-responsive insect assay if not already performed (e.g. the Chironomid Life Cycle Toxicity Test – OECD TG 233).	If a new *in vitro* mechanistic assay is conducted, note that a negative does not mean that the test material has no JH/Ec activity. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
I	+	Eq/0	Eq/0	Strong evidence for adverse *in vivo* effects in crustaceans, possibly but not necessarily caused by JH or Ec (ant)agonists.	Given the absence or equivocal nature of the *in vitro* mechanistic data, it might also be helpful to conduct an *in vitro* screen for JH/Ec activity. It might also be sensible to conduct a JH/Ec-responsive insect assay (e.g. the Chironomid Life Cycle Toxicity Test – OECD TG 233).	If a new *in vitro* mechanistic assay is conducted, note that a negative does not mean that the test material has no JH/Ec activity. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
J	–	+	+	The test chemical is probably a JH or Ec (ant)agonist without adverse effects in crustaceans, although it is possible that *Amphiascus* responds atypically in this case.	Some regulatory authorities may conclude that no further evidence is required. However, it might also be sensible to conduct a JH/Ec-responsive insect assay if not already performed (e.g. the Chironomid Life Cycle Toxicity Test – OECD TG 233).	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is a JH/Ec (ant)agonist.
K	–	+	–	The test chemical is probably a JH or Ec (ant)agonist without adverse effects in crustaceans or other taxa, although it is possible that *Amphiascus* responds atypically in this case.	If there is no activity in crustaceans or insects, further evidence is probably not needed.	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is a JH/Ec (ant)agonist.

Scenarios	Result of OECD GD 201	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
L	–	+	Eq/0	The test chemical is probably a JH or Ec (ant)agonist without adverse effects in crustaceans, although it is possible that *Amphiascus* responds atypically in this case.	Some regulatory authorities may conclude that no further evidence is required, but if insect data are absent, it might be desirable to conduct a Chironomid Life Cycle Toxicity Test (OECD TG 233).	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is a JH/Ec agonist. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
M	–	–	+	The test chemical is probably without JH or Ec activity in crustaceans, although it is possible that *Chironomus* responds atypically in this case.	Some regulatory authorities may conclude that no further evidence is required. However, it might be desirable to obtain data from insects (e.g. the Chironomid Life Cycle Toxicity Test – OECD TG 233) if these are not already available.	The lack of *in vitro* JH/Ec activity is not evidence against any JH activity, due to the limited nature of current *in vitro* JH screens.
N	–	–	–	The test chemical is probably without JH or Ec activity in arthropods.	No further action is necessary.	–
O	–	–	Eq/0	The test chemical is probably without JH or Ec activity in arthropods.	Some regulatory authorities may conclude that no further evidence is required. However, it might be desirable to obtain data from insects (e.g. Chironomid Life Cycle Toxicity Test – OECD TG 233) if these are not already available.	The lack of *in vitro* JH/Ec activity is not evidence against any JH activity, due to the limited nature of current *in vitro* JH screens. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
P	–	Eq/0	+	The test chemical is probably without JH or Ec activity in crustaceans, although it is possible that *Amphiascus* responds atypically in this case.	Some regulatory authorities may conclude that no further evidence is required. Also, if clear *in vitro* mechanistic data are missing, it might be desirable to obtain some.	If a new *in vitro* mechanistic assay is conducted, note that a negative does not mean that the test material has no JH/Ec activity. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

Scenarios	Result of OECD GD 201	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
Q	–	Eq/0	–	The test chemical is probably without JH or Ec activity in arthropods.	No further action is necessary.	It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
R	–	Eq/0	Eq/0	The test chemical is probably without JH or Ec activity in crustaceans and possibly insects.	Some regulatory authorities may conclude that no further evidence is required. However, it might be desirable to obtain data from insects (e.g. the Chironomid Life Cycle Toxicity Test – OECD TG 233) if these are not already available.	It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

C.2.20. *Daphnia* Multigeneration Test for Assessment of Endocrine-Active Chemicals (DMGT) (proposed OECD TG)

Status: Assay proposed for validation by the OECD.

574. Modality detected/endpoints: This long-term *in vivo* assay with *Daphnia magna* is responsive to juvenile hormone (JH) agonists which lead to the production of male offspring. It exposes the test organisms over two generations. The lack of internationally validated mechanistic assays for endocrine activity in crustaceans may prevent firm conclusions about whether test chemicals are endocrine disruptors (EDs) in this taxon, although *in vitro* assays for JH and ecdysteroid (Ec) activity are available in the literature. However, the data from the test may nevertheless be of value for classification and hazard identification/characterisation.

Background to the assay

575. This *in vivo* assay has been proposed but not yet approved as an OECD project. The validation by the OECD has not yet begun and the proposal is not expected to be approved as a test guideline (TG) until 2019 at the earliest. As such, the guidance in this section should be regarded as provisional, and it should be noted that the protocol outlined below may change. The DMGT at present consists of three linked exposure experiments. It begins with <24-hour-old neonates, exposes them continuously to dilutions of the test chemical, allows them to grow to adulthood, then produce at least three successive broods (termed the "F1 test", run for 21 days). The second test (termed the "F2: F1 exposed test") takes neonates from the third or subsequent brood in each concentration of the F1 test and exposes them to the same range of test concentrations for a further 21 days. The third test (termed the "F2: F1 unexposed test") solely takes control neonates from the third or subsequent brood in the F1 test and again exposes them for 21 days to the same range of concentrations as in the other tests. At the end of each test, all individual neonates are sexed by observation of their longer first antenna. JH and other JH agonists cause the production of males due to exposure during a short critical period (52-53 hours after ovulation). An adverse outcome pathway for this process has been described[1] – significant male production in a population would be expected to lead to its decline. The production in this test of a sex ratio significantly skewed towards males can therefore probably be considered as an adverse apical endpoint.

576. Limited data produced during the development of this test suggest that JH agonists may give have a more pronounced effect on sex ratio in the "F2: F1 exposed" test than the "F2: F1 unexposed" test. However, more data are required for this to be substantiated.

When/why the assay may be used

577. Although the DMGT could, in principle, be used at any stage in the hazard assessment process, the most likely use scenario will be when there are already some *in vitro* and *in vivo* data available about the possible JH-disrupting properties of a chemical. Given the significant degree of endocrine system conservation across the arthropods,

endocrine-linked effects in the DMGT may also indicate the possibility of related activity in other arthropods such as copepods, decapods and insects.

578. It is possible that no endocrine-relevant data are available before the DMGT is deployed (i.e. if the DMGT has been used as a primary test, or has been deployed to test a chemical for non-endocrine related chronic toxicity), but in that case a positive result in the test should probably be followed up with relevant *in vitro* screening, if available, to investigate the suspected mode of action (MOA) in more detail. However, it should be noted that there are no standardised *in vitro* screens for JH agonists, although some are described in the scientific literature (e.g. Cherbas, Koehler and Cherbas [1989]).

579. In order to provide information relevant for assessing whether or not a chemical may fulfil the WHO/IPCS (2002) definition of an ED, the study design has to be sufficiently robust to demonstrate the presence or absence of effects. In the dose selection, the investigator should also consider and ensure that data generated are adequate to fulfil the regulatory requirement across OECD countries as appropriate (e.g. hazard and risk assessment and labelling, ED assessment, etc.). The top dose or concentration should be sufficiently high to give clear systemic (i.e. non endocrine-specific) toxicity in order to ensure that a wide range of exposures (high to low) is tested. However, endocrine effects observed solely in the presence of clear systemic toxicity should be interpreted with caution and may be disregarded when sufficiently justified to be caused by secondary effects which are unlikely to be due to endocrine activity. The reason for this advice is a concern that some endocrine active substance (EAS) sensitive assays are being run at doses/concentrations of EASs that are too low to trigger direct impacts on the endocrine system. This guidance document is not the place to address this issue directly, but it should be considered when EAS-sensitive TGs are revised in the future. In addition, the number and spacing of dose/concentration levels should also be adequate to fulfil the objectives of the study (e.g. to demonstrate dose response relationships if this is required).

Existing data to be considered

580. Existing information on endocrine-related effects from other arthropods should also be considered before deployment of the DMGT, given the commonality of endocrine mechanisms in these taxa. Existing data available might also include one or more of a range of *in silico* or *in vitro* results which suggest that JH disruption may occur *in vivo* (but note the limitations of this approach, as indicated above). Such indicators of possible JH activity might include quantitative structure activity relationship (QSAR) predictions of JH activity, "read-across" from *in vivo* results obtained with structurally related chemicals or positive results from an *in vitro* screen for JH agonist activity. In addition, *in vivo* data should ideally be available from one of two assays, the Short-Term Juvenile Hormone Activity Screening Assay (SJHASA) or the *Daphnia magna* Reproduction Test with male neonate option (OECD TG 211).

Scenarios: Positive and negative results combined with existing data

581. The scenarios (A to R) presented in Table C.2.20 represent all the possibilities of positive or negative results in combination with the presence or absence of existing data. The action taken will also depend on the regulatory environment, but the considerations given here are generally science based. Wherever possible, the recommended "next step which could be taken" avoids unnecessary animal testing. However, sometimes conducting an animal test will be indicated and then the relevance of species, strain and exposure route should

always be considered. Further considerations, specific to each scenario are given in the table.

582. Positive results obtained with the DMGT (Table C.2.20, Scenarios A-I) result in the conclusion that the test chemical is a possible JH disrupter *in vivo* with adverse apical effects, at least in crustaceans. However, although a positive response of the DMGT indicates that the chemical is a possible JH agonist with adverse effects in crustaceans, it should be noted that *Daphnia*'s parthenogenetic reproductive strategy is not shared by many other arthropods. Therefore, if countries need further evidence concerning growth and sexual development, etc. in this phylum, a Harpacticoid Copepod Development and Reproduction Test (OECD GD 201) and/or the Sediment-Water Chironomid Life Cycle Toxicity Test (OECD TG 233) would be able to provide information on adverse effects in sexually reproducing species. In other words, in order to strengthen weight of evidence, a positive result in the DMGT could be followed by OECD GD 201 (Level 4) and/or OECD TG 233 (Level 5). Existing data suggesting endocrine-specific activity (e.g. positive *in vitro* data, or positive *in vivo* data from other species) will strengthen the case for additional testing still further.

583. The situation in which the DMGT gives a negative result (Table C.2.20, Scenarios J-R) needs careful consideration of any existing data. If these data suggest that the chemical is endocrine active both *in vitro* and *in vivo* (Scenario J), then it is possible that the DMGT is simply insufficiently sensitive.

584. If the DMGT and existing *in vivo* data are all negative, but *in vitro* data reveal some JH activity (Scenario K), the probability is that the test chemical is not sufficiently potent to produce JH agonism *in vivo* in arthropods, or it may be rapidly metabolised. In such a situation, further testing is probably not necessary. However, if the chemical is known to bioaccumulate slowly, it may be that exposures in the *in vivo* tests have been insufficiently prolonged, in which case longer term testing with OECD GD 201 or OECD TG 233 might be justified.

585. On the other hand, if the DMGT and the *in vitro* tests are negative (Scenario M), but there are positive existing *in vivo* data, the nature of those existing data should be considered. Unless the existing data are from another crustacean, the chemical is possibly not a JH agonist acting in crustaceans, but it may be more potent in species (e.g. insects) or life stages that have not been tested. In this situation, the existing *in vivo* data should be used to guide decisions about whether to conduct any further testing (e.g. with OECD TG 233).

586. Finally, a negative DMGT, set against a background of negative *in vitro* and *in vivo* data (Scenario N), suggests that the test chemical is probably not a JH agonist *in vitro* or *in vivo*, and further action is unnecessary.

587. In each of the above scenarios, it is possible that existing data will be equivocal, or there may be no existing data. This will weaken the conclusions which can be drawn about a negative DMGT, and this is reflected in Table C.2.20. However, a lack of mechanistic data on JH activity should ideally be rectified before any further *in vivo* testing is finally conducted, although as indicated above, *in vitro* JH screens have not yet been internationally standardised. On the other hand, if the DMGT is positive, further *in vivo* testing would generally be needed to quantify any adverse effects in crustaceans and/or insects, even if all existing data are equivocal, or if there are no existing data. Again, however, it may be useful to obtain some mechanistic information before conducting further *in vivo* testing. There is also the possibility that equivocal mechanistic data may be the result

of multiple modes of endocrine action. Under some circumstances, two opposite modes of simultaneous action (e.g. JH agonistic and antagonistic) could, depending on dose, lead to a minimisation or abolition of adverse effects, while in others two different MOA could potentially reinforce effects on the DMGT endpoint. If multiple MOA are suspected, either from the existing results or based on QSAR/read-across/integrated approaches, this situation should be investigated further if needed for regulatory decision making.

588. The scenario in which the results of the DMGT are themselves equivocal has not been dealt with in Table C.2.20, for reasons of brevity. In this context, an equivocal result might be an inconsistent concentration-response (e.g. no effect at a high concentration but effects at a lower concentration), or a result which borders on statistical significance. Without knowing the exact circumstances, reliable advice cannot be given, but the opinions of an experienced ecotoxicologist should be sought. Clearly, however, such equivocal results do not necessarily rule out the existence of *in vivo* endocrine activity. If possible reasons for false negatives are suspected, the DMGT could be repeated (e.g. conduct it at lower concentrations which avoid systemic toxicity). It should also be borne in mind that changing environmental conditions such as shortening photoperiod, temperature, and food shortages can also cause the production of male neonates in *D. magna*, so if these have accidentally occurred during the test, the results should be treated as suspect.

589. In summary, positive results in the DMGT indicate that a chemical is a probable endocrine disrupter with adverse effects in crustaceans via JH agonism. This may need to be followed up with further apical testing with sexually reproducing arthropods. Negative results in the DMGT do not necessarily mean that the chemical is not a potential ED – a judgement about the endocrine disruption potential in other arthropods will have to be made based on a weight of evidence evaluation of existing *in vitro* and *in vivo* data.

Note

1. See: https://aopwiki.org/wiki/index.php/Aop:201.

References

Cherbas, L., M.M.D. Koehler and P. Cherbas (1989), "Effects of juvenile hormone on the ecdysone response of *Drosophila* Kc cells", *Developmental Genetics*, Vol. 10/3, pp. 177-188, https://doi.org/10.1002/dvg.1020100307.

WHO/IPCS (2002), "Global assessment of the state-of-the-science of endocrine disrupters", Damstra, T. et al. (eds.) WHO/PCS/EDC/02.2, World Health Organization, Geneva, www.who.int/ipcs/publications/new_issues/endocrine_disruptors/en.

Table C.2.20. ***Daphnia*** **Multigeneration Test for Assessment of Endocrine-Active Chemicals (DMGT) (proposed OECD TG – SPSF not yet agreed): Guidance for scenarios of combinations of results with existing data**

This table represents possible conclusions to be drawn from assay data, and a next step which could be taken if further evidence is required about possible endocrine disrupting properties and/or effects. The guidance offered is not meant to be prescriptive, but provides science-based considerations. It encourages the use of all available data and expert judgement in a weight of evidence approach. Regional and national interpretation of results and "next steps" may vary.

The conclusions are grouped into a series of scenarios (A-R), each scenario representing a different combination of assay results, existing *in vitro* data and existing *in vivo* data. The symbol "+" indicates that the data in question represent a positive result, "'-" indicates a negative result and "Eq/0" indicates that the data are either equivocal or are not available.

Existing results: * "Mechanism (*in vitro* mechanistic data)" assumes that mechanistic data are available from available from juvenile hormone (JH-) based assays. JH assays concerning mechanisms of JH disruption may be available, but they are have not yet been internationally standardised. In practice, data from all assays may not be available and therefore this must be taken into account when deciding on the "next step".

Existing results: ** "Effects (*in vivo* effects of concern)" assumes effects have been observed in other *in vivo* screens/tests which give rise to concern that the test chemical may be a JH disrupter.

Scenarios	Result of DMGT	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
A	+	+	+	Evidence for adverse *in vivo* effects in crustaceans caused by juvenile hormone (JH) and JH mimics, plus possible JH effects in other arthropods.	As *Daphnia* are parthenogenetic, it would be desirable (if not already conducted) to perform an additional apical test with sexually reproducing crustaceans and/or insects (e.g. the Harpacticoid Copepod Development and Reproduction Test – OECD GD 201; and/or the Sediment-Water Chironomid Life Cycle Toxicity Test – OECD TG 233).	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is a JH agonist.
B	+	+	–	Evidence for adverse *in vivo* effects in crustaceans caused by JH and JH mimics.	As *Daphnia* are parthenogenetic, it would be desirable (if not already conducted) to perform an additional apical test with sexually reproducing crustaceans and/or insects (e.g. the Harpacticoid Copepod Development and Reproduction Test – OECD GD 201; and/or the Sediment-Water Chironomid Life Cycle Toxicity Test – OECD TG 233).	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is a JH agonist.
C	+	+	Eq/0	Evidence for adverse *in vivo* effects in crustaceans caused by JH and JH mimics.	As *Daphnia* are parthenogenetic, it would be desirable (if not already conducted) to perform an additional apical test with sexually reproducing crustaceans and/or insects (e.g. the Harpacticoid Copepod Development and Reproduction Test – OECD GD 201; and/or the Sediment-Water Chironomid Life Cycle Toxicity Test – OECD TG 233).	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is a JH agonist. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple modes of action (MOA). If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
D	+	–	+	Evidence for adverse *in vivo* effects in crustaceans caused by JH and JH mimics, plus possible JH effects in other arthropods.	As *Daphnia* are parthenogenetic, it would be desirable (if not already conducted) to perform an additional apical test with sexually reproducing crustaceans and/or insects (e.g. the Harpacticoid Copepod Development and Reproduction Test – OECD GD 201; and/or the Sediment-Water Chironomid Life Cycle Toxicity Test – OECD TG 233).	The lack of *in vitro* JH activity is not evidence against any JH activity, due to the limited nature of current *in vitro* JH screens.
E	+	–	–	Evidence for adverse *in vivo* effects in crustaceans caused by JH and JH mimics.	As *Daphnia* are parthenogenetic, it would be desirable (if not already conducted) to perform an additional apical test with sexually reproducing crustaceans and/or insects (e.g. the Harpacticoid Copepod Development and Reproduction Test – OECD GD 201; and/or the Sediment-Water Chironomid Life Cycle Toxicity Test – OECD TG 233).	The lack of *in vitro* JH activity is not evidence against any JH activity, due to the limited nature of current *in vitro* JH screens.

Scenarios	Result of DMGT	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
F	+	–	Eq/0	Evidence for adverse *in vivo* effects in crustaceans caused by JH and JH mimics.	Given the absence or equivocal nature of existing *in vivo* data, and the fact that *Daphnia* are parthenogenetic, it might also be sensible to conduct a JH-responsive insect assay (e.g. the Sediment-Water Chironomid Life Cycle Toxicity Test – OECD TG 233) or crustacean assay (e.g. the Harpacticoid Copepod Development and Reproduction Test – OECD GD 201).	The lack of *in vitro* JH activity is not evidence against any JH activity, due to the limited nature of current *in vitro* JH screens. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
G	+	Eq/0	+	Evidence for adverse *in vivo* effects in crustaceans caused by JH and JH mimics, plus possible JH effects in other arthropods.	Given the absence or equivocal nature of existing *in vitro* data, it would be desirable to obtain further *in vitro* data on JH activity if possible. Also, as *Daphnia* are parthenogenetic, it might also be desirable to conduct a JH-responsive insect assay (e.g. the Sediment-Water Chironomid Life Cycle Toxicity Test – OECD TG 233) or crustacean assay (e.g. the Harpacticoid Copepod Development and Reproduction Test – OECD GD 201).	If a new *in vitro* mechanistic assay is conducted, note that a negative does not mean that the test material has no JH activity. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
H	+	Eq/0	–	Evidence for adverse *in vivo* effects in crustaceans caused by JH and JH mimics.	Given the absence or equivocal nature of the *in vitro* mechanistic data, it might also be helpful to conduct an *in vitro* screen for JH activity. Also, as *Daphnia* are parthenogenetic, it might also be desirable to conduct a JH-responsive insect assay (e.g. the Sediment-Water Chironomid Life Cycle Toxicity Test – OECD TG 233) or crustacean assay (e.g. the Harpacticoid Copepod Development and Reproduction Test – OECD GD 201).	If a new *in vitro* mechanistic assay is conducted, note that a negative does not mean that the test material has no JH activity. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
I	+	Eq/0	Eq/0	Evidence for adverse *in vivo* effects in crustaceans caused by JH and JH mimics.	Given the absence or equivocal nature of the *in vitro* mechanistic data, it might also be helpful to conduct an *in vitro* screen for JH activity. Also, as *Daphnia* are parthenogenetic, it might also be desirable to conduct a JH-responsive insect assay (e.g. the Sediment-Water Chironomid Life Cycle Toxicity Test – OECD TG 233) or crustacean assay (e.g. the Harpacticoid Copepod Development and Reproduction Test – OECD GD 201).	If a new *in vitro* mechanistic assay is conducted, note that a negative does not mean that the test material has no JH activity. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
J	–	+	+	The test chemical is probably a JH agonist without adverse effects in crustaceans, although it is possible that *Daphnia magna* responds atypically in this case.	Some regulatory authorities may conclude that no further evidence is required. However, it might be desirable to obtain data from insects (e.g. the Sediment-Water Chironomid Life Cycle Toxicity Test – OECD TG 233) if these are not already available.	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is a JH agonist.

Scenarios	Result of DMGT	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
K	–	+	–	The test chemical is probably a JH agonist without adverse effects in crustaceans or other taxa, although it is possible that *Daphnia magna* responds atypically in this case.	If there is no activity in crustaceans or insects, further evidence is probably not needed.	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is a JH agonist.
L	–	+	Eq/0	The test chemical is probably a JH agonist without adverse effects in crustaceans, although it is possible that *Daphnia magna* responds atypically in this case.	Some regulatory authorities may conclude that no further evidence is required, but if insect data are absent, it might be desirable to conduct a Sediment-Water Chironomid Life Cycle Toxicity Test (OECD TG 233).	Based on the limited scope of current *in vitro* screens, the positive *in vitro* data suggest that the test chemical is a JH agonist. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
M	–	–	+	The test chemical is probably without JH activity in crustaceans, although it is possible that *Daphnia magna* responds atypically in this case.	Some regulatory authorities may conclude that no further evidence is required. However, it might be desirable to obtain data from insects (e.g. the Sediment-Water Chironomid Life Cycle Toxicity Test – OECD TG 233) if these are not already available.	The lack of *in vitro* JH activity is not evidence against any JH activity, due to the limited nature of current *in vitro* JH screens.
N	–	–	–	The test chemical is probably without JH activity in arthropods.	No further action is necessary.	–
O	–	–	Eq/0	The test chemical is probably without JH activity in arthropods.	Some regulatory authorities may conclude that no further evidence is required. However, it might be desirable to obtain data from insects (e.g. the Sediment-Water Chironomid Life Cycle Toxicity Test – OECD TG 233) if these are not already available.	The lack of *in vitro* JH activity is not evidence against any JH activity, due to the limited nature of current *in vitro* JH screens. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
P	–	Eq/0	+	The test chemical is probably without JH activity in crustaceans, although it is possible that *Daphnia magna* responds atypically in this case.	Some regulatory authorities may conclude that no further evidence is required. Also, if clear *in vitro* mechanistic data are missing, it might be desirable to obtain some.	If a new *in vitro* mechanistic assay is conducted, note that a negative does not mean that the test material has no JH activity. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

Scenarios	Result of DMGT	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
Q	–	Eq/0	–	The test chemical is probably without JH activity in arthropods.	No further action is necessary.	It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
R	–	Eq/0	Eq/0	The test chemical is probably without JH activity in crustaceans and possibly insects.	Some regulatory authorities may conclude that no further evidence is required. However, it might be desirable to obtain data from insects (e.g. the Sediment-Water Chironomid Life Cycle Toxicity Test – OECD TG 233) if these are not already available.	It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

C.2.21. Fish Life Cycle Toxicity Test (FLCTT) (US EPA OPPTS 850.1500)

Status: Assay validated at national level.

590. Modality detected/endpoints: The basic FLCTT as described by Benoit (1981), US EPA (1996) and others does not contain endpoints which solely respond to endocrine disrupters. However, many of the endpoints in this apical test are nevertheless affected by estrogen/androgen/thyroid/steroidogenesis (E,A,T,S) endocrine disruptors (EDs). Of particular interest in the context of estrogens, androgens and steroidogenesis disrupters are time to sexual maturity, sex ratio of adults, fecundity and fertility, but other endpoints may also be responsive to some EDs (e.g. growth may respond to some thyroid disrupters). It should be noted that no cases are known in which altered sex ratio was caused by a substance other than an ED.

Background to the assay

591. This assay is designed primarily as an apical test for chemicals with suspected reproductive or long-term toxicity. It has not been adopted for publication as an OECD test gideline (TG), but has been widely used for several decades by regulatory agencies for assessing possible chronic effects in fish. The endpoints are all apical measures of development, growth or reproduction. Exposure of the test organisms (fathead minnow *Pimephales promelas*, in the case of Benoit [1981], but other species can be successfully used with minor changes in the protocol, including sheepshead minnow *Cyprinodon variegatus*, zebrafish *Danio rerio* and medaka *Oryzias latipes*) usually continues from the freshly fertilised eggs of the F0 generation to the fry or young fish of the F1 generation (four to eight weeks post-hatch in the case of fathead minnow [Benoit, 1981]).

592. It should be noted that it would be relatively straightforward to include ED-specific endpoints in this test. However, this is no longer necessary as a fish life cycle test that includes such endpoints, the Medaka Extended One-Generation Reproduction Test (MEOGRT), has recently been validated and published by the OECD (TG 240). A similar test using zebrafish – the Zebrafish Extended One-Generation Reproduction Test (ZEOGRT) – is currently in validation. If there is a need to examine the apical effects of a suspected ED, it would therefore be preferable to use the MEOGRT or ZEOGRT rather than the FLCTT.

When/why the assay may be used

593. As stated above, the FLCTT has essentially been superseded by the MEOGRT or ZEOGRT for the purposes of evaluating endocrine active substances (EAS), because it does not have any EAS-specific endpoints. Therefore, if new life cycle data are required in an assessment, the MEOGRT or ZEOGRT would be the assays of choice. However, if FLCTT data on adverse effects are already available when conducting a hazard assessment of an E,A,T,S chemical, they should certainly be considered in a weight of evidence

evaluation as "*in vivo* effects of concern" alongside any other relevant *in vitro* and *in vivo* data.

594. In order to provide information relevant for assessing whether or not a chemical may fulfil the WHO/IPCS (2002) definition of an ED, the study design has to be sufficiently robust to demonstrate the presence or absence of effects. In the dose selection, the investigator should also consider and ensure that data generated are adequate to fulfil the regulatory requirement across OECD countries as appropriate (e.g. hazard and risk assessment and labelling, ED assessment, etc.). The top dose or concentration should be sufficiently high to give clear systemic (i.e. non endocrine-specific) toxicity in order to ensure that a wide range of exposures (high to low) is tested. However, endocrine effects observed solely in the presence of clear systemic toxicity should be interpreted with caution and may be disregarded when sufficiently justified to be caused by secondary effects which are unlikely to be due to endocrine activity. The reason for this advice is a concern that some EAS-sensitive assays are being run at doses/concentrations of EASs that are too low to trigger direct impacts on the endocrine system. This guidance document is not the place to address this issue directly, but it should be considered when EAS-sensitive TGs are revised in the future. In addition, the number and spacing of dose/concentration levels should also be adequate to fulfil the objectives of the study (e.g. to demonstrate dose response relationships if this is required).

Existing data to be considered

595. Given the commonality of endocrine mechanisms in the vertebrates, relevant existing data available before deployment of the FLCTT for endocrine disruption hazard assessment might include *in vivo* results obtained with other vertebrates (e.g. a positive Uterotrophic Bioassay with rodents, positive findings for endocrine endpoints in mammalian repeat dose toxicity or reproductive studies), as well as information on possible modes of action (MOA) from quantitative structure activity relationships (QSARs) and/or *in vitro* screens. These will probably be accompanied by *in vivo* fish assay data from OECD TG 229, TG 230 or EASZY, and may also include data from the Fish Sexual Development Test (OECD TG 234). It would not be advisable or ethically desirable to conduct an unmodified FLCTT without mechanistic or *in vivo* screening data because it would then not be possible to link any apical effects with endocrine disruption. Furthermore, data from OECD TG 229 and/or TG 234 (FSDT) could be of use in focusing attention in the FLCTT on particularly vulnerable parts of the life cycle. Given the high ethical and financial cost of the FLCTT, it is important to make full use of existing endocrine-related data, both before the test is begun and during data evaluation.

Scenarios: Positive and negative results combined with existing data

596. The scenarios (A to R) presented in Table C.2.21 represent all the possibilities of positive or negative results in combination with the presence or absence of existing data. The action taken will also depend on the regulatory environment, but the considerations given here are generally science based. Wherever possible, the recommended "next step which could be taken" avoids unnecessary animal testing. However, sometimes conducting an animal test will be indicated and then the relevance of species, strain and exposure route should always be considered. Further considerations, specific to each scenario are given in the table.

597. Positive results obtained with one of the FLCTT endpoints result in the conclusion that the test chemical is able to cause adverse effects *in vivo* (Table C.2.21, Scenarios A-I), but not necessarily that it is an ED. Note that if doubt exists about the test performance (e.g. highly unusual results in controls), a comparison with historical control data with respect to overall test performance might be helpful. However, the nature of these effects and any existing data will require careful consideration. If *in vitro* and/or *in vivo* data already exist which reveal possible endocrine disrupting properties (Scenarios A, B and D), a positive endpoint in the FLCTT could lead to a conclusion that the test chemical is an actual ED. Such a conclusion will be strengthened considerably if the endocrine modality previously identified is plausibly linked to the responding endpoint. For example, if the chemical has estrogenic properties and reduced fecundity of the F0 adults has been observed in the FLCTT, this gives added confidence in this conclusion. On the other hand, it may be harder to argue a plausible link between estrogenic properties on the one hand, and an endpoint such as growth or survival on the other, although it is known that some estrogens are able to cause changes in growth rates (Knacker et al., 2010). In this example, an effect solely on growth or survival, while potentially of concern from the viewpoint of environmental hazard identification/characterisation, would not on its own lead to a conclusion that the chemical is an ED in fish.

598. If a plausible link of a responding FLCTT endpoint with previously identified endocrine activity can be made, regulatory authorities may conclude that sufficient evidence is available to categorise the chemical as an ED (i.e. interference with the endocrine system has caused adverse effects *in vivo*), and no further information might then be required. Of course, if the intention is to conduct an environmental hazard identification/characterisation, it may also be necessary to consider whether or not effects observed are relevant at the population level (e.g. reproduction, growth, development). On the other hand, if data from prior endocrine screens and tests are negative (Scenario E), a positive response in the FLCTT would not, in general, support the hypothesis that the chemical is an ED in fish (although it could be argued that a change in sex ratio is likely to have been caused by an ED). It could, of course, still be subjected to an environmental hazard identification/characterisation.

599. The scenarios in which the FLCTT gives a negative result (Table C.2.21, Scenarios J-R) lead to a tentative conclusion that the test chemical is not an ED in fish, and this conclusion is strengthened considerably if prior screens have failed to reveal endocrine activity (Scenario N). In the latter circumstances, regulatory authorities would be justified in concluding that no further action is needed. On the other hand, if one or more of those screens was positive (Scenarios J-M and P), the probable reasons for lack of effects in the FLCTT might be metabolism to an inactive chemical, or failure to reach the active site, and no further action would be indicated.

600. In each of the above scenarios, it is possible that existing data will be equivocal (Table C.2.21, Scenarios C, F-I, L and O-R), or there may be no existing data. This will weaken the conclusions which can be drawn about a positive FLCTT, and this is reflected in Table C.2.21. However, as indicated above, it would be undesirable to proceed with an FLCTT if prior data on endocrine activity are equivocal or absent, and if there are no other effect- or exposure-related reasons for considering such a comprehensive test. On the other hand, if the FLCTT is positive, it would be essential to obtain some reliable mechanistic data before reaching a conclusion about whether or not the chemical is an ED in fish. There is also the possibility that equivocal mechanistic data may be the result of multiple modes of endocrine action. Under some circumstances, two opposite modes of simultaneous action (e.g. estrogenic and anti-estrogenic) could, depending on dose, lead to a

minimisation or abolition of adverse effects, while in others two different MOA (e.g. estrogenic and anti-androgenic) could potentially reinforce effects on certain apical endpoints. If multiple MOA are suspected, either from the existing results or based on QSAR/read-across/integrated approaches, this situation should be investigated further if needed for regulatory decision making.

601. The scenario in which the results of the FLCTT are themselves equivocal has not been dealt with in Table C.2.21, for reasons of brevity. In this context, an equivocal result might be an inconsistent concentration-response (e.g. no effect at a high concentration but effects at a lower concentration), or a result which borders on statistical significance. Without knowing the exact circumstances, reliable advice cannot be given, but the opinions of an experienced ecotoxicologist should be sought. However, if a comprehensive set of prior screens are all negative, it is doubtful whether further action is needed, because the chemical is unlikely to be an ED. If an endocrine screen is positive, some types of equivocal FLCTT results would have to be taken more seriously. For example, a non-monotonic concentration-response would not necessarily rule out the test chemical as an ED in fish. An example of this would be a chemical like ethinylestradiol which causes adverse effects on fish reproduction at low doses, but reduced reproductive success at very high doses, thus potentially giving a U-shaped response curve. Ideally, concentrations causing systemic toxicity of this type should not be tested in an FLCTT, but such toxicity may have been missed in earlier screens.

602. In summary, positive results in the FLCTT indicate that a chemical is a probable ED if they can be plausibly linked to an endocrine MOA established on the basis of prior mechanistic screening or concurrent observation of mechanistic effects or their biochemical/physiological manifestations. If such screening data are unavailable or negative, it should not be concluded that a positive FLCTT is the result of endocrine disruption (although it is likely that biased sex ratio will be the result of ED). On the other hand, a negative FLCTT combined with a sufficiently comprehensive set of negative screening data could lead to a firm conclusion that a chemical is not an ED in fish. A negative FLCTT set against a background of a positive screen might, however, raise concerns (e.g. if the chemical is known to be involved in epigenesist).

References

Benoit, D.A. (1981), "User's guide for conducting life-cycle chronic toxicity tests with fathead minnows (*Pimephales promelas*)", EPA-600/8-81-011, Environmental Protection Agency, Washington, DC.

Knacker, T. et al. (2010), "Environmental effect assessment for sexual endocrine-disrupting chemicals: Fish testing strategy", *Integrated Environmental Assessment and Management*, Vol. 6/4, pp. 653-662, https://doi.org/10.1002/ieam.92.

US EPA (1996), "Fish Life Cycle Toxicity", US EPA Ecological Effects Test Guidelines, OPPTS 850.1500, Environmental Protection Agency, Washington, DC, https://www.epa.gov/sites/production/files/2015-07/documents/850-1500.pdf.

WHO/IPCS (2002), "Global assessment of the state-of-the-science of endocrine disrupters", Damstra, T. et al. (eds.) WHO/PCS/EDC/02.2, World Health Organization, Geneva, www.who.int/ipcs/publications/new_issues/endocrine_disruptors/en.

Table C.2.21. **Fish Life Cycle Toxicity Test (FLCTT) (US EPA OPPTS 850.1500): Guidance for scenarios of combinations of results with existing data**

This table represents possible conclusions to be drawn from assay data, and a next step which could be taken if further evidence is required about possible endocrine disrupting properties and/or effects. The guidance offered is not meant to be prescriptive, but provides science-based considerations. It encourages the use of all available data and expert judgement in a weight of evidence approach. Regional and national interpretation of results and "next steps" may vary.

The conclusions are grouped into a series of scenarios (A-R), each scenario representing a different combination of assay results, existing *in vitro* data and existing *in vivo* data. The symbol "+" indicates that the data in question represent a positive result, "'-" indicates a negative result and "Eq/0" indicates that the data are either equivocal or are not available.

Existing results: * "Mechanism (*in vitro* mechanistic data)" assumes that mechanistic data are available from estrogen receptor (ER-), androgen receptor (AR-) and steroidogenesis-based assays (Level 2). Thyroid hormone receptor (TR) and other assays concerning mechanisms of thyroid disruption may be available, but they are not in common use. In practice, data from all assays may not be available and therefore this must be taken into account when deciding on the "next step". Quantitative structure activity relationship (QSAR) predictions of estrogen and androgen binding/activation may be made for some substances. There is no evidence at present that equivalent *in vitro* assays with systems derived from fish offer advantages over their mammalian counterparts.

Existing results: ** "Effects (*in vivo* effects of concern)" assumes effects have been observed in other *in vivo* screens/tests which give rise to concern that the test chemical may be an endocrine disrupter.

Note that although this assay has been used for many years to assess the chronic effects of chemicals, no attempt has been made to validate it for use with potential endocrine disruptors, and it has not been published as an OECD test guideline.

Scenario	Result of FLCTT	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
A	+	+	+	The test chemical is probably an endocrine disruptor (ED) if the modality identified in existing screens/tests can be plausibly linked to the affected endpoint.	Further evidence is probably not required.	If the affected endpoint in the FLCTT cannot be plausibly linked to the known modality, the test chemical is unlikely to be an ED. The FLCTT is unlikely to detect epigenetic effects.
B	+	+	–	The test chemical is probably an ED in fish if the modality identified in existing screens/tests can be plausibly linked to the affected endpoint.	Further evidence is probably not required.	If the affected endpoint in the FLCTT cannot be plausibly linked to the known modality, the test chemical is unlikely to be an ED. The FLCTT is unlikely to detect epigenetic effects.
C	+	+	Eq/0**	The test chemical is probably an ED in fish if the modality identified in existing screens can be plausibly linked to the affected endpoint.	Further evidence is probably not required.	If the affected endpoint in the FLCTT cannot be plausibly linked to the known modality, the test chemical is unlikely to be an ED. The FLCTT is unlikely to detect epigenetic effects. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple modes of action (MOA). If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
D	+	–	+	The test chemical may be an ED, but the negative mechanistic data reduce the confidence in this conclusion. However, if the endocrine disruption effects in existing *in vivo* tests can be plausibly linked to the FLCTT responses, this increases the probability that the chemical is an ED.	Further evidence is probably not required.	If the affected endpoint in the FLCTT cannot be plausibly linked to the endocrine effects in existing *in vivo* tests, the test chemical is unlikely to be an ED. The FLCTT is unlikely to detect epigenetic effects.
E	+	–	–	The test chemical is unlikely to be an ED.[1]	Further evidence is probably not required.	It is possible that the effects observed in the FLCTT have been caused by an unknown endocrine mechanism. This would not, however, prevent the chemical being subjected to hazard identification/characterisation. The FLCTT is unlikely to detect epigenetic effects.
F	+	–	Eq/0	The test chemical is unlikely to be an ED, but the relevance of any equivocal existing *in vivo* data to the FLCTT results should be examined.	Further evidence is probably not required.	It is possible that the effects observed in the FLCTT have been caused by an unknown endocrine mechanism – equivocal existing *in vivo* data may throw some light on this. The absence of data on a possible endocrine mechanism would, however, not prevent the chemical being subjected to hazard identification/characterisation. The FLCTT is unlikely to detect epigenetic effects. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

Note: 1. However, note that if biased sex ratio is observed, it is likely to have been caused by an endocrine disrupting chemical.

Scenario	Result of FLCTT	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
G	+	Eq/0	+	The test chemical may be an ED, but the equivocal or absent mechanistic data reduce the confidence in this conclusion. However, if the endocrine disruption effects in existing *in vivo* tests can be plausibly linked to the FLCTT responses, this increases the probability that the chemical is an ED.	If reliable mechanistic data are not available, it would be desirable to obtain some.	The test chemical is probably an ED if a modality identified in the newly commissioned mechanistic screens, or in the existing *in vivo* data, can be plausibly linked to the affected endpoint. The FLCTT is unlikely to detect epigenetic effects. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
H	+	Eq/0	–	The test chemical may be an ED, but the equivocal or absent mechanistic data reduce the confidence in this conclusion.	If reliable mechanistic data are not available, it would be desirable to obtain some.	The test chemical is probably an ED if a modality identified in the newly commissioned mechanistic screens can be plausibly linked to the affected endpoint. The FLCTT is unlikely to detect epigenetic effects. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
I	+	Eq/0	Eq/0	The test chemical may be an ED, but the equivocal or absent mechanistic and *in vivo* data reduce the confidence in this conclusion.	If reliable mechanistic data are not available, it would be desirable to obtain some.	The test chemical is probably an ED if a modality identified in the newly commissioned mechanistic screens can be plausibly linked to the affected endpoint. The FLCTT is unlikely to detect epigenetic effects. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
J	–	+	+	The chemical is probably not an ED in fish, unless this conclusion is contradicted by existing *in vivo* data.	No further action.	–
K	–	+	–	The chemical is probably not an ED in fish.	No further action.	–
L	–	+	Eq/0	The chemical is probably not an ED in fish.	No further action.	It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information
M	–	–	+	The chemical is probably not an ED in fish.	No further action.	–
N	–	–	–	The chemical is probably not an ED.	Further evidence is probably not required.	–

Scenario	Result of FLCTT	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
O	–	–	Eq/0	The chemical is probably not an ED in fish.	Further evidence is probably not required.	It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
P	–	Eq/0	+	The chemical is probably not an ED in fish.	If reliable mechanistic data are not available, it would be desirable to obtain some.	It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
Q	–	Eq/0	–	The chemical is probably not an ED, but confidence in this conclusion is reduced by the lack of clear mechanistic data.	Further evidence is probably not required, but confidence in the conclusion would be increased by the provision of reliable negative mechanistic data.	It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
R	–	Eq/0	Eq/0	The chemical may not be an ED, but confidence in this conclusion is reduced by the lack of clear mechanistic and existing *in vivo* data.	Further evidence is probably not required, but confidence in the conclusion would be increased by the provision of reliable negative mechanistic data.	It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

C.2.22. Zebrafish Extended One-Generation Reproduction Test (ZEOGRT) (draft OECD TG 240)

Status: Assay being validated by the OECD.

603. Modality detected/endpoints: This draft fish life cycle test was specifically designed to investigate the apical effects of endocrine disrupters, and has several endpoints which can be considered diagnostic of some types of estrogen/androgen/thyroid/steroidogenesis (E,A,T,S) activity. This gives it a potential advantage over other currently standardised life cycle tests apart from the Medaka Extended One-Generation Reproduction Test (MEOGRT). If it is successfully validated, its use for evaluating endocrine disruptors (EDs) is to be preferred to the Fish Life Cycle Toxicity Test (FLCTT; see Section C.2.21) which, although sensitive to the apical effects of some EDs, contains no endocrine-sensitive endpoints. In view of the inclusion of certain ED-specific endpoints, the ZEOGRT can contribute useful evidence about the probable causality of apical effects, which is a key issue in the definition of EDs.

Background to the assay

604. This assay is a comprehensive test using zebrafish (*Danio rerio*) exposed continuously from the adult stage of the first generation (F0) to the newly hatched stage of the third generation (F2). In other words, it includes two phases of reproductive activity, and two phases of embryonic development and hatching, separated by a full phase of growth and sexual development. The test differs from the MEOGRT in that zebrafish are group spawners, whereas medaka are pair spawners. It begins with spawning groups of 5 male and 5 female sexually mature F0 fish (approximately 15 weeks post-fertilisation, or wpf) reproducing for 3 weeks, brings their F1 offspring to sexual maturity (13-20 weeks), then allows the F1 adults to breed, and finally follows their offspring (F2) to hatching (up to 14 days post-fertilisation, or dpf). The main emphasis of the assay concerns population-relevant apical endpoints (e.g. survival, development, growth, sex ratio and reproduction). However, in order to obtain mechanistic information, additional endpoints include measurements of vitellogenin (either as protein – VTG, or as mRNA coding for vitellogenin – *vtg*), phenotypic sex ratio and gonadal histopathology. Histopathology of liver and kidney may also be measured in order to distinguish between endocrine effects and possible systemic or other toxicity. Unlike the MEOGRT, the assay may be able to distinguish relatively small changes in the sex ratio of the F1 generation as it includes a large number of F1 fish (36 per replicate). On the other hand, it does not have a genetic sex endpoint, which may diminish the power to measure changes in sex ratio. It is also worth noting that because sex determination in zebrafish is polygenic (i.e. it is not driven by a single gene as in medaka), a range of environmental influences such as crowding, hypoxia and temperature fluctuations can have an influence on sex ratio. Finally, as zebrafish are not sexually dimorphic, this assay is not able to measure secondary sexual characteristics.

605. It should be noted that the ZEOGRT has only just begun validation by the OECD (Standard Project Submission Form approved by the OECD in September 2015) and has

not yet been widely used. It is expected that there will be a significant risk of test failure because of its length and difficulty. Currently, however, few testing laboratories have experience with the ZEOGRT.

606. Only zebrafish are recommended for use in this test design. The related assay using medaka (*Oryzias latipes*) – the MEOGRT – has been an OECD test guideline since 2015 (OECD TG 240).

When/why the assay may be used

607. Although the ZEOGRT could, in principle, be used at any stage in the hazard assessment process, the most likely use scenario will be when there are already some data available to suggest possible endocrine disruption properties. In other words, the ZEOGRT will generally be used to investigate whether such potential properties result in adverse apical effects on development, growth or reproduction over an entire life cycle. It is unlikely (and undesirable) that the ZEOGRT will be the first ED-responsive test procedure to be applied to a chemical. Furthermore, the conduct of a ZEOGRT **in addition to** a MEOGRT is not likely to be necessary (for example, to address perceived sensitivity differences). Before either assay is initiated, careful thought should be given to which is more appropriate in the circumstances. For example, if previous data are available with zebrafish and the ZEOGRT is sufficiently powerful for the expected endpoint of concern, then conducting a ZEOGRT may be the correct choice. However, if a genetic sex marker or secondary sexual characters are desired, it may be more beneficial to consider a MEOGRT.

608. This is a comprehensive test which examines a range of potentially adverse apical effects, but also considers several ED-specific endpoints. It is therefore suitable for helping to define whether a test chemical is an ED, and the results could be used in an environmental risk assessment for fish. Given the high degree of endocrine system conservation across the vertebrates, adverse endocrine-linked effects in the ZEOGRT may also indicate the possibility of related activity in other organisms such as amphibians, reptiles, birds or mammals.

609. In order to provide information relevant for assessing whether or not a chemical may fulfil the WHO/IPCS (2002) definition of an ED, the study design has to be sufficiently robust to demonstrate the presence or absence of effects. In the dose selection, the investigator should also consider and ensure that data generated are adequate to fulfil the regulatory requirement across OECD countries as appropriate (e.g. hazard and risk assessment and labelling, ED assessment, etc.). The top dose or concentration should be sufficiently high to give clear systemic (i.e. non endocrine-specific) toxicity in order to ensure that a wide range of exposures (high to low) is tested. However, endocrine effects observed solely in the presence of clear systemic toxicity should be interpreted with caution and may be disregarded when sufficiently justified to be caused by secondary effects which are unlikely to be due to endocrine activity. The reason for this advice is a concern that some endocrine active substance (EAS) sensitive assays are being run at doses/concentrations of EASs that are too low to trigger direct impacts on the endocrine system. This guidance document is not the place to address this issue directly, but it should be considered when EAS-sensitive test guidelines (TGs) are revised in the future. In addition, the number and spacing of dose/concentration levels should also be adequate to fulfil the objectives of the study (e.g. to demonstrate dose response relationships if this is required).

Existing data to be considered

610. Existing data available before deployment of the ZEOGRT for endocrine disruption hazard assessment are likely to include information on possible modes of action (MOA) from quantitative structure activity relationships (QSARs), adverse outcome pathways and/or *in vitro* screens. These may be accompanied by *in vivo* fish assay data from EASZY, the Juvenile Medaka Anti-Androgen Screening Assay, OECD TG 229 and/or OECD TG 230, and may also include data from the Fish Sexual Development Test (FSDT – TG 234). In addition, existing information on endocrine-related effects from other vertebrates (up to and including mammals, e.g. positive findings for endocrine endpoints in mammalian repeat dose toxicity or reproductive studies) should also be considered, given the commonality of endocrine mechanisms in these taxa. It would not be advisable or ethically desirable to conduct a ZEOGRT without mechanistic or *in vivo* screening data because it would then be less straightforward to link any apical effects with endocrine disruption. Furthermore, data from OECD TG 229 and/or TG 234 (FSDT), especially if obtained with zebrafish, could be of use in focusing attention in the ZEOGRT on particularly vulnerable parts of the life cycle. Given the high ethical and financial cost of the ZEOGRT, it is important to make full use of existing endocrine-related data, both before the test is begun and during data evaluation.

Scenarios: Positive and negative results combined with existing data

611. The advice given for the following scenarios is largely based on experience gained with the MEOGRT, and so should be treated with caution.

612. The scenarios (A to R) presented in Table C.2.22 represent all the possibilities of positive or negative results in combination with the presence or absence of existing data. The action taken will also depend on the regulatory environment, but the considerations given here are generally science based. Wherever possible, the recommended "next step which could be taken" avoids unnecessary animal testing. However, sometimes conducting an animal test will be indicated and then the relevance of species, strain and exposure route should always be considered. Further considerations, specific to each scenario are given in the table.

613. Positive results obtained with one of the ZEOGRT apical endpoints result in the conclusion that the test chemical is able to cause adverse effects *in vivo* (Table C.2.22, Scenarios A-I), but not necessarily that it is an ED. Note that if doubt exists about the test performance (e.g. highly unusual results in controls), a comparison with historical control data with respect to overall test performance might be helpful. However, the nature of these effects and any existing data will require careful consideration. If *in vitro* and/or *in vivo* data already exist which reveal possible endocrine disrupting properties (Scenarios A, B and D), a positive apical endpoint in the ZEOGRT could lead to a conclusion that the test chemical is an actual ED if adverse population effects are expected as a consequence. This conclusion will, of course, be reinforced if mechanistic endpoints in the ZEOGRT itself also respond. The probability that the test chemical is an ED will also be strengthened considerably if the endocrine modality identified in the present or earlier tests is plausibly linked to the responding endpoint. For example, if the chemical has estrogenic properties (such as the induction of VTG in males) and there is observed to be reduced fecundity of the F0 or F1 adults in the ZEOGRT, this gives added confidence in this conclusion. On the other hand, it may be harder to argue a plausible link between estrogenic properties on the one hand, and an endpoint such as growth or survival on the other, although it is known that some estrogens are able to cause changes in growth rates (Knacker et al., 2010). In this example, an effect

solely on growth or survival, while potentially of concern from the viewpoint of environmental hazard identification/characterisation, would not on its own lead to a conclusion that the chemical is an ED in fish.

614. If a plausible link of a responding ZEOGRT apical endpoint with identified endocrine activity can be made, regulatory authorities may conclude that sufficient evidence is available to categorise the chemical as an ED (i.e. interference with the endocrine system has caused adverse effects *in vivo*), and no further information might then be required. It may also be necessary to consider whether or not effects observed are relevant at the population level (e.g. reproduction, growth, development). On the other hand, if data from prior endocrine screens and tests are negative, including negative mechanistic data from the ZEOGRT itself (Scenario E), a positive apical response in the ZEOGRT would not in general support the hypothesis that the chemical is an ED in fish (although a change in sex ratio may have been caused by an ED). The chemical could, of course, still be subjected to an environmental hazard identification/characterisation.

615. The scenarios in which the ZEOGRT gives a negative apical result (Table C.2.22, Scenarios J-R) lead to a tentative conclusion that the test chemical is not an ED in fish, and this conclusion is strengthened considerably if prior screens, or the ZEOGRT itself, have failed to reveal endocrine activity (Scenario N). In the latter circumstances, regulatory authorities would be justified in concluding that no further action is needed. On the other hand, if one or more of those screens was positive (Scenarios J-M and P), the bioconcentration factor (BCF) of the chemical should be checked. If the BCF indicates that the chemical is strongly bioaccumulative and equilibrium is reached slowly, it would be worth considering the conduct of an extended ZEOGRT (but no TG is available for this), although as indicated above, there is little evidence at present that EDs with a high BCF would be consistently more potent in such a test. If a chemical which screened positive is not bioaccumulative, the probable reasons for lack of effects in the ZEOGRT might be metabolism to an inactive chemical, or failure to reach the active site, and no further action would be indicated.

616. In each of the above scenarios, it is possible that existing data will be equivocal (Table C.2.22, Scenarios C, F-I, L and O-R), or there may be no existing data. This will weaken the conclusions which can be drawn about a positive apical endpoint in the ZEOGRT, and this is reflected in Table C.2.22. However, as indicated above, it would be undesirable to proceed with a ZEOGRT if prior data on endocrine activity are equivocal or absent, and if there are no other effect- or exposure-related reasons for considering such a comprehensive test. On the other hand, if the ZEOGRT shows a positive apical endpoint, it would be essential to obtain some reliable mechanistic data before reaching a conclusion about whether or not the chemical is an ED in fish. There is also the possibility that equivocal mechanistic data may be the result of multiple modes of endocrine action. Under some circumstances, two opposite modes of simultaneous action (e.g. estrogenic and anti-estrogenic) could, depending on dose, lead to a minimisation or abolition of adverse effects, while in others two different MOA (e.g. estrogenic and anti-androgenic) could potentially reinforce effects on certain apical endpoints. If multiple MOA are suspected, either from the existing results or based on QSAR/read-across/integrated approaches, this situation should be investigated further if needed for regulatory decision making.

617. The scenario in which the results of the ZEOGRT are themselves equivocal has not been dealt with in Table C.2.22, for reasons of brevity. In this context, an equivocal result might be an inconsistent concentration-response (e.g. no effect at a high concentration but effects at a lower concentration), or a result which borders on statistical significance.

Without knowing the exact circumstances, reliable advice cannot be given, but the opinions of an experienced ecotoxicologist should be sought. However, if a comprehensive set of prior screens are all negative, it is doubtful whether further action is needed, because the chemical is unlikely to be an ED. If an endocrine screen is positive, some types of equivocal ZEOGRT apical results would have to be taken more seriously. For example, a non-monotonic concentration-response would not necessarily rule out the test chemical as an ED in fish. An example of this would be a chemical like ethinylestradiol which causes adverse effects (elevated fecundity) on fish reproduction at low doses, but reduced reproductive success at very high doses, thus potentially giving an inverted U-shaped response curve (e.g. Jobling et al. [2004]). Ideally, concentrations causing systemic toxicity of this type should not be tested in ZEOGRT, but such toxicity may have been missed in earlier screens.

618. In summary, positive apical results in the ZEOGRT indicate that a chemical is a probable ED if they can be plausibly linked to an endocrine MOA established on the basis of prior mechanistic screening or concurrent observation of mechanistic effects or their biochemical/physiological manifestations. If such screening data are unavailable or negative, it should not be concluded that a positive ZEOGRT is the result of endocrine disruption (although a biased sex ratio may have been the result of ED). On the other hand, a negative ZEOGRT combined with a sufficiently comprehensive set of negative screening data could lead to a firm conclusion that a chemical is not an ED in fish. A negative ZEOGRT set against a background of a positive screen might, however, raise concerns (e.g. if the chemical is strongly bioaccumulative or known to be involved in epigenesist). In this case, an extended ZEOGRT should be considered, although this is not expected to be covered by a ZEOGRT test guideline.

References

Jobling, S. et al. (2004), "Comparative responses of molluscs and fish to environmental estrogens and an estrogenic effluent", *Aquatic Toxicology*, Vol. 66/2, pp. 207-222, https://doi.org/10.1016/j.aquatox.2004.01.002.

Knacker, T. et al. (2010), "Environmental effect assessment for sexual endocrine-disrupting chemicals: Fish testing strategy", *Integrated Environmental Assessment and Management*, Vol. 6/4, pp. 653-662, https://doi.org/10.1002/ieam.92.

WHO/IPCS (2002), "Global assessment of the state-of-the-science of endocrine disrupters", Damstra, T. et al. (eds.) WHO/PCS/EDC/02.2, World Health Organization, Geneva, www.who.int/ipcs/publications/new_issues/endocrine_disruptors/en.

Table C.2.22. **Zebrafish Extended One-Generation Reproduction Test (ZEOGRT) (draft OECD TG): Guidance for scenarios of combinations of results with existing data**

It should be noted that this assay has not yet been validated, so the advice given in the table is provisional and may change. The guidance offered is not meant to be prescriptive, but provides science-based considerations. It encourages the use of all available data and expert judgement in a weight of evidence approach. Regional and national interpretation of results and "next steps" may vary.

This table represents possible conclusions to be drawn from assay data, and a next step which could be taken if further evidence is required about possible endocrine disrupting properties and/or effects. The conclusions are grouped into a series of scenarios (A-R), each scenario representing a different combination of assay results, existing mechanistic data and existing *in vivo* effects data. The symbol "+" indicates that the data in question represent a positive result, «-" indicates a negative result, and "Eq/0" indicates that the data are either equivocal or are not available.

Results of the ZEOGRT: * Apical results of the ZEOGRT include effects on survival, growth, development and reproduction. The other ZEOGRT endpoints, including VTG, sex ratio and gonadal histopathology, can be indicative of endocrine mechanisms which may have caused the apical effect.

Existing results: ** "Mechanism (*in vitro* and/or *in vivo* mechanistic data)" assumes that mechanistic data are available from estrogen receptor (ER-), androgen receptor (AR-) and steroidogenesis-based assays (Level 2). Thyroid hormone receptor (TR) and other assays concerning mechanisms of thyroid disruption may also be available. In practice, data from all assays may not be available and therefore this must be taken into account when deciding on the "next step". Quantitative structure activity relationship (QSAR) predictions of estrogen and androgen binding/activation may be made for some substances. There is no evidence at present that equivalent *in vitro* assays with systems derived from fish offer advantages over their mammalian counterparts.

Existing results: *** "Effects (*in vivo* effects of concern)" assumes effects have been observed in other *in vivo* screens/tests which give rise to concern that the test chemical may be an endocrine disrupter.

Scenario	Apical result of ZEOGRT*	Existing results		Possible conclusions: 1) Indicators of endocrine activity and apical endpoints positive 2) Indicators of endocrine activity positive and apical endpoints negative 3) Indicators of endocrine activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* and/or *in vivo* mechanistic data)**	Effects (*in vivo* effects of concern)***			
A	+	+	+	1) Strong evidence for adverse effects in fish and other organisms by an endocrine mechanism. 2) Strong evidence for endocrine effects, but they do not appear adverse in fish. 3) Strong evidence for adverse effects in fish and other organisms. There is a possibility that the apical endpoint sex ratio is more sensitive to the test chemical than the mechanistic endpoint VTG, or mechanism may hypothetically not be via direct interaction with estrogen receptor (ER), androgen receptor (AR) or by aromatase inhibition, even though it is noted that currently there is no evidence for sex ratio change in fish caused by other mechanisms than those mentioned here at otherwise non-toxic concentrations of chemicals.	Further evidence is probably not required.	If the affected apical endpoint in the ZEOGRT cannot be plausibly linked to the known modality, the test chemical is unlikely to be an endocrine disruptor (ED). The ZEOGRT is unlikely to detect epigenetic effects. If these are suspected, extending the test beyond F2 hatching could be considered, although this would depart from the expected test guideline (TG).
B	+	+	–	1) Strong evidence for adverse effects in fish by an endocrine mechanism. 2) Strong evidence for endocrine effects in fish, but they do not appear adverse. 3) Strong evidence for adverse effects in fish and other organisms. There is a possibility that the apical endpoint sex ratio is more sensitive to the test chemical than the mechanistic endpoint VTG, or mechanism may hypothetically not be via direct interaction with ER, AR or by aromatase inhibition, even though it is noted that currently there is no evidence for sex ratio change in fish caused by other mechanisms than those mentioned here at otherwise non-toxic concentrations of chemicals.	Further evidence is probably not required.	If the affected apical endpoint in the ZEOGRT cannot be plausibly linked to the known modality, the test chemical is unlikely to be an ED. The ZEOGRT is unlikely to detect epigenetic effects. If these are suspected, extending the test beyond F2 hatching could be considered, although this would depart from the expected TG.
C	+	+	Eq/0	1) Strong evidence for adverse effects in fish by an endocrine mechanism. 2) Strong evidence for endocrine effects in fish, but they do not appear adverse. 3) Strong evidence for adverse effects in fish and other organisms. There is a possibility that the apical endpoint sex ratio is more sensitive to the test chemical than the mechanistic endpoint VTG, or mechanism may hypothetically not be via direct interaction with ER, AR or by aromatase inhibition, even though it is noted that currently there is no evidence for sex ratio change in fish caused by other mechanisms than those mentioned here at otherwise non-toxic concentrations of chemicals.	Further evidence is probably not required.	If the affected apical endpoint in the ZEOGRT cannot be plausibly linked to the known modality, the test chemical is unlikely to be an ED. The ZEOGRT is unlikely to detect epigenetic effects. If these are suspected, extending the test beyond F2 hatching could be considered, although this would depart from the expected TG. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple modes of action (MOA). If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

Scenario	Apical result of ZEOGRT*	Existing results		Possible conclusions: 1) Indicators of endocrine activity and apical endpoints positive 2) Indicators of endocrine activity positive and apical endpoints negative 3) Indicators of endocrine activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* and/or *in vivo* mechanistic data)**	Effects (*in vivo* effects of concern)***			
D	+	–	+	1) Strong evidence for adverse effects in fish and other organisms, possibly by an unknown endocrine mechanism. 2) Medium-strong evidence for endocrine effects in fish, but they do not appear to be adverse. 3) Strong evidence for adverse effects in fish and other organisms. There is a possibility that the apical endpoint sex ratio is more sensitive to the test chemical than the mechanistic endpoint VTG, or mechanism may hypothetically not be via direct interaction with ER, AR or by aromatase inhibition, even though it is noted that currently there is no evidence for sex ratio change in fish caused by other mechanisms than those mentioned here at otherwise non-toxic concentrations of chemicals.	Further evidence is probably not required.	If the affected apical endpoint in the ZEOGRT cannot be plausibly linked to the known modality, the test chemical is unlikely to be an ED. The ZEOGRT is unlikely to detect epigenetic effects. If these are suspected, extending the test beyond F2 hatching could be considered, although this would depart from the expected TG.
E	+	–	–	1) Strong evidence for adverse effects in fish, possibly by an unknown endocrine mechanism. 2) Medium-strong evidence for endocrine effects in fish, but they do not appear to be adverse. 3) Strong evidence for adverse effects in fish and other organisms. There is a possibility that the apical endpoint sex ratio is more sensitive to the test chemical than the mechanistic endpoint VTG, or mechanism may hypothetically not be via direct interaction with ER, AR or by aromatase inhibition, even though it is noted that currently there is no evidence for sex ratio change in fish caused by other mechanisms than those mentioned here at otherwise non-toxic concentrations of chemicals.	Further evidence is probably not required.	It is possible that the effects observed in the ZEOGRT have been caused by an unknown endocrine mechanism. This would not, however, prevent the chemical being subjected to hazard identification/characterisation. The ZEOGRT is unlikely to detect epigenetic effects. If these are suspected, extending the test beyond F2 hatching could be considered, although this would depart from the expected TG.

Scenario	Apical result of ZEOGRT*	Existing results		Possible conclusions: 1) Indicators of endocrine activity and apical endpoints positive 2) Indicators of endocrine activity positive and apical endpoints negative 3) Indicators of endocrine activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* and/or *in vivo* mechanistic data)**	Effects (*in vivo* effects of concern)***			
F	+	–	Eq/0	1) Moderate evidence for adverse effects in fish, possibly by an unknown endocrine mechanism. 2) Medium-strong evidence for endocrine effects in fish, but they do not appear to be adverse. 3) Strong evidence for adverse effects in fish and other organisms. There is a possibility that the apical endpoint sex ratio is more sensitive to the test chemical than the mechanistic endpoint VTG, or mechanism may hypothetically not be via direct interaction with ER, AR or by aromatase inhibition, even though it is noted that currently there is no evidence for sex ratio change in fish caused by other mechanisms than those mentioned here at otherwise non-toxic concentrations of chemicals.	Further evidence is probably not required.	It is possible that the effects observed in the ZEOGRT have been caused by an unknown endocrine mechanism or not by an endocrine mechanism at all – equivocal existing *in vivo* data may throw some light on this. The absence of data on a possible endocrine mechanism would, however, not prevent the chemical being subjected to hazard identification/characterisation. The ZEOGRT is unlikely to detect epigenetic effects. If these are suspected, extending the test beyond F2 hatching could be considered, although this would depart from the expected TG. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
G	+	Eq/0	+	1) Strong evidence for adverse effects in more than one organism, possibly by an unknown endocrine mechanism. 2) Medium-strong evidence for endocrine effects, but they do not appear to be adverse in fish. 3) Strong evidence for adverse effects in fish and other organisms. There is a possibility that the apical endpoint sex ratio is more sensitive to the test chemical than the mechanistic endpoint VTG, or mechanism may hypothetically not be via direct interaction with ER, AR or by aromatase inhibition, even though it is noted that currently there is no evidence for sex ratio change in fish caused by other mechanisms than those mentioned here at otherwise non-toxic concentrations of chemicals.	If reliable mechanistic data are not available, it would be desirable to obtain some.	The test chemical is probably an ED if a modality identified in the newly commissioned mechanistic screens, or in the existing *in vivo* data, can be plausibly linked to the affected endpoint. The ZEOGRT is unlikely to detect epigenetic effects. If these are suspected, extending the test beyond F2 hatching could be considered, although this would depart from the expected TG. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

Scenario	Apical result of ZEOGRT*	Existing results		Possible conclusions: 1) Indicators of endocrine activity and apical endpoints positive 2) Indicators of endocrine activity positive and apical endpoints negative 3) Indicators of endocrine activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* and/or *in vivo* mechanistic data)**	Effects (*in vivo* effects of concern)***			
H	+	Eq/0	–	1) Strong evidence for adverse effects in fish, possibly by an unknown endocrine mechanism. 2) Medium-strong evidence for endocrine effects in fish, but they do not appear to be adverse. 3) Strong evidence for adverse effects in fish and other organisms. There is a possibility that the apical endpoint sex ratio is more sensitive to the test chemical than the mechanistic endpoint VTG, or mechanism may hypothetically not be via direct interaction with ER, AR or by aromatase inhibition, even though it is noted that currently there is no evidence for sex ratio change in fish caused by other mechanisms than those mentioned here at otherwise non-toxic concentrations of chemicals.	If reliable mechanistic data are not available, it would be desirable to obtain some.	The test chemical is probably an ED if a modality identified in the newly commissioned mechanistic screens can be plausibly linked to the affected endpoint. The ZEOGRT is unlikely to detect epigenetic effects. If these are suspected, extending the test beyond F2 hatching could be considered, although this would depart from the expected TG. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
I	+	Eq/0	Eq/0	1) Strong evidence for adverse effects in fish, possibly by an unknown endocrine mechanism. 2) Moderate-strong evidence for endocrine effects in fish, but they do not appear to be adverse. 3) Strong evidence for adverse effects in fish and other organisms. There is a possibility that the apical endpoint sex ratio is more sensitive to the test chemical than the mechanistic endpoint VTG, or mechanism may hypothetically not be via direct interaction with ER, AR or by aromatase inhibition, even though it is noted that currently there is no evidence for sex ratio change in fish caused by other mechanisms than those mentioned here at otherwise non-toxic concentrations of chemicals.	If reliable mechanistic data are not available, it would be desirable to obtain some.	The test chemical is probably an ED if a modality identified in the newly commissioned mechanistic screens can be plausibly linked to the affected endpoint. The ZEOGRT is unlikely to detect epigenetic effects. If these are suspected, extending the test beyond F2 hatching could be considered, although this would depart from the expected TG. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
J	–	+	+	The chemical is probably not an ED in fish, unless this conclusion is contradicted by existing *in vivo* data.	If the chemical is strongly bioaccumulative, or if epigenetic effects are suspected, an extended version of the ZEOGRT could be considered, although this would depart from the expected TG.	If any effects in an extended ZEOGRT can be plausibly linked with mechanistic data, the test chemical is probably an ED.

Scenario	Apical result of ZEOGRT*	Existing results		Possible conclusions: 1) Indicators of endocrine activity and apical endpoints positive 2) Indicators of endocrine activity positive and apical endpoints negative 3) Indicators of endocrine activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* and/or *in vivo* mechanistic data)**	Effects (*in vivo* effects of concern)***			
K	–	+	–	The chemical is probably not an ED in fish.	If the chemical is strongly bioaccumulative, or if epigenetic effects are suspected, an extended version of the ZEOGRT could be considered, although this would depart from the expected TG.	If any effects in an extended ZEOGRT can be plausibly linked with mechanistic data, the test chemical is probably an ED.
L	–	+	Eq/0	The chemical is probably not an ED in fish.	If the chemical is strongly bioaccumulative, or if epigenetic effects are suspected, an extended version of the ZEOGRT could be considered, although this would depart from the expected TG.	If any effects in an extended ZEOGRT can be plausibly linked with mechanistic data, the test chemical is probably an ED. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
M	–	–	+	The chemical is probably not an ED in fish.	If the chemical is strongly bioaccumulative, or if epigenetic effects are suspected, an extended version of the ZEOGRT could be considered, although this would depart from the expected TG.	If any effects in an extended ZEOGRT can be plausibly linked with *in vivo* data which provide information on ED properties, the test chemical is probably an ED, but likely not by a mechanism covered by the existing *in vitro* screens.
N	–	–	–	The chemical is probably not an ED.	Further evidence is probably not required.	–
O	–	–	Eq/0	The chemical is probably not an ED in fish.	Further evidence is probably not required.	It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to any mechanistic information.

Scenario	Apical result of ZEOGRT*	Existing results		Possible conclusions: 1) Indicators of endocrine activity and apical endpoints positive 2) Indicators of endocrine activity positive and apical endpoints negative 3) Indicators of endocrine activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* and/or *in vivo* mechanistic data)**	Effects (*in vivo* effects of concern)***			
P	–	Eq/0	+	The chemical is probably not an ED in fish.	If reliable mechanistic data are not available, it would be desirable to obtain some.	If the newly commissioned mechanistic data are positive and the chemical is strongly bioaccumulative, or if epigenetic effects are suspected, consider conducting an extended ZEOGRT, although this would depart from the expected TG. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
Q	–	Eq/0	–	The chemical is probably not an ED, but confidence in this conclusion is reduced by the lack of clear mechanistic data.	Further evidence is probably not required, but confidence in the conclusion would be increased by the provision of reliable negative mechanistic data.	If the newly commissioned mechanistic data are positive and the chemical is strongly bioaccumulative, or if epigenetic effects are suspected, consider conducting an extended ZEOGRT, although this would depart from the expected TG. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
R	–	Eq/0	Eq/0	The chemical may not be an ED, but confidence in this conclusion is reduced by the lack of clear mechanistic and existing *in vivo* data.	Further evidence is probably not required, but confidence in the conclusion would be increased by the provision of reliable negative mechanistic data.	If the newly commissioned mechanistic data are positive and the chemical is strongly bioaccumulative, or if epigenetic effects are suspected, consider conducting an extended ZEOGRT, although this would depart from the expected TG. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

C.2.23. Avian Two-Generation Toxicity Test in the Japanese Quail (ATGT) (US EPA OCSPP 890.2100/740-C-15-003)

Status: Assay validated at national level.

619. Modality detected/endpoints: This avian multigeneration test was specifically designed to investigate the apical effects of endocrine disrupters, and has several endpoints which can be considered diagnostic of some types of estrogen/androgen/thyroid/steroidogenesis (E,A,T,S) activity. In view of the inclusion of certain endocrine disruptor (ED-) specific endpoints, the ATGT can contribute useful evidence about the probable causality of apical effects, which is a key issue in the definition of EDs.

Background to the assay

620. The assay is a comprehensive test using Japanese quail (*Coturnix japonica*). The F0 generation are exposed to a range of test chemical concentrations in their food for 49 days from 28 days post-hatch (dph). The F1 generation is exposed via the egg and orally from hatch to 70 dph. The F2 generation is only exposed via the egg, not via the food, and is followed to 14 dph (although there is an option to continue the test to F2 sexual maturity at 42 dph). The complete test therefore takes a minimum of 19 weeks.

621. A large range of endpoints is measured, including growth; development; reproduction; histopathology of multiple organs including gonads; phenotypic and genotypic sex; and various hormone titres including thyroid hormone (T4), estradiol (E2) and testosterone (T). Gonadal histopathology, hormone titres and sex ratio can all be used to provide information about possible endocrine modes of action (MOA).

622. It should be noted that the ATGT is a relatively new test (adopted by the United States Environmental Protection Agency in 2015) which has not yet been widely used and has not been validated by the OECD because it is time-consuming and technically challenging, and requires considerable resources. There is a significant risk of test failure because of its length and difficulty. Currently, few testing laboratories have experience with the ATGT.

When/why the assay may be used

623. Although the ATGT could, in principle, be used at any stage in the hazard assessment process, the most likely use scenario will be when there are already some data available to suggest possible endocrine disruption properties *in vitro* and/or *in vivo*. In other words, the ATGT will generally be used to investigate whether such potential properties result in adverse apical effects on development, growth or reproduction over two generations. It is unlikely (and undesirable) that the ATGT will be the first ED-responsive test procedure to be applied to a chemical.

624. This is a comprehensive test which examines a range of potentially adverse apical effects, but also considers several ED-specific endpoints. It is therefore suitable for helping to define whether a test chemical is an ED, and the results could be used in an

environmental hazard identification/characterisation for birds. Given the high degree of endocrine system conservation across the vertebrates, adverse endocrine-linked effects in the ATGT may also indicate the possibility of related activity in other organisms such as fish, amphibians, reptiles or mammals.

625. In order to provide information relevant for assessing whether or not a chemical may fulfil the WHO/IPCS (2002) definition of an ED, the study design has to be sufficiently robust to demonstrate the presence or absence of effects. In the dose selection, the investigator should also consider and ensure that data generated are adequate to fulfil the regulatory requirement across OECD countries as appropriate (e.g. hazard and risk assessment and labelling, ED assessment, etc.). The top dose or concentration should be sufficiently high to give clear systemic (i.e. non endocrine-specific) toxicity in order to ensure that a wide range of exposures (high to low) is tested. However, endocrine effects observed solely in the presence of clear systemic toxicity should be interpreted with caution and may be disregarded when sufficiently justified to be caused by secondary effects which are unlikely to be due to endocrine activity. The reason for this advice is a concern that some endocrine active substance (EAS) sensitive assays are being run at doses/concentrations of EASs that are too low to trigger direct impacts on the endocrine system. This guidance document is not the place to address this issue directly, but it should be considered when EAS-sensitive test guidelines (TGs) are revised in the future. In addition, the number and spacing of dose/concentration levels should also be adequate to fulfil the objectives of the study (e.g. to demonstrate dose response relationships if this is required).

Existing data to be considered

626. Existing data available before deployment of the ATGT for endocrine disruption hazard assessment are likely to include information on possible MOA from quantitative structure activity relationships (QSARs), adverse outcome pathways and/or *in vitro* screens. These may be accompanied by *in vivo* bird assay data from the Avian Reproduction Test (OECD TG 206). In addition, existing information on endocrine-related effects from other vertebrates (up to and including mammals, e.g. positive findings for endocrine endpoints in mammalian repeat dose toxicity or reproductive studies) should also be considered, given the commonality of endocrine mechanisms in these taxa. It would not be advisable or ethically desirable to conduct an ATGT without mechanistic or *in vivo* screening data because it would then be less straightforward to link any apical effects with endocrine disruption. Furthermore, data from OECD TG 206 could be of use in focusing attention in the ATGT on particularly vulnerable parts of the life cycle. Given the high ethical and financial cost of the ATGT, it is important to make full use of existing endocrine-related data, both before the test is begun and during data evaluation.

Scenarios: Positive and negative results combined with existing data

627. The scenarios (A to R) presented in Table C.2.23 represent all the possibilities of positive or negative results in combination with the presence or absence of existing data. The action taken will also depend on the regulatory environment, but the considerations given here are generally science based. Wherever possible, the recommended "next step which could be taken" avoids unnecessary animal testing. However, sometimes conducting an animal test will be indicated and then the relevance of species, strain and exposure route should always be considered. Further considerations, specific to each scenario are given in the table.

628. Positive results obtained with one of the ATGT apical endpoints result in the conclusion that the test chemical is able to cause adverse effects *in vivo* (Table C.2.23, Scenarios A-I), but not necessarily that it is an ED. Note that if doubt exists about the test performance (e.g. highly unusual results in controls), a comparison with historical control data with respect to overall test performance might be helpful. However, the nature of these effects and any existing data will require careful consideration. If *in vitro* and/or *in vivo* data already exist which reveal possible endocrine disrupting properties (Scenarios A, B and D), a positive apical endpoint in the ATGT could lead to a conclusion that the test chemical is an actual ED if adverse population effects are expected as a consequence. This conclusion will, of course, be reinforced if mechanistic endpoints in the ATGT itself also respond. The probability that the test chemical is an ED will also be strengthened considerably if the endocrine modality identified in the present or earlier tests is plausibly linked to the responding endpoint. For example, if the chemical has estrogenic properties and there is observed to be reduced fecundity of the F0 or F1 adults in the ATGT, this gives added confidence in this conclusion. On the other hand, it may be harder to argue a plausible link between estrogenic properties on the one hand, and an endpoint such as growth or survival on the other, although it is known that some estrogens are able to cause changes in growth rates. In this example, an effect solely on growth or survival, while potentially of concern from the viewpoint of environmental hazard identification/characterisation, would not on its own lead to a conclusion that the chemical is an ED in birds.

629. If a plausible link of a responding ATGT apical endpoint with identified endocrine activity can be made, regulatory authorities may conclude that sufficient evidence is available to categorise the chemical as an ED (i.e. interference with the endocrine system has caused adverse effects *in vivo*), and no further information might then be required. Of course, if the intention is to conduct an environmental hazard identification/characterisation, it may also be necessary to consider whether or not effects observed are relevant at the population level (e.g. reproduction, growth, development). On the other hand, if data from prior endocrine screens and tests are negative, including negative mechanistic data from the ATGT itself (Scenario E), a positive apical response in the ATGT would not, in general, support the hypothesis that the chemical is an ED in birds (although it could be argued that a change in sex ratio is likely to have been caused by an ED). The chemical could, of course, still be subjected to an environmental hazard identification/characterisation.

630. The scenarios in which the ATGT gives a negative apical result (Table C.2.23, Scenarios J-R) lead to a tentative conclusion that the test chemical is not an ED in birds, and this conclusion is strengthened considerably if prior screens, or the ATGT itself, have failed to reveal endocrine activity (Scenario N). In the latter circumstances, regulatory authorities would be justified in concluding that no further action is needed. On the other hand, if one or more of those screens was positive (Scenarios J-M and P), the test chemical may simply be inactive in *Coturnix japonica*. If a chemical screened positive, the probable reasons for lack of effects in the ATGT might be metabolism to an inactive chemical, or failure to reach the active site, and no further action would be indicated.

631. In each of the above scenarios, it is possible that existing data will be equivocal (Table C.2.23, Scenarios C, F-I, L and O-R), or there may be no existing data. This will weaken the conclusions which can be drawn about a positive apical endpoint in the ATGT, and this is reflected in Table C.2.23. However, as indicated above, it would be undesirable to proceed with an ATGT if prior data on endocrine activity are equivocal or absent, and if there are no other effect- or exposure-related reasons for considering such a comprehensive test. On the other hand, if the ATGT shows a positive apical endpoint, it would be essential

to obtain some reliable mechanistic data before reaching a conclusion about whether or not the chemical is an ED in birds. There is also the possibility that equivocal mechanistic data may be the result of multiple modes of endocrine action. Under some circumstances, two opposite modes of simultaneous action (e.g. estrogenic and anti-estrogenic) could, depending on dose, lead to a minimisation or abolition of adverse effects, while in others two different MOA (e.g. estrogenic and anti-androgenic) could potentially reinforce effects on certain apical endpoints. If multiple MOA are suspected, either from the existing results or based on QSAR/read-across/integrated approaches, this situation should be investigated further if needed for regulatory decision making.

632. The scenario in which the results of the ATGT are themselves equivocal has not been dealt with in Table C.2.23, for reasons of brevity. In this context, an equivocal result might be an inconsistent concentration-response (e.g. no effect at a high concentration but effects at a lower concentration), or a result which borders on statistical significance. Without knowing the exact circumstances, reliable advice cannot be given, but the opinions of an experienced ecotoxicologist should be sought. However, if a comprehensive set of prior screens are all negative, it is doubtful whether further action is needed, because the chemical is unlikely to be an ED. If an endocrine screen is positive, some types of equivocal ATGT apical results would have to be taken more seriously. For example, a non-monotonic concentration-response would not necessarily rule out the test chemical as an ED in birds. Ideally, concentrations causing systemic toxicity should not be tested in the ATGT, but such toxicity may have been missed in earlier screens.

633. In summary, positive apical results in the ATGT indicate that a chemical is a probable ED if they can be plausibly linked to an endocrine MOA established on the basis of prior mechanistic screening or concurrent observation of mechanistic effects or their biochemical/physiological manifestations. If such screening data are unavailable or negative, it should not be concluded that a positive ATGT is the result of endocrine disruption (although it is likely that biased sex ratio will be the result of ED). On the other hand, a negative ATGT combined with a sufficiently comprehensive set of negative screening data could lead to a firm conclusion that a chemical is not an ED in birds. A negative ATGT set against a background of a positive screen might, however, raise concerns.

Reference

WHO/IPCS (2002), "Global assessment of the state-of-the-science of endocrine disrupters", Damstra, T. et al. (eds.) WHO/PCS/EDC/02.2, World Health Organization, Geneva, www.who.int/ipcs/publications/new_issues/endocrine_disruptors/en.

Table C.2.23. **Avian Two-Generation Toxicity Test in the Japanese Quail (ATGT) (US EPA OCSPP 890.2100/740-C-15-003): Guidance for scenarios of combinations of results with existing data**

This table represents possible conclusions to be drawn from assay data, and a next step which could be taken if further evidence is required about possible endocrine disrupting properties and/or effects. The guidance offered is not meant to be prescriptive, but provides science-based considerations. It encourages the use of all available data and expert judgement in a weight of evidence approach. Regional and national interpretation of results and "next steps" may vary.

The conclusions are grouped into a series of scenarios (A-R), each scenario representing a different combination of assay results, existing *in vitro* data and existing *in vivo* data. The symbol "+" indicates that the data in question represent a positive result, «-" indicates a negative result, and "Eq/0" indicates that the data are either equivocal or are not available.

Existing results: * "Mechanism (*in vitro* mechanistic data)" assumes that mechanistic data are available from endocrine receptor (ER-), androgen receptor (AR-) and steroidogenesis-based assays (Level 2). Thyroid hormone receptor (TR) and other assays concerning mechanisms of thyroid disruption may be available, but they are not in common use. In practice, data from all assays may not be available and therefore this must be taken into account when deciding on the "next step". Quantitative structure activity relationship (QSAR) predictions of estrogen and androgen binding/activation may be made for some substances. There is no evidence at present that equivalent *in vitro* assays with systems derived from birds offer advantages over their mammalian counterparts.

Existing results: ** "Effects (*in vivo* effects of concern)" assumes effects have been observed in other *in vivo* screens/tests which give rise to concern that the test chemical may be an endocrine disrupter.

The assay under discussion could either be positive for both apical endpoints and indicators of endocrine activity (e.g. mechanistic endpoints such as hormone titres and gonad histopathology), or positive just for apical endpoints, or positive just for indicators of endocrine activity. For each scenario, each of these three possibilities is addressed separately in the possible conclusions column.

Scenario	Result of ATGT	Existing results		Possible conclusions: 1) Indicators of endocrine activity and apical endpoints positive 2) Indicators of endocrine activity positive and apical endpoints negative 3) Indicators of endocrine activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
A	+	+	+	1) Strong evidence for adverse effects in birds and other organisms by an endocrine mechanism. 2) Strong evidence for endocrine effects, but they do not appear adverse in birds. 3) Strong evidence for adverse effects in more than one organism, but mechanism may not be via direct interaction with endocrine receptor (ER) or androgen receptor (AR), or by aromatase inhibition or thyroid disruption.	Probably no need for additional data.	–
B	+	+	–	1) Strong evidence for adverse effects in birds by an endocrine mechanism. 2) Strong evidence for endocrine effects in birds, but they do not appear adverse. 3) Strong evidence for adverse effects in birds, but mechanism may not be via direct interaction with ER or AR, or by aromatase inhibition or thyroid disruption.	Probably no need for additional data.	–
C	+	+	Eq/0**	1) Strong evidence for adverse effects in birds by an endocrine mechanism. 2) Strong evidence for endocrine effects in birds, but they do not appear adverse. 3) Strong evidence for adverse effects in birds, but mechanism may not be via direct interaction with ER or AR, or by aromatase inhibition or thyroid disruption.	Probably no need for additional data.	It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple modes of action (MOA). If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
D	+	–	+	1) Strong evidence for adverse effects in birds and other organisms, possibly by an unknown endocrine mechanism. 2) Medium-strong evidence for endocrine effects in birds, but they do not appear to be adverse. 3) Strong evidence for adverse effects in more than one organism, but mechanism may not be by endocrine disruption.	Probably no need for additional data, but see right-hand column.	Negative *in vitro* mechanistic data combined with positive endocrine-specific *in vivo* endpoints suggests that an unknown endocrine MOA is causing any adverse effects. Some regulatory authorities may consider that the MOA needs further investigation before it can be concluded that the chemical in this scenario (sub-section 1) is an ED.
E	+	–	–	1) Strong evidence for adverse effects in birds, possibly by an unknown endocrine mechanism. 2) Medium-strong evidence for endocrine effects in birds, but they do not appear to be adverse. 3) Strong evidence for adverse effects in birds, but mechanism may not be by endocrine disruption.	Probably no need for additional data, but see right-hand column.	Negative *in vitro* mechanistic data combined with positive endocrine-specific *in vivo* endpoints suggests that an unknown endocrine MOA is causing any adverse effects. Some regulatory authorities may consider that the MOA needs further investigation before it can be concluded that the chemical in this scenario (sub-section 1) is an ED, especially as existing *in vivo* data are negative.

Scenario	Result of ATGT	Existing results		Possible conclusions: 1) Indicators of endocrine activity and apical endpoints positive 2) Indicators of endocrine activity positive and apical endpoints negative 3) Indicators of endocrine activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
F	+	–	Eq/0	1) Strong evidence for adverse effects in birds, possibly by an unknown endocrine mechanism. 2) Medium-strong evidence for endocrine effects in birds, but they do not appear to be adverse. 3) Strong evidence for adverse effects in birds, but mechanism may not be by endocrine disruption.	Probably no need for additional data, but see right-hand column.	Negative *in vitro* mechanistic data combined with positive endocrine-specific *in vivo* endpoints suggests that an unknown endocrine MOA is causing any adverse effects. Some regulatory authorities may consider that the MOA needs further investigation before it can be concluded that the chemical in this scenario (sub-section 1) is an ED, especially as existing *in vivo* data are equivocal or absent. However, such a conclusion appears likely on the basis of the positive endocrine-sensitive endpoints *in vivo*. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
G	+	Eq/0	+	1) Strong evidence for adverse effects in more than one organism, possibly by an unknown endocrine mechanism. 2) Medium-strong evidence for endocrine effects, but they do not appear to be adverse in birds. 3) Strong evidence for adverse effects in more than one organism, but mechanism may not be by endocrine disruption.	It would be desirable to obtain some clear mechanistic data before concluding that the chemical is an ED. See right-hand column.	Some regulatory authorities may consider that the MOA needs further investigation before it can be concluded that the chemical in sub-section 1 is an ED. However, such a conclusion appears likely on the basis of the positive endocrine-sensitive endpoints *in vivo*. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
H	+	Eq/0	–	1) Strong evidence for adverse effects in birds, possibly by an unknown endocrine mechanism. 2) Medium-strong evidence for endocrine effects in birds, but they do not appear to be adverse. 3) Strong evidence for adverse effects in birds, but mechanism may not be by endocrine disruption.	It would be desirable to obtain some clear mechanistic data before concluding whether the chemical is an ED. See right-hand column.	Some regulatory authorities may consider that the MOA needs further investigation before it can be concluded that the chemical in sub-section 1 is an ED. However, such a conclusion appears likely on the basis of the positive endocrine-sensitive endpoints *in vivo*. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

Scenario	Result of ATGT	Existing results		Possible conclusions: 1) Indicators of endocrine activity and apical endpoints positive 2) Indicators of endocrine activity positive and apical endpoints negative 3) Indicators of endocrine activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
I	+	Eq/0	Eq/0	1) Strong evidence for adverse effects in birds, possibly by an unknown endocrine mechanism. 2) Moderate-strong evidence for endocrine effects in birds, but they do not appear to be adverse. 3) Strong evidence for adverse effects in birds, but mechanism may not be by endocrine disruption.	It would be desirable to obtain some clear mechanistic data before concluding whether the chemical is an ED. See right-hand column.	Some regulatory authorities may consider that the MOA needs further investigation before it can be concluded that the chemical in sub-section 1 is an ED. However, such a conclusion appears likely on the basis of the positive endocrine-sensitive endpoints *in vivo*. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
J	–	+	+	The chemical is an ED *in vivo* in other species but does not appear to act on growth, sexual development or reproduction in birds. If any other bird tests are also negative, birds may not be responsive at all to the test chemical.	Regulatory authorities may consider that further evidence is not required.	The fact that the chemical has endocrine properties *in vitro* and in other species *in vivo* suggests that it may be an ED, but probably not in birds. If the existing positive *in vivo* data are from a lower tier bird assay, note that it is generally considered that a negative higher tier test trumps a positive lower tier test.
K	–	+	–	Despite the *in vitro* mechanistic data for possible endocrine activity, there is no evidence for endocrine disruption *in vivo*. This may be because the chemical is degraded to an inactive metabolite, or because it only interacts very weakly with endocrine receptors.	Regulatory authorities may consider that further evidence is not required.	–
L	–	+	Eq/0	The chemical is not an ED in birds, but it may be active in other species as there is only one unequivocal *in vivo* test result (a negative).	Regulatory authorities may consider that further evidence is not required.	It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
M	–	–	+	The chemical is probably not an ED acting on growth, sexual development or reproduction in birds, but it does have endocrine activity in other species. However, it may act through MOA not covered by the available *in vitro* assays, or it may be more potent in a bird species other than that tested.	Regulatory authorities may consider that further evidence is not required.	The fact that the chemical has endocrine properties in other species *in vivo* suggests that it may be an ED, but probably not in birds. If the existing positive *in vivo* data are from a lower tier bird assay, note that it is generally considered that a negative higher tier test trumps a positive lower tier test.
N	–	–	–	The chemical is probably not an ED in birds or other species.	Regulatory authorities may consider that further evidence is not required.	–
O	–	–	Eq/0	The chemical is probably not an ED in birds.	Regulatory authorities may consider that further evidence is not required.	It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

Scenario	Result of ATGT	Existing results		Possible conclusions: 1) Indicators of endocrine activity and apical endpoints positive 2) Indicators of endocrine activity positive and apical endpoints negative 3) Indicators of endocrine activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
P	–	Eq/0	+	The chemical is probably not an ED acting on growth, sexual development or reproduction in birds, but it does have endocrine activity in other species. However, it may act through MOA not covered by the available *in vitro* assays, or it may be more potent in a bird species other than that tested.	Regulatory authorities may consider that further evidence is not required.	The fact that the chemical has endocrine properties in other species *in vivo* suggests that it may be an ED, but probably not in birds. If the existing positive *in vivo* data are from a lower tier bird assay, note that it is generally considered that a negative higher tier test trumps a positive lower tier test. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
Q	–	Eq/0	–	The chemical is probably not an ED acting on growth, sexual development or reproduction in birds, or *in vivo* on other species.	Regulatory authorities may consider that further evidence is not required, although negative *in vitro* data would strengthen the conclusion that the chemical is probably not an ED.	It should be borne in mind that equivocal data may be due to a variety of causes including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
R	–	Eq/0	Eq/0	The chemical is probably not an ED acting on growth, sexual development or reproduction in birds.	Regulatory authorities may consider that further evidence is not required, although negative *in vitro* data would strengthen the conclusion that the chemical is probably not an ED.	It should be borne in mind that equivocal data may be due to a variety of causes including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

C.2.24. RADAR: Rapid Androgen Disruption Adverse Outcome Reporter Assay (draft OECD TG)

Status: Assay being validated by the OECD.

634. Modality detected/endpoints: This draft *in vivo* transfected medaka assay is sensitive to androgen receptor agonists and androgen receptor antagonists and to chemicals interfering with androgen biosynthesis. In principle, it can also be used to identify estrogen agonists and antagonists, as well as aromatase inhibitors. However, this guidance will restrict itself to the detection of receptor-mediated androgenicity and anti-androgenicity alone, as data on responses to the other modalities are not yet available.

Background to the assay

635. This assay is started validation by the OECD in June 2017 for possible approval as a test guideline (TG), a Standard Project Submission Form (SPSF) having been approved by the Working Group of National Coordinators of the Test Guidelines Programme in April 2017. No validation data have yet been produced, but some published data on development and use of the assay are available (Sébillot et al., 2014). It is planned to have a TG ready by 2020 at the earliest. The assay is based on freshly hatched embryonic medaka (*Oryzias latipes*), stably transfected with the *spiggin1* promoter cloned upstream of a green fluorescent protein coding sequence. The presence of the spiggin promoter linked to androgen receptor alpha (ARα) means that the transparent transgenic fish fry will fluoresce green when exposed to an androgen for up to six days. The presence of an anti-androgen can be detected by exposing the fish in combination with an androgen such as 17-methyl testosterone (17MT) and measuring the decrease in expected fluorescence. The assay is relatively cheap to operate by comparison with *in vivo* screening assays using juvenile or adult fish. Furthermore, its sensitivity to anti-androgens is expected to be broadly similar to the Androgenised Female Stickleback Screen (AFSS – OECD GD 148) (Sébillot et al., 2014), and it is expected that the metabolic capability of medaka embryos, while limited by comparison with adult fish, will allow the detection of some metabolically activated endocrine active substances (EASs). It can be run in multiwell plates and is potentially suitable for use in a robotic screening programme.

When/why the assay may be used

636. Although data from RADAR could, in principle, be available at any stage in the hazard assessment process, the most likely scenario will be when there are relatively few data available about the possible endocrine disrupting properties of a chemical. The assay is most likely to be used either as part of a battery of *in vitro* and *in vivo* screens, or to follow up on existing data which suggest possible endocrine disruption activity. Given the high degree of endocrine system conservation across the vertebrates, adverse endocrine-linked effects in the RADAR assay may also indicate the possibility of related activity in other organisms such as amphibians, reptiles, birds or mammals. It is also possible that no

existing endocrine-relevant data are available (i.e. RADAR has been used as a primary screen), but in that case a positive result in the screen should ideally be followed up with relevant *in vitro* screening in an attempt to confirm the precise mode of action (MOA). Furthermore, a positive RADAR result would also need to be followed up with an additional *in vivo* fish test such as the Fish Short-Term Reproduction Assay (FSTRA – OECD TG 229) or Fish Sexual Development Test (FSDT – OECD TG 234), which will give some indication of any adverse apical effects. Possible conclusions to be derived from the results of RADAR, and guidance about potential additional studies to strengthen weight of evidence, are summarised in Table C.2.24.

637. In order to provide information relevant for assessing whether or not a chemical may fulfil the WHO/IPCS (2002) definition of an endocrine disruptor (ED), the study design has to be sufficiently robust to demonstrate the presence or absence of effects. In the dose selection, the investigator should also consider and ensure that data generated are adequate to fulfil the regulatory requirement across OECD countries as appropriate (e.g. hazard and risk assessment and labelling, ED assessment, etc.). The top dose or concentration should be sufficiently high to give clear systemic (i.e. non endocrine-specific) toxicity in order to ensure that a wide range of exposures (high to low) is tested. However, endocrine effects observed solely in the presence of clear systemic toxicity should be interpreted with caution and may be disregarded when sufficiently justified to be caused by secondary effects which are unlikely to be due to endocrine activity. The reason for this advice is a concern that some EAS-sensitive assays are being run at doses/concentrations of EASs that are too low to trigger direct impacts on the endocrine system. This guidance document is not the place to address this issue directly, but it should be considered when EAS-sensitive TGs are revised in the future. In addition, the number and spacing of dose/concentration levels should also be adequate to fulfil the objectives of the study (e.g. to demonstrate dose response relationships if this is required).

Existing data to be considered

638. Existing information on endocrine-related effects from other vertebrates (up to and including mammals, e.g. positive findings for endocrine endpoints in mammalian repeat dose toxicity or reproductive studies) should always be considered, given the commonality of endocrine mechanisms in these taxa. Existing data available before deployment of RADAR might include *in vivo* results obtained with other vertebrates (e.g. a Hershberger Bioassay with rodents – OECD TG 441), or one or more of a range of *in silico* or *in vitro* results which suggest that androgenicity or anti-androgenicity may occur *in vivo*. Such indicators of possible *in vivo* activity might include quantitative structure activity relationship (QSAR) predictions of endocrine activity, high throughput screening data, "read-across" from *in vivo* results obtained with structurally related chemicals or positive results from an *in vitro* screen for androgen receptor-mediated activity, or for effects on androgen biosynthesis.

639. It should be noted that a sensitive *in vivo* assay for anti-androgenicity is already available, the AFSS (OECD GD 148). This is longer than RADAR (21 day), and relies on the pre-treatment of adult female sticklebacks (*Gasterosteus aculeatus*) with an androgen before measuring anti-androgenic effects of the test chemical (reduction in induced spiggin glue protein).

Scenarios: Positive and negative results combined with existing data

640. The scenarios (A to R) presented in Table C.2.24 represent all the possibilities of positive or negative results in combination with the presence or absence of existing data. The action taken will also depend on the regulatory environment, but the considerations given here are generally science based. Wherever possible, the recommended "next step which could be taken" avoids unnecessary animal testing. However, sometimes conducting an animal test will be indicated and then the relevance of species, strain and exposure route should always be considered. Further considerations specific to each scenario are given in the table.

641. Positive results obtained with the fluorescence endpoint (Table C.2.24, Scenarios A-I) result in the conclusion that the test chemical is a possible androgen or anti-androgen *in vivo*. This would ideally need to be followed up with more comprehensive testing to show whether adverse apical effects related to endocrine impacts occur at any part of the life cycle (and hence to discover whether the chemical is an ED acting through certain estrogen/androgen/thyroid/steroidogenesis [E,A,T,S] pathways). In other words, a positive result in the RADAR assay may trigger OECD TG 234 (FSDT) at Level 4 or fish life cycle testing (e.g. Medaka Extended One-Generation Reproduction Test [MEOGRT] – OECD TG 240) at Level 5. Existing data suggesting androgenic or anti-androgenic activity will strengthen the case for additional testing.

642. The situation in which the RADAR assay gives a negative result (Table C.2.24, Scenarios J-R) needs careful consideration of any existing data. If the weight of evidence of these data suggests that the chemical is endocrine active both *in vitro* and *in vivo* in other species (Scenario J), then the probability is that RADAR may simply be insufficiently responsive in that case, or fish in general may be unresponsive. For example, this might be the case if the medaka embryos have not transformed a chemical to an active metabolite. In some of these circumstances, it might be appropriate to conduct an FSDT (OECD TG 234), or alternatively, a fish life cycle test (e.g. MEOGRT OECD TG 240) to confirm that there is no endocrine activity in fish.

643. If the RADAR and existing *in vivo* data are all negative, but *in vitro* data reveal some endocrine activity (Scenario K), the probability is that the test chemical is not sufficiently potent to produce endocrine effects *in vivo* in fish, or it may be rapidly metabolised. In such a situation, further testing may or may not be necessary. If the chemical is known to bioaccumulate slowly, it may be that exposures in the *in vivo* tests are not of sufficient duration, in which case longer term testing might be justified.

644. On the other hand, if the RADAR and the *in vitro* tests are negative, but there are positive existing *in vivo* data (Scenario M), the chemical is probably not a potential ED with androgenic or anti-androgenic activity, but it may act via androgen-related MOA not covered by the *in vitro* screens, or it may be more potent in species or life stages that have not been tested. In this situation, the relevant existing *in vitro* and *in vivo* data should be used to guide decisions about whether to conduct any further testing, including life stages represented in OECD TG 234 (FSDT) or TG 240 (MEOGRT).

645. Finally, a negative RADAR, set against a background of negative *in vitro* and *in vivo* data (Scenario N) that includes relevant *in vivo* data for fish, suggests that the test chemical is not a potential ED in fish or other vertebrates, and no further testing for androgenic or anti-androgenic MOA will generally be necessary.

646. In each of the above scenarios, it is possible that existing data will be equivocal, or there may be no existing data (Scenarios C, F-I, L and O-R). This will weaken the conclusions which can be drawn about a negative RADAR, and this is reflected in Table C.2.24. However, a lack of mechanistic data on endocrine activity should usually be rectified before any further *in vivo* testing is finally decided on. Indeed, as a general principle, it is desirable to obtain mechanistic data before any *in vivo* testing. On the other hand, if RADAR is positive, further *in vivo* testing is generally indicated, particularly when existing data are equivocal, or if there are no existing data. There is also the possibility that equivocal mechanistic data may be the result of multiple modes of endocrine action. Under some circumstances, two opposite modes of simultaneous action (e.g. androgenic and anti-androgenic) could, depending on dose, lead to a minimisation or abolition of adverse effects, while in others two different MOA (e.g. anti-steroidogenic and androgenic) could potentially reinforce effects on the RADAR. If multiple MOA are suspected, either from the existing results or based on QSAR/read-across/integrated approaches, this situation should be investigated further if needed for regulatory decision making.

647. The scenario in which the results of the RADAR are themselves equivocal has not been dealt with in Table C.2.24, for reasons of brevity. In this context, an equivocal result might be an inconsistent concentration-response (e.g. no effect at a high concentration but effects at a lower concentration), or a result which borders on statistical significance. Without knowing the exact circumstances, reliable advice cannot be given, but the opinions of an experienced ecotoxicologist should be sought. Clearly, however, such equivocal results do not necessarily rule out the existence of *in vivo* endocrine activity. If false negatives (e.g. systemic toxicity) are suspected with good reason, the screen could be repeated if none of the test concentrations have given reliable data (e.g. conduct it at lower concentrations which avoid systemic toxicity). However, note that a repeat test in the event of systemic toxicity would not be needed providing at least one tested concentration was not subject to such effects.

648. In summary, positive results in the RADAR assay indicate that a chemical is a possible endocrine disrupter. More predictive *in vivo* testing would then be necessary to produce a long-term no-observed-effect-concentration/x% effect concentration (NOEC/ECx) and/or to confirm whether or not the chemical is an actual endocrine disrupter with adverse effects *in vivo*. Negative results in the RADAR do not necessarily mean that the chemical is not a potential ED – a judgement about its endocrine disruption potential and the possible need for additional testing will have to be made based on a weight of evidence evaluation of existing *in vitro* and *in vivo* data.

References

Sébillot, A. et al. (2014), "Rapid fluorescent detection of (anti)androgens with spiggin-gfp medaka", *Environmental Science & Technology*, Vol. 48/18, pp. 10919-10928, http://dx.doi.org/10.1021/es5030977.

WHO/IPCS (2002), "Global assessment of the state-of-the-science of endocrine disrupters", Damstra, T. et al. (eds.) WHO/PCS/EDC/02.2, World Health Organization, Geneva, www.who.int/ipcs/publications/new_issues/endocrine_disruptors/en.

Table C.2.24. **RADAR: Rapid Androgen Disruption Adverse Outcome Reporter Assay (draft OECD TG): Guidance for scenarios of combinations of results with existing data**

This table represents possible conclusions to be drawn from assay data, and a next step which could be taken if further evidence is required about possible endocrine disrupting properties and/or effects. The guidance offered is not meant to be prescriptive, but provides science-based considerations. It encourages the use of all available data and expert judgement in a weight of evidence approach. Regional and national interpretation of results and "next steps" may vary.

The conclusions are grouped into a series of scenarios (A-R), each scenario representing a different combination of assay results, existing *in vitro* data and existing *in vivo* data. The symbol "+" indicates that the data in question represent a positive result, «-" indicates a negative result, and "Eq/0" indicates that the data are either equivocal or are not available. Note that there are no apical endpoints in this assay considered to be diagnostic of an estrogen/androgen/thyroid/steroidogenesis (E,A,T,S) modality.

Existing results: * "Mechanism (*in vitro* mechanistic data)" assumes that mechanistic data are available from estrogen receptor (ER-), androgen receptor (AR-), and steroidogenesis-based assays (Level 2). Thyroid hormone receptor (TR) and other assays concerning mechanisms of thyroid disruption may be available, but they are not in common use. In practice, data from all assays may not be available and therefore this must be taken into account when deciding on the "next step". Quantitative structure activity relationship (QSAR) predictions of estrogen and androgen binding/activation may be made for some substances. There is no evidence at present that equivalent *in vitro* assays with systems derived from fish offer advantages over their mammalian counterparts.

Existing results: ** "Effects (*in vivo* effects of concern)" assumes effects have been observed in other *in vivo* screens/tests which give rise to concern that the test chemical may be an endocrine disrupter.

Scenarios	Result of RADAR	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
A	+	+	+	Strong evidence for *in vivo* androgenic or anti-androgenic activity in fish and other organisms.	Consider performing a fish life cycle test (e.g. MEOGRT – OECD TG 240), especially if the intention is to obtain precise data on a reproductive or developmental no-observed-effect-concentration/x% effect concentration (NOEC/ECx).	An alternative approach would be to deploy OECD TG 234 (Fish Sexual Development Test), especially if sexual development is expected to give a response at lower concentrations than reproduction.
B	+	+	–	Strong evidence for *in vivo* androgenic or anti-androgenic activity in fish, despite lack of *in vivo* effects in existing tests.	Consider performing a fish life cycle test (e.g. MEOGRT – OECD TG 240), especially if the intention is to obtain precise data on a reproductive or developmental NOEC/ECx.	An alternative approach would be to deploy OECD TG 234 (FSDT), especially if sexual development is expected to give a response at lower concentrations than reproduction. If the negative *in vivo* data are from a fish test (e.g. OECD TG 229), consider possible reasons for the disparity (e.g. differences in species sensitivity) before conducting a life cycle test.
C	+	+	Eq/0**	Strong evidence for *in vivo* androgenic or anti-androgenic activity in fish, despite equivocal or absent *in vivo* data in other species.	Consider performing a fish life cycle test (e.g. MEOGRT – OECD TG 240), especially if the intention is to obtain precise data on a reproductive or developmental NOEC/ECx.	If no existing fish data are available, it may be worth performing OECD TG 234 (FSDT) before a possible life cycle test in order to obtain information on whether sexual development is a sensitive part of the life cycle. Such information could influence the design of the life cycle test. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple modes of action (MOA). If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
D	+	–	+	Strong evidence for *in vivo* androgenic or anti-androgenic activity in fish and other species, but confidence about MOA is reduced by negative mechanistic data.	Consider performing a fish life cycle test (e.g. MEOGRT – OECD TG 240), especially if the intention is to obtain precise data on a reproductive or developmental NOEC/ECx.	The negative *in vitro* data suggest that the test chemical may be metabolically activated *in vivo*, or it may operate via mechanisms not covered by the *in vitro* screens. An alternative approach to a life cycle test would be to deploy OECD TG 234 (FSDT), especially if sexual development is expected to give a response at lower concentrations than reproduction.
E	+	–	–	Moderate-strong evidence for *in vivo* androgenic or anti-androgenic activity in fish, but confidence is reduced by negative *in vitro* data and negative *in vivo* activity in other species.	Consider performing a fish life cycle test (e.g. MEOGRT – OECD TG 240), especially if the intention is to obtain precise data on a reproductive or developmental NOEC/ECx.	The negative *in vitro* data suggest that the test chemical may be metabolically activated *in vivo*, or it may operate via mechanisms not covered by the *in vitro* screens. An alternative approach to a life cycle test would be to deploy OECD TG 234 (FSDT), especially if sexual development is expected to give a response at lower concentrations than reproduction. If the negative *in vivo* data are from a fish test (e.g. OECD TG 229), consider possible reasons for the disparity (e.g. differences in species sensitivity) before conducting a life cycle test.

Scenarios	Result of RADAR	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
F	+	–	Eq/0	Moderate-strong evidence for *in vivo* androgenic or anti-androgenic activity in fish, but confidence is reduced by negative *in vitro* data and equivocal or absent *in vivo* activity in other species.	Consider performing a fish life cycle test (e.g. MEOGRT – OECD TG 240), especially if the intention is to obtain precise data on a reproductive or developmental NOEC/ECx.	The negative *in vitro* data suggest that the test chemical may be metabolically activated *in vivo*, or it may operate via mechanisms not covered by the *in vitro* screens. If no existing fish data are available, it may be worth performing OECD TG 234 (FSDT) before a possible life cycle test in order to obtain information on whether sexual development is a sensitive part of the life cycle. Such information could influence the design of the life cycle test. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
G	+	Eq/0	+	Strong evidence for *in vivo* androgenic or anti-androgenic activity in fish, but mechanism unconfirmed.	Obtain mechanistic data, then consider performing a fish life cycle test (e.g. MEOGRT – OECD TG 240).	An alternative approach to a life cycle test would be to deploy OECD TG 234 (FSDT), especially if sexual development is expected to give a response at lower concentrations than reproduction. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
H	+	Eq/0	–	Strong-moderate evidence for *in vivo* androgenic or anti-androgenic activity in fish, but mechanism unconfirmed.	Obtain mechanistic data, then consider performing a fish life cycle test (e.g. MEOGRT – OECD TG 240).	An alternative approach to a life cycle test would be to deploy OECD TG 234 (FSDT), especially if sexual development is expected to give a response at lower concentrations than reproduction. If the negative *in vivo* data are from a fish test (e.g. OECD TG 229), consider possible reasons for the disparity (e.g. differences in species sensitivity) before possibly conducting a life cycle test. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
I	+	Eq/0	Eq/0	Moderate evidence for *in vivo* endocrine activity in fish, but mechanism unconfirmed.	Obtain mechanistic data, then consider performing a fish life cycle test (e.g. MEOGRT – OECD TG 240).	If no existing fish data are available, it may be worth performing OECD TG 234 (FSDT) before a possible life cycle test in order to obtain information on whether sexual development is a sensitive part of the life cycle. Such information could influence the design of the life cycle test. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

Scenarios	Result of RADAR	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
J	–	+	+	Based on the existing data, the chemical has androgenic or anti-androgenic activity *in vivo*. The lack of response in RADAR suggests that fish are not responsive, unless the existing data are from fish.	Consider performing OECD TG 234 (FSDT).	It is possible that the failure to give a positive result in RADAR was caused by the relatively short exposure time (up to six days). If this is suspected (e.g. the chemical only bioaccumulates slowly), or if the existing *in vivo* data are from a fish, OECD TG 234 (FSDT) or potentially a life cycle test (e.g. OECD TG 240 – MEOGRT) would be able to study the effects of longer exposure and confirm whether there is a hazard to fish. Choice of test should be guided by the existing *in vivo* data.
K	–	+	–	There is no evidence that the chemical is a possible androgenic or anti-androgenic ED *in vivo*, probably because it is very weakly acting or rapidly metabolised.	Probably no further action, but see comments in right-hand column.	It is possible that EDs which bioaccumulate slowly may only cause effects *in vivo* after exposure times longer than 28 days. If this is suspected, and depending on which part of the life cycle is suspected of being the most sensitive, consider performing OECD TG 234 (FSDT).
L	–	+	Eq/0	The chemical may not be an androgenic or anti-androgenic ED *in vivo*, but the confidence in this conclusion is relatively low as there is only one unequivocal *in vivo* test result (a negative).	Probably no further action, but see comments in right-hand column.	If the equivocal existing data are from a fish assay, consider performing a fish assay (e.g. OECD TG 229 or TG 230) with a different species, or a longer term test (e.g. OECD TG 234 [FSDT] or life cycle test [MEOGRT – TG 240]) if the chemical is a slow bioaccumulator. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
M	–	–	+	The chemical is apparently not a possible androgenic or anti-androgenic ED in fish but it does have activity in another species.	Use the existing *in vivo* data to help decide whether a longer term test with an appropriate fish species is indicated.	Use the existing *in vivo* data to guide any further testing.
N	–	–	–	The chemical is probably not a possible androgenic or anti-androgenic ED *in vivo*.	No further action with respect to androgenic or anti-androgenic MOA.	
O	–	–	Eq/0	The chemical is probably not a possible androgenic or anti-androgenic ED in fish.	Probably no further action. However, see comments in right-hand column.	If the paucity of *in vivo* data is a concern, performance of a screening test (OECD TG 229 or TG 230) with a different species, or a longer term test (i.e. OECD TG 234 [FSDT]) or life cycle test [MEOGRT]) could be considered. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

Scenarios	Result of RADAR	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
P	–	Eq/0	+	The chemical is probably not a possible androgenic or anti-androgenic ED in fish, but confidence in this conclusion is low given the lack of mechanistic *in vitro* data and the availability of positive existing *in vivo* data.	Obtain mechanistic data, then consider whether further testing is desirable.	If the mechanistic data confirm that the chemical has potential androgenic or anti-androgenic action, consider conducting a fish assay (OECD TG 229 or TG 230) with another species, or a longer term test (OECD TG 234 [FSDT] or life cycle test [MEOGRT]). Use the existing *in vivo* data as a guide to test choice. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
Q	–	Eq/0	–	The chemical is probably not a possible androgenic or anti-androgenic ED in fish, but the lack of mechanistic *in vitro* data are a concern, even though the existing *in vivo* data are negative.	Obtain mechanistic data, then consider whether further testing is desirable.	If the mechanistic data confirm that the chemical has potential endocrine action, consider conducting a fish assay (OECD TG 229 or TG 230) with another species, or a longer term test (OECD TG 234 [FSDT] or life cycle test [MEOGRT]). Use the existing *in vivo* data as a guide to test choice. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.
R	–	Eq/0	Eq/0	The chemical is probably not a possible androgenic or anti-androgenic ED in fish, but confidence in this conclusion is low given the lack of mechanistic *in vitro* and existing *in vivo* data.	Obtain mechanistic data, then consider whether further testing is desirable.	If the mechanistic data confirm that the chemical has potential androgenic or anti-androgenic action, consider conducting a fish assay (OECD TG 229 or TG 230) with another species, or a longer term test (OECD TG 234 [FSDT] or life cycle test [MEOGRT]). Use the existing *in vivo* data as a guide to test choice. It should be borne in mind that equivocal data may be due to a variety of causes, including experimental error, very weak endocrine activity or multiple MOA. If the latter case is suspected, it may be necessary to investigate the matter further and/or increase the weight given to the mechanistic information.

OECD mammalian screens and tests (Conceptual Framework Levels 3-5)

C.3.1. Uterotrophic Bioassay in Rodents (UT assay) (OECD TG 440) (including OECD GD 71 on the procedure to test for anti-estrogenicity)

Status: Assay validated by the OECD.

649. Modality detected/endpoints: estrogens (uterine wet weight and blotted weight ↑); anti-estrogens (stimulated uterine weight ↓); optional others (e.g. histopathologic changes in uterus/vagina).

Background to the assay

650. This assay is a short-term *in vivo* screening assay in female rodents for chemicals that interact with the estrogen receptor (ER). It is based on the increase in uterine weight (or uterotrophic response) that is elicited by ER agonists in animal models where endogenous estrogen levels are minimal. There are two variants of the assay; one uses immature animals, the other uses ovariectomised animals. The immature rodent assay may detect modalities acting via mechanisms other than ER, as the animals have an intact hypothalamic/pituitary/gonadal (HPG) axis, but the ability to detect these is limited. The assay may be conducted using rats or mice, but the there is more experience with the rat assay and this species was used in the OECD validation of this assay (OECD, 2006). Route of administration of test substance is via oral gavage or subcutaneous injection. This assay has been considered to be the "gold standard" bioassay screen for identifying ER agonists. A recently curated database of bioactivity with results from over 2 500 Uterotrophic Bioassays in rats and mice provides comprehensive information on this assay (Kleinstreuer et al., 2016).

651. Although this assay is a "screen", some authorities may regard an increase in uterine weight as possibly adverse. If this occurs in immature animals at a point in time when this should not occur naturally then this could represent an adverse effect in a sensitive life stage. Likewise, the ovariectomised UT assay may be regarded as a model for immature animals and therefore a uterine weight increase could be regarded as adverse. Interpretations of the results of this assay may vary according to region and regulation and should always utilise all data in a weight of evidence approach.

652. Non-aromatisable (steroidal and non-steroidal) androgens and aromatisable androgens that may be metabolised to estrogens have also been shown to increase uterine weight. In immature animals, aromatisable androgens like testosterone elicit histopathologic changes very similar to that of estradiol, suggesting that the observed changes are mediated through estrogen. For all other conditions, the observed histopathologic changes are different and are considered to be mediated via the androgen receptor (AR). In practical terms, this issue is of minor importance. Potentially aromatisable androgens can easily be identified based on their structural features, and non-steroidal androgenic chemicals are currently considered to be rare in the chemical universe. In addition, progesterone and synthetic progestins may also give a positive response (Jones and Edgren, 1973).

653. The OECD test guideline (TG 440) was adopted in October 2007 and is specific for estrogen agonists only. The validation of the assay was not considered adequate for anti-estrogens as there were insufficient pure anti-estrogens available. The test for anti-estrogens, however, is frequently used and is available as OECD GD 71 (OECD, 2007). Its use as an assay was reviewed during the validation of the UT assay (Owens and Ashby, 2002) and it continues to be used to date.

When/why the assay may be used

654. Although OECD TG 440 can be used at any stage in the assessment process, the most likely use scenario will be following a positive result in an ER transactivation assay (ER STTA) and/or an ER binding assay, in order to determine whether the positive result *in vitro* is translated into a positive result *in vivo*. It may also be used as a screen in the absence of positive *in vitro* data, when a chemical that is negative in the *in vitro* ER interaction screens is suspected of producing estrogenic metabolites *in vivo*. In this case, the first option would be to use an additional metabolising system in the *in vitro* tests, but the Uterotrophic Bioassay as an *in vivo* test will include all metabolising systems. Another possible scenario is following observation of effects in higher tier tests, for example acceleration of puberty onset in females, but which are not exclusively indicative of an effect on ER. In the European Union, chemicals included in REACH, Plant Protection Products and Biocides legislation may have been tested in OECD TG 421/422, OECD TG 416 (Two-Generation Reproductive Toxicity Study) or the Extended One-Generation Reproduction Toxicity Study (EOGRTS – OECD TG 443), the UT assay may then be used as a follow up to clarify the mode of action (MOA). The UT assay is also likely to be carried out as part of the United States Environmental Protection Agency's Endocrine Disruptor Screening Program Tier 1 screening battery. Selection of the most appropriate tests has to be on a case-by-case basis, but also considering the need to minimise animal testing.

655. It should be noted that the UT assay was designed to be sensitive and will detect weak and strong ER modulators. In the validation of the UT assay, ethinylestradiol and oestradiol were defined as "strong" estrogens whilst nonylphenol and genistein were defined as "weak" estrogens (OECD, 2006). Weakly acting chemicals may not always be detected as endocrine disruptors (EDs) when tested in higher level tests because the endocrine system in intact/adult animals has a greater ability to compensate than in the UT assay where the HPG axis is disrupted/immature. Furthermore, in case of repeat dose studies, dose levels may need adjustment to lower doses in order to cope with general toxicity.

656. The route of exposure is also an important consideration for the UT assay. OECD TG 440 states that chemicals may be administered by oral or subcutaneous routes but suggests that the route most relevant for human exposure should be used. The route will have consequences for absorption, distribution, metabolism and excretion and is an important consideration when interpreting results. Methoxychlor, for example, gave negative results when administered by subcutaneous injection but positive results when given orally (due to metabolism to estrogenic metabolites) (Laws et al., 2000).

657. In order to provide information relevant for assessing whether or not a chemical may fulfil the WHO/IPCS (2002) definition of an ED, the study design has to be sufficiently robust to demonstrate the presence or absence of effects. In the dose selection, the investigator should also consider and ensure that data generated are adequate to fulfil the regulatory requirement across OECD countries as appropriate (e.g. hazard and risk assessment and labelling, ED assessment, etc.). The top dose or concentration should be

sufficiently high to give clear systemic (i.e. non endocrine-specific) toxicity in order to ensure that a wide range of exposures (high to low) is tested. However, endocrine effects observed solely in the presence of clear systemic toxicity should be interpreted with caution and may be disregarded when sufficiently justified to be caused by secondary effects which are unlikely to be due to endocrine activity. The reason for this advice is a concern that some endocrine active substance (EAS) sensitive assays are being run at doses/concentrations of EASs that are too low to trigger direct impacts on the endocrine system. This guidance document is not the place to address this issue directly, but it should be considered when EAS-sensitive test guidelines (TGs) are revised in the future. In addition, the number and spacing of dose/concentration levels should also be adequate to fulfil the objectives of the study (e.g. to demonstrate dose response relationships if this is required).

Introduction to the table of scenarios

658. Table C.3.1 gives guidance on a further step to take in the event of a positive (+) or negative (-) result in the UT assay and in the presence of positive (+), negative (-) or equivocal/absent (Eq/0) existing results. "Existing results" are subdivided into "mechanism" and "effects" data (third and fourth columns). The table is divided horizontally into a series of scenarios that represent all the combinations of these events.

659. The results of OECD TG 440 are given in the second column. Criteria for positive results in OECD TG 440 are given in the test guideline itself (i.e. a statistically significant increase in uterine weight compared to the solvent control). A positive result in the assay for anti-estrogenicity would be a statistically significant decrease in uterine weight compared to the estrogen-stimulated control group. Negative results are no (statistically significant) changes in wet and blotted uterine weight. It is important that quality criteria for control uterine weights are demonstrated. It is also of note that a uterotrophic response may not always be entirely of estrogenic origin (e.g. testosterone may give a positive result, chemicals interacting with other endocrine axes may give a positive result in the immature rodent assay, diets high in phytoestrogens or energy sources may also give a positive result). Further guidance is provided in the TG. Optional endpoints may include histopathologic changes in uterus/vagina or vaginal cornification in the ovariectomised rat assay. These endpoints should supplement the uterotrophic response. Changes in these endpoints in the absence of a uterotrophic response should be considered equivocal.

660. Equivocal results for the guideline are not included in Table C.3.1 because these data require further interrogation about the result itself. In the event of an equivocal result, the considerations mentioned above about uterine weights in control animals, non ER-related changes, possible effects of phytoestrogens or high energy diets should be taken into account and further investigations made.

Existing data to be considered

661. Existing "mechanism" *in vitro* data are assumed to be available from ER (ER binding and ER STTA), AR (AR binding and AR STTA) and steroidogenesis-based assays (Level 2). Assays may also be available for interference with thyroid modalities. In practice, it is possible that data from all of these assays may not be available, so judgement will need to be used to decide which assays to perform. Thyroid hormone receptor (TR-) based assays are less relevant for the UT assay. Although the current *in vitro* TGs do not incorporate metabolic activation, published information on use of metabolic activation

systems is available in Jacobs et al. (2008, 2013) and OECD (2008). These methods, however, have not yet been validated.

662. Existing "effects" data refer to *in vivo* effects that may come from varied sources and will depend on the type of chemical (e.g. new chemicals, high production volume chemical, pesticide). Thus, available data may range from repeated dose toxicity studies (28-day, 90-day) or combined repeat dose/reproductive screening assays to chronic toxicity studies and multigeneration reproductive tests. Some studies fail to identify EDs that weakly affect estrogen or androgen receptors as was demonstrated on the basis of data generated in the validation process of OECD TG 407 assay with endocrine endpoints. In this validation, only moderate EDs such as nonylphenol and DDE, and strong EDs such as ethinylestradiol and flutamide (acting via ER and AR respectively), were detected. Thus OECD TG 407 cannot be regarded as a screening assay for endocrine activity. This means that when a relatively insensitive test is positive for both endocrine-specific and apical endpoints, this should be taken as an indication that the substance is a potential ED. Caution should be exercised, however, because endocrine endpoints may be impacted secondary to non-endocrine toxicity and *in vivo* apical endpoints can be affected by many MOA including endocrine modalities. The ability of a given assay to detect endocrine disruption will also vary depending on the version of the test guideline used. Older test guidelines may contain fewer endocrine-sensitive endpoints than more recent ones. If data are available from single or multigeneration studies that are adequately conducted with updated guidelines that include endpoints sensitive to EASs, then there should be no reason to conduct a UT assay as the higher tier test will provide stronger evidence for hazard identification/characterisation. Multigeneration studies conducted prior to the introduction of these endpoints will still provide valuable information on reproductive and endocrine organ toxicity, reproduction and development, but may not be sufficiently sensitive to EASs, in which case the UT assay would provide further valuable information. Data may also be available on effects in mammalian and non-mammalian wildlife species although caution should be used when extrapolating between taxa. A chemical causing endocrine effects in non-mammalian environmental species (fish, for example) may also have endocrine effects in mammals, but the physiological consequences of the effects are likely to be different.

663. When considering the results of the UT assay, all available data should be used in order to reach a conclusion and a weight of evidence approach taken. This may include high throughput screening data, read-across data from structural analogues and quantitative structure activity relationship (QSAR). Several QSAR models for ER and AR binding/activation are now available (see Sections B.1.1.1 and B.1.1.2).

Scenarios: Positive and negative results combined with existing data

664. The scenarios (A to R) presented in Table C.3.1 represent all the possibilities of positive or negative results in combination with the presence or absence of existing data. The action taken will also depend on the regulatory environment, but the considerations given here are generally science based. Although OECD TG 440 uses rats, the well-conserved nature of ER across taxa should be a strong indication that results in this assay are relevant to other vertebrate species. Results in laboratory mammal tests are also highly relevant for environmental mammalian species. Wherever possible, the recommended "next step which could be taken" avoids unnecessary animal testing. However, sometimes conducting an animal test will be indicated and then the relevance of species, strain, exposure route and species-specific metabolism should always be considered. The sensitivity and

physiological function of the hormone under investigation in the test species should also be considered. In general, lower level tests should be conducted before higher level tests in order to avoid unnecessary animal, usage unless it is apparent that a Level 5 test will be required anyway or will be needed to establish the evidence to conclude on ED properties. At Level 5, the EOGRTS (OECD TG 443) is the most sensitive reproduction assay for detecting endocrine disruption because it includes evaluation of a number of endocrine endpoints not included in the two-generation study (OECD TG 416) adopted in 2001. It is recognised, however, that some jurisdictions may require a two-generation study. Further considerations specific to each scenario are given in the table.

665. Scenarios A to C represent positive results in the UT assay in the presence of positive *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. A positive result in ER-based assays in combination with a positive UT assay is strong evidence for (anti)estrogenic activity that may or may not be supported by the *in vivo* effects data. Effects on endocrine endpoints in OECD TGs 407, 408, 453 or 421/422 may provide sufficient evidence to conclude concern for endocrine disruption and therefore there is no need for further screening. Positive results in the UT assay may also indicate similar (anti)estrogenicity in lower vertebrates. These could be followed up with partial life cycle tests such as the Fish Sexual Development Test (FSDT) or the Larval Amphibian Growth and Development Assay (LAGDA). *In vivo* assays/tests with negative results should be interpreted with caution as they may either indicate that the tests used do not have sufficient power to detect weak effects or, alternatively, that the effects do not present a concern for endocrine disruption. Selection of the dose level and the strain of animal should also be considered. Generally, a conclusion of lack of concern for endocrine disruption in the presence of positive effects data may only be made given adequate Level 5 assays. MOA data to provide a clear interpretation may be required by some regulatory agencies. The possibility of other mechanisms should also not be overlooked (e.g. positive AR-based assays may indicate an aromatisable androgen and a positive Steroidogenesis Assay could indicate a chemical that alters endogenous estrogen levels, both situations may give a positive result in the immature rat UT assay). Other (non-E,A,T,S) mechanisms may also be considered (e.g. involving other receptors or endocrine axes).

666. Scenarios D to F represent positive results in the UT assay in the presence of negative *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. Negative results in the *in vitro* assays should be viewed with caution in case a metabolite is responsible for the positive UT assay. Unless the metabolic profile of the test substance is known, then one option may be to conduct these assays with an added metabolising system. If the metabolic profile is known, then a higher level *in vivo* test may be advisable. The choice of tests will depend on the available *in vivo* effects data. Positive results in the UT assay may also indicate similar (anti)estrogenicity in lower vertebrates. As in scenarios A-C, *in vivo* assays/tests with negative results should be interpreted with caution as they may either indicate that the tests used do not have sufficient power to detect weak effects or, alternatively, that the effects do not present a concern for endocrine disruption. Selection of the dose level and the strain of animal should also be considered. Generally, a conclusion of lack of concern for endocrine disruption in the presence of positive effects data (Scenario E) may only be made given adequate Level 5 assays. Information on some endocrine-related tumours may be detected more comprehensively in carcinogenicity studies (OECD TG 451/453) (Level 4); for example, detection of certain types of thyroid tumors in the absence of reproductive or developmental effects, as well as substances causing tumors in other endocrine-sensitive tissues. MOA data to provide a clear interpretation may be required by some regulatory agencies.

667. Scenarios G to I represent positive results in the UT assay in the presence of various combinations of missing or equivocal data. Positive results in the UT assay may also indicate similar (anti)estrogenicity in lower vertebrates. The next step to take in these eventualities will depend on the nature of the other available data and the jurisdiction in which it is being used. In some cases, equivocal data may be viewed as positive, whilst in others it may or may not contribute to the weight of evidence. The interpretation may also depend on the MOA in question and why the data are considered equivocal, e.g. a study that is equivocal for thyroid effects may still be of value in evaluating (anti)androgenic effects. In all three scenarios, the recommended first step is to obtain reliable mechanistic (*in vitro*) data rather than proceed further with *in vivo* testing. Equivocal and missing data are alternative scenarios and two possibilities for the next step are given in most cases, but the nature of equivocal data means that decisions need to be taken on a case-by-case basis. In all cases, the role of metabolism, route of exposure and data from structural analogues should be considered before deciding on the next step. As above, generally a conclusion of lack of concern for endocrine disruption in the presence of some positive effects data (Scenario H) may only be made given adequate Level 5 assays. Information on some endocrine-related tumours may be detected more comprehensively in carcinogenicity studies (OECD TG 451/453) (Level 4); for example, detection of certain types of thyroid tumors in the absence of reproductive or developmental effects, as well as substances causing tumors in other endocrine-sensitive tissues. MOA data to provide a clear interpretation may be required by some regulatory agencies.

668. Scenarios J to L represent negative results in the UT assay in the presence of positive *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. The *in vitro* mechanistic data given in the table could be any of the E,A,T,S tests (e.g. the AR binding or Steroidogenesis Assay). A weak aromatase inhibitor, for example, could give Scenario J from a positive result in the Steroidogenesis Assay and a positive result in the female Peripubertal Assay. All three scenarios could also arise from a chemical that binds to ER but is metabolised to a non-estrogenic metabolite leading to negative results in the UT assay and this should be considered first when investigatingthe next step. Endocrine active potency may also explain differences between *in vitro* and *in vivo* results (e.g. a chemical with weak endocrine activity may give a positive result *in vitro* but may be negative *in vivo*). Positive *in vivo* effects data may involve other E,A,T,S, non-E,A,T,S mechanisms (e.g. involving other receptors or endocrine axes), more sensitive endpoints, greater statistical power or life stages that are more sensitive to the substance than immature/ovariectomised animals in the UT assay.

669. Scenarios M to O represent negative results in the UT assay in the presence of negative *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. Negative results for all tests (Scenario N) may be sufficient to enable a conclusion of no concern for endocrine disruption. This will depend on the weight of evidence and may not be possible in some cases. However, in the presence of negative data from robust Level 4 and 5 assays, further animal testing is probably not justified. Where there are positive *in vivo* effects data, there could still be an estrogen-related mechanism. These effects may be related to length of exposure, route of exposure or exposure at different life stages. Other E,A,T,S or non-E,A,T,S mechanisms may also be involved.

Scenarios P to R represent negative results in the UT assay in the presence of various combinations of missing or equivocal data. As with the positive result scenarios above (see Paragraph 665), the next step to take in these eventualities will have to be decided on a case-by-case basis. However, the recommended first step is generally to obtain reliable mechanistic (*in vitro*) data rather than proceed directly to *in vivo* testing. In all cases, the role of metabolism, route of exposure and data from structural analogues should be considered before deciding on the next step.

670. In all scenarios (A-R), the next step to take to strengthen weight of evidence will depend on the existing information. Table C.3.1 is meant to provide a succinct guide and may not cover all circumstances or possibilities. The scenarios may also suggest that chemicals have simple or single MOA, when in practice they may have multiple endocrine and non-endocrine MOA. In some cases, for example, two opposite modes of simultaneous action (e.g. estrogenic and anti-estrogenic) could, depending on dose, lead to a minimisation or abolition of effects, while in others two different MOA (e.g. estrogenic and anti-androgenic) could potentially reinforce effects. Endocrine pathways interact, mixed effects are common and there are many pathways that cannot be distinguished with currently available TGs. If multiple MOA are suspected, either from the existing results or based on QSAR/read-across/integrated approaches, this should be investigated further if needed for regulatory decision making.

References

Jacobs, M. et al. (2013), "*In vitro* metabolism and bioavailability tests for endocrine active substances: What is needed next for regulatory purposes?", *ALTEX – Alternatives to Animal Experimentation*, Vol. 30/3, pp. 331-351.

Jacobs, M.N. et al. (2008), "The use of metabolising systems for *in vitro* testing of endocrine disrupters", *Current Drug Metabolism*, Vol. 9/8, pp. 796-826.

Jones, R.C. and R.A. Edgren (1973), "The effects of various sterols on the vaginal histology in the rat", *Fertility and Sterility*, Vol. 24, pp. 284-291.

Kleinstreuer, N.C. et al. (2017), "Development and validation of a computational model for androgen receptor activity", *Chemical Research in Toxicology*, Vol. 30/4, pp. 946-964, http://dx.doi.org/10.1021/acs.chemrestox.6b00347.

Laws, S.C. et al. (2000), "Estrogenic activity of octylphenol, nonylphenol, bisphenol A and methoxychlor in rats", *Toxicological Sciences*, Vol. 54/1, pp. 154-167, https://doi.org/10.1093/toxsci/54.1.154.

OECD (2008), *Detailed Review Paper on the Use of Metabolising Systems for* In Vitro *Testing of Endocrine Disrupters*, OECD Series on Testing and Assessment, No. 97, OECD Publishing, Paris, https://doi.org/10.1787/9789264085497-en.

OECD (2007), "Guidance document on the procedure to test for antioestrogenicity", OECD Series on Testing and Assessment, No. 71, OECD, Paris, www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2007)15&doclanguage=en.

OECD (2006), "Detailed review paper on thyroid hormone disruption assays", OECD Series on Testing and Assessment, No. 57, OECD, Paris, www.oecd.org/env/ehs/testing/37235405.pdf.

Owens, J.W. and J. Ashby (2002), "Critical review and evaluation of the uterotrophic bioassay for the identification of possible estrogen agonists and antagonists: In support of the validation of the OECD uterotrophic protocols for the laboratory rodent", *Critical Reviews in Toxicology*, Vol. 32/6, pp. 445-520, http://dx.doi.org/10.1080/20024091064291.

WHO/IPCS (2002), "Global assessment of the state-of-the-science of endocrine disrupters", Damstra, T. et al. (eds.) WHO/PCS/EDC/02.2, World Health Organization, Geneva, www.who.int/ipcs/publications/new_issues/endocrine_disruptors/en.

Table C.3.1. **Uterotrophic Bioassay in Rodents (UT assay) (OECD TG 440) (including OECD GD 71 on the Procedure to Test for Anti-estrogenicity): Guidance for scenarios of combinations of results with existing data**

This table represents possible conclusions to be drawn from assay data, and a next step which could be taken if further evidence is required about possible endocrine disrupting properties and/or effects. The guidance offered is not meant to be prescriptive, but provides science-based considerations. It encourages the use of all available data and expert judgement in a weight of evidence approach. Regional and national interpretation of results and "next steps" may vary.

The conclusions are grouped into a series of scenarios (A-R), each scenario representing a different combination of assay results, existing *in vitro* data and existing *in vivo* data. The symbol "+" indicates that the data in question represent a positive result, «-" indicates a negative result, and "Eq/0" indicates that the data are either equivocal or are not available.

Existing results: * "Mechanism (*in vitro* mechanistic data)" assumes that mechanistic data are available from estrogen receptor (ER-), androgen receptor (AR-) and steroidogenesis-based assays (Level 2). Thyroid hormone receptor (TR) and other assays concerning mechanisms of thyroid disruption may be available, but they are not in common use. In practice, data from all assays may not be available and therefore this must be taken into account when deciding on the "next step". Quantitative structure activity relationship (QSAR) predictions of estrogen and androgen binding/activation may be made for some substances.

Existing results: ** "Effects (*in vivo* effects of concern)" assumes effects have been observed in other *in vivo* screens/tests which give rise to concern that the test chemical may be an endocrine disrupter. These may be repeated dose toxicity tests (e.g. OECD TG 407, TG 408 28-day and 90-day studies), reproductive tests (e.g. reproduction screening assays or two-generation studies) or read-across from chemical analogues.

Scenarios	Result of OECD TG 440 (UT assay)	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
A	+	+	+	Strong evidence for estrogenic/anti-estrogenic (E/anti-E) activity with (potential for) adverse effects via estrogen receptor (ER) mechanism.	Perform assay from Level 4 (e.g. female pubertal assay) or Level 5 (e.g. Extended One-Generation Reproduction Toxicity Study [EOGRTS] or two-generation) assay.	If existing data are from Level 4 or 5 (or less sensitive) assays, there may be sufficient information to conclude evidence of concern for endocrine disruption (the EOGRTS provides the most information; however, for endocrine disrupting chemicals [EDCs] with a carcinogenic potential, OECD TG 451-3 may be more sensitive). Consider route of exposures for UT assay and existing effects data and possible implications of absorption, distribution, metabolism and excretion (ADME) characteristics of the chemical. The chemical itself may give positive *in vitro* results and *in vivo* results, but may also be metabolised to a metabolite that also has positive results *in vitro* and *in vivo*. A positive result could have arisen from other (E,A,T,S or non-E,A,T,S) mechanisms. E/anti-E activity possible in lower vertebrates. Consider performing the Fish Sexual Development Test (FSDT) or Larval Amphibian Growth and Development Assay (LAGDA).
B	+	+	–	Strong evidence for E/anti-E activity via ER but effects not detected in other *in vivo* studies in intact animals.	Perform assay from Level 4 (e.g. female pubertal assay) or Level 5 (e.g. EOGRTS or two-generation) assay.	If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from a less sensitive assay, a higher level test may be required. Consider route of exposures for UT assay and existing effects data and possible implications of ADME characteristics of the chemical. Check data on chemical analogues. A positive result could have arisen from other (E,A,T,S or non-E,A,T,S) mechanisms. E/anti-E activity possible in lower vertebrates. Consider performing an FSDT or LAGDA.
C	+	+	Eq/0	Strong evidence for E/anti-E activity via ER, but no or equivocal data from other *in vivo* studies.	Perform assay from Level 4 (e.g. female pubertal assay) or Level 5 (e.g. EOGRTS or two-generation) assay.	Check data on chemical analogues. Consider route of exposures for UT assay and existing effects data and possible implications of ADME characteristics of the chemical. Depending on route/kinetic and existing data considerations, may perform assay from Levels 4 or 5. A positive result could have arisen from other (E,A,T,S or non-E,A,T,S) mechanisms. E/anti-E activity possible in lower vertebrates. Consider performing an FSDT or LAGDA. Equivocal results may indicate chemical has multiple modes of action (MOA).

Scenarios	Result of OECD TG 440 (UT assay)	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
D	+	–	+	Strong evidence for E/anti-E activity. Acts via ER mechanism, but requires metabolic activation. Acts via non-ER mechanism and may or may not require metabolic activation.	Perform ER transactivation assay or binding assay with added metabolising system.	If existing data are from Level 4 or 5 (or less sensitive) assays, there may be sufficient information to conclude evidence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). Further mechanistic studies may help determine MOA. A positive result could have arisen from other (E,A,T,S or non-E,A,T,S) mechanisms e.g. hypothalamic/pituitary/gonadal (HPG) axis. E/anti-E activity possible in lower vertebrates. Consider performing an FSDT or LAGDA.
E	+	–	–	Weak evidence for E/anti-E activity. Acts via non-ER mechanism. Chemical requires metabolic activation and metabolite has weak activity. Weak E/anti-E activity via ER does not result in adverse effects.	Perform ER transactivation assay or binding assay with added metabolising system OR Perform assay from Levels 4 or 5.	If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from a less sensitive assay, a higher level test may be required. Consider route of exposures for UT assay and existing effects data and possible implications of ADME characteristics of the chemical. Check data on chemical analogues A positive result could have arisen from other (E,A,T,S or non-E,A,T,S) mechanisms, e.g. HPG axis. E/anti-E activity possible in lower vertebrates. Consider performing an FSDT or LAGDA.
F	+	–	Eq/0	Weak evidence for E/anti-E activity via ER. Acts via non-ER mechanism. Requires metabolic activation and metabolite has weak/equivocal activity.	Perform ER transactivation assay or binding assay with added metabolising system OR Perform assay from Levels 4 or 5.	Check data on chemical analogues. Further mechanistic studies may help determine MOA. Level 4 or 5 studies will provide hazard data. A positive result could have arisen from other (E,A,T,S or non-E,A,T,S) mechanisms, e.g. HPG axis. E/anti-E activity possible in lower vertebrates. Consider performing an FSDT or LAGDA. Equivocal results may indicate chemical has multiple MOA.
G	+	Eq/0	+	Moderate or strong evidence for E/anti-E activity via ER. May act via ER, metabolic activation is required. Has potential for adverse effects via ER mechanism. May acts via non-ER mechanism and may or may not require metabolic activation.	For the "0" scenario, perform ER transactivation assay or binding assay. For the "Eq" scenario, perform ER transactivation assay or binding assay with added metabolising system.	If existing data are from Level 4 or 5 (or less sensitive) assays, there may be sufficient information to conclude evidence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). Check data on chemical analogues. Further mechanistic studies may help determine MOA. A positive result could have arisen from other (E,A,T,S or non-E,A,T,S) mechanisms, e.g. HPG axis. E/anti-E activity possible in lower vertebrates. Consider performing an FSDT or LAGDA. Equivocal results may indicate chemical has multiple MOA.

Scenarios	Result of OECD TG 440 (UT assay)	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
H	+	Eq/0	–	Weak evidence for E/anti-E activity. May act via ER, metabolic activation is required. E/anti-E activity does not result in adverse effects.	For the "0" scenario, perform ER transactivation assay or binding assay. For the "Eq" scenario, perform ER transactivation assay or binding assay with added metabolising system.	If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from a less sensitive assay, a higher level test may be required. Consider route of exposures for UT assay and existing effects data and possible implications of ADME characteristics of the chemical. Check data on chemical analogues. Further mechanistic studies may help determine MOA. E/anti-E activity possible in lower vertebrates. Consider performing an FSDT or LAGDA. Equivocal results may indicate chemical has multiple MOA.
I	+	Eq/0	Eq/0	E/anti-E activity of unknown potency. May act via ER, metabolic activation is required. Unknown potential for adverse effects.	For the "0" scenario, perform ER transactivation assay or binding assay. For the "Eq" scenario, perform ER transactivation assay or binding assay with added metabolising system, or Level 4 or 5 assay if existing data indicate this is needed.	Check data on chemical analogues. Further mechanistic studies may help determine MOA. E/anti-E activity possible in lower vertebrates. Consider performing an FSDT or LAGDA. Equivocal results may indicate chemical has multiple MOA.
J	–	+	+	No evidence for E/anti-E activity *in vivo* via ER. Route of exposure, metabolic differences or potency explain differences between UT assay and existing *in vitro*/*in vivo* studies. Effects seen in existing studies are via non-ER mechanism.	Perform ER transactivation assay or binding assay with added metabolising system.	If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). Effects seen in existing studies may be in a more sensitive life stage. Consider route of exposure for UT assay and possible implications of ADME characteristics of the chemical. Check data on chemical analogues. Further mechanistic studies may help determine MOA.
K	–	+	–	No evidence for E/anti-E activity *in vivo* via ER. Metabolic differences or potency explain *in vitro*/*in vivo* differences.	Perform ER transactivation assay or binding assay with added metabolising system.	If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from a less sensitive assay, a higher level test may be required. Metabolic deactivation of chemical may occur *in vivo* so that possible *in vitro* activity is not realised. Consider possible routes of exposure implications of metabolism.

Scenarios	Result of OECD TG 440 (UT assay)	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
L	–	+	Eq/0	No evidence for E/anti-E activity *in vivo* via ER. Metabolic differences or potency explain *in vitro*/*in vivo* difference. Unknown potential for adverse effects.	Perform ER transactivation assay or binding assay with added metabolising system OR Perform assay from Levels 4 or 5.	Metabolic deactivation of chemical may occur *in vivo* so that potential *in vitro* activity is not realised. Consider possible routes of exposure implications of metabolism. Equivocal results may indicate chemical has multiple MOA.
M	–	–	+	No evidence for E/anti-E activity *in vivo* or *in vitro* via ER. Metabolic differences or route of exposure explain *in vitro*/*in vivo* existing differences. Effects seen in existing studies are via non-ER mechanism.	Perform *in vitro* assays with added metabolising system.	Effects seen in existing studies may be in a more sensitive life stage. Check data on chemical analogues. Further mechanistic studies may help determine MOA.
N	–	–	–	No evidence for E/anti-E activity *in vivo* or *in vitro* via ER. No evidence of adverse effects.	Possibly no need for further testing. If there is uncertainty, may perform assay from Level 4.	If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). Check data on chemical analogues.
O	–	–	Eq/0	No evidence for E/anti-E activity *in vivo* or *in vitro* via ER. Unknown potential for adverse effects via other non-ER mechanisms.	Perform assay from Levels 4 or 5.	Consider route of exposure for UT assay and possible implications for ADME characteristics of the chemical in follow-up assay.
P	–	Eq/0	+	No evidence for E/anti-E activity *in vivo* via ER. Unknown potential for adverse effects via other mechanisms.	For the "0" scenario, perform *in vitro* E,A,T,S assays, otherwise Eq result available.	Consider route of exposure for UT assay and possible implications for differences from existing assay. Effects seen in existing studies may be in a more sensitive life stage. Check data on chemical analogues. Further mechanistic studies may help determine MOA. Equivocal results may indicate chemical has multiple MOA.
Q	–	Eq/0	–	No evidence for E/anti-E activity *in vivo* via ER. No evidence of adverse effects.	For the "0" scenario, perform *in vitro* E,A,T,S assays, otherwise Eq result available.	If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from a less sensitive assay, a higher level test may be required. Check data on chemical analogues. Further mechanistic studies may help determine MOA.
R	–	Eq/0	Eq/0	No evidence for E/anti-E activity *in vivo* via ER.	For the "0" scenario, perform *in vitro* E,A,T,S assays, otherwise Eq result available.	Consider route of exposure for UT assay and possible implications for differences from existing assay. Check data on chemical analogues. Further mechanistic studies may help determine MOA. Equivocal results may indicate chemical has multiple MOA.

C.3.2. Hershberger Bioassay in Rats (H assay) (OECD TG 441) (including OECD GD 115 on the Weanling Hershberger Bioassay)

Status: Assay validated by the OECD.

671. Modality detected/endpoints: androgens (weights of ventral prostate, seminal vesicles, LABC [levator ani plus bulbocavernosus muscle complex], cowpers glands, glans penis ↑); anti-androgens (weights of testosterone stimulated ventral prostate, seminal vesicles, LABC, cowpers glands, glans penis ↓); optional others (e.g. liver, paired kidney, paired adrenal and testis weights, changes in serum hormones including thyroid hormones). *Note:* weanling H assay does not include glans penis.

Background to the assay

672. This assay is a short-term *in vivo* screening assay in male rats for chemicals that interact with the androgen receptor (AR) and chemicals that inhibit the enzyme 5-alpha reductase. Route of administration of test substance is via oral gavage or subcutaneous injection. It is based on changes in weight of the accessory tissues of the male reproductive tract in response to androgens and anti-androgens in animal models where endogenous androgens are minimal as a result of castration or because the animals are immature. The surgically castrated peripubertal rat is the primary model validated for the assay and is described in OECD TG 441 (adopted in September 2007). This model is sensitive to androgens and anti-androgens. An alternative model – the intact (uncastrated) weanling rat – was also validated due to animal welfare concerns with the castration procedure, but did not seem to consistently detect weak anti-androgenic chemicals at the doses tested, although androgenic chemicals were detected. The castrated peripubertal model is therefore more commonly used because both androgenic and anti-androgenic protocols can be run in the same experiment. The use of the weanling H assay is described in a guidance document (OECD, 2009). The castrated peripubertal rat model utilises the weights of five androgen-dependent sex accessory tissues (ventral prostate, seminal vesicles, LABC, cowpers glands and glans penis) as the primary endpoints, whilst for the weanling rat model the list does not include the glans penis because the weanling male has not yet achieved preputial separation. Testis weight is an optional endpoint in the weanling model, although it should be noted that the weight changes with androgens and anti-androgens are opposite to those seen with the other sex accessory tissues. Serum hormone levels are also optional for both models. These include the thyroid hormones (T3 and T4) so that additional information on thyroid effects may also be obtained, and luteinising hormone (LH), follicle stimulating hormone (FSH) and testosterone.

673. The castrated peripubertal rat does not have an intact hypothalamic/pituitary/gonadal (HPG) axis and therefore chemicals acting through this mechanism will not be detected. The HPG axis in the weanling rat is intact and therefore it is possible that such chemicals may be detected. In practice, this has not been tested and the immaturity of the animals, plus the co-administration of testosterone in the anti-androgen test, makes this unlikely.

674. Although this assay is a "screen", some authorities may regard a decrease in the weight of sex accessory tissues as possibly adverse; for example OECD GD 43 (OECD, 2008c) states that "a significant change in absolute testis weight (increase or decrease) can indicate an adverse effect". If this occurs in immature animals at a time when this should not occur naturally, then this could represent an adverse effect in a sensitive life stage. Likewise, the castrated H assay may be regarded as a model for immature animals and therefore a decrease sex accessory tissue weights could be regarded as adverse. Interpretations of the results of this assay may vary according to region and regulation, and should always utilise all data in a weight of evidence approach. Androgenic chemicals cause growth of the sex accessory tissues whilst anti-androgenic chemicals inhibit the growth caused by co-administration of testosterone. Anti-androgens may act either via AR antagonism (e.g. flutamide) or they may act via inhibition of the enzyme 5-alpha reductase (e.g. finasteride), which converts testosterone to the more potent dihydrotestosterone. 5-alpha reductase inhibitors may be distinguished from AR antagonists in the H assay by a more pronounced effect on the ventral prostate. AR antagonists can also be distinguished from 5-alpha reductase inhibitors by the use of *in vitro* assays as 5-alpha reductase inhibitors do not generally interact with AR. At present there are no validated assays for 5-alpha reductase inhibition although literature methods are available (Lo et al., 2007).

675. The growth of the sex accessory tissues may not always be entirely of androgenic origin. High doses of other hormones may give similar responses (e.g. potent estrogens may increase the weight of seminal vesicles). Chemicals affecting steroid metabolism could also conceivably affect the anti-androgen assay.

When/why the assay may be used

676. Although OECD TG 441 can be used at any stage in the hazard assessment process, the most likely use scenario will be following a positive result in an AR transactivation assay or AR Binding Assay, in order to determine whether the positive result *in vitro* is translated into a positive result *in vivo*. It may also be used as a screen in the absence of positive *in vitro* data, when a chemical that is negative in the *in vitro* AR-interaction screens is suspected of producing androgenic metabolites *in vitro*. In this case, the first option would be to use an additional metabolising system in the *in vitro* tests, but the H assay as an *in vivo* assay will include all metabolising systems. Another possible scenario is following observation of effects in higher tier tests, for example delayed puberty onset in males, but which are not exclusively indicative of an effect on AR. In the European Union, chemicals included in REACH, Plant Protection Products and Biocides legislation may have been tested in OECD TG 421/422, TG 416 (Two-Generation Reproductive Toxicity Study) or the Extended One-Generation Reproduction Toxicity Study (EOGRTS – OECD TG 443), the H assay may then be used as a follow up to clarify the mode of action (MOA). The H assay is also likely to be carried out as part of the United States Environmental Protection Agency's Endocrine Disruptor Screening Program (US EPA EDSP) Tier 1 screening battery. The castrated peripubertal rat assay (as described in OECD TG 441) is mandatory for the US EPA EDSP Tier 1 screening battery and is most likely to be the assay of choice in other testing strategies. Selection of the most appropriate assays has to be on a case-by-case basis, but also considering the need to minimise animal testing.

677. It should be noted that the H assay was designed to be sensitive and will detect weak and strong AR modulators and 5-alpha reductase inhibitors. In the validation of the H assay, trenbolone acetate and testosterone were defined as "potent" androgens whilst finasteride was a "potent" anti-androgen. Linuron and vinclozolin were defined as "weak" anti-

androgens (OECD, 2008b), but no weak androgens were tested. Weakly acting chemicals may not always be detected as endocrine disruptors (EDs) when tested in higher level tests because the endocrine system in intact/adult animals has a greater ability to compensate than in the H assay where the HPG axis is disrupted/immature and in the case of repeat dose studies dose levels may need adjustment to lower doses in order to cope with general toxicity.

678. The route of exposure is also an important consideration for the H assay. OECD TG 441 states that the test substance may be administered by oral or subcutaneous routes, but suggests that the route most relevant for human exposure should be used. The route will have consequences for absorption, distribution, metabolism and excretion (ADME) and is an important consideration when interpreting results.

679. In order to provide information relevant for assessing whether or not a chemical may fulfil the WHO/IPCS (2002) definition of an ED, the study design has to be sufficiently robust to demonstrate the presence or absence of effects. In the dose selection, the investigator should also consider and ensure that data generated are adequate to fulfil the regulatory requirement across OECD countries as appropriate (e.g. hazard and risk assessment and labelling, ED assessment, etc.). The top dose or concentration should be sufficiently high to give clear systemic (i.e. non endocrine-specific) toxicity in order to ensure that a wide range of exposures (high to low) is tested. However, endocrine effects observed solely in the presence of clear systemic toxicity should be interpreted with caution and may be disregarded when sufficiently justified to be caused by secondary effects which are unlikely to be due to endocrine activity. The reason for this advice is a concern that some endocrine active substance (EAS) sensitive assays are being run at doses/concentrations of EASs that are too low to trigger direct impacts on the endocrine system. This guidance document (GD) is not the place to address this issue directly, but it should be considered when EAS-sensitive test guidelines (TGs) are revised in the future. In addition, the number and spacing of dose/concentration levels should also be adequate to fulfil the objectives of the study (e.g. to demonstrate dose response relationships if this is required).

Introduction to the table of scenarios

680. Table C.3.2 gives guidance on a further step to take in the event of a positive (+) or negative (-) result in the H assay and in the presence of positive (+), negative (-) or equivocal/absent (Eq/0) existing results. “Existing results” are subdivided into “mechanism” and “effects” data (third and fourth columns). The table is divided horizontally into a series of scenarios that represent all the combinations of these events.

681. The results of OECD TG 441 are given in the second column. Criteria for positive results in OECD TG 441 are given in the test guideline itself, i.e. a statistically significant increase (agonism) or decrease (antagonism or 5-alpha reductase inhibition) in weights of two or more of the sex accessory tissues compared to the relevant control and all target tissues showing some change in the relevant direction. In the case of agonists, the control is only treated with vehicle for the test substance whilst for antagonists and 5-alpha reductase inhibitors, the control is treated with testosterone plus vehicle for the test substance. Negative results are no (statistically significant) changes in weights of the sex accessory tissues compared to the relevant control. Single, isolated changes would also be considered negative. The guideline suggests that combined evaluation of all sex accessory tissue responses could be achieved using appropriate multivariate data analysis. It is important that quality criteria (coeffecients of variation) for the weights of control sex

accessory tissues are demonstrated. Details are given in the test guideline. Note that in the weanling assay, testis weight decreases with agonists and increases with antagonists. Details of the criteria for positive results in this assay are given in the GD (OECD, 2009).

682. Optional endpoints may include measurement of serum LH, FSH or testosterone. These endpoints should supplement the sex accessory tissue weights and the assay should not be considered to be positive if changes in these endpoints occur in the absence of weight changes in the primary tissues. In addition, serum T3 and T4 levels may provide useful information on possible effects on the thyroid, although measurement of thyroid weight and serum TSH levels would be also useful in this case. They are not considered further here as this is not the primary use of the assay. Measurement of serum testosterone may be useful if induction of liver xenobiotic metabolising enzymes is suspected. Experience with of serum hormone determinations in rodent assays has revealed that their detection/measurement in rodent studies can be challenging. A recent workshop on "Practicability of Hormonal Measurements" was organised by the BfR (Germany) and the finding from this workshop will be published (Kucheryavenko et al., 2018). The optional endpoint of liver weight would also be very useful. In these cases, increased clearance of testosterone may lead to an apparent anti-androgenic effect on the sex accessory tissues that does not result from interaction with AR.

683. Equivocal results for the guideline are not included in the table because these data require further interrogation about the result itself. This assay is a screen and therefore a clear positive or negative result should be obtained. In the event of an equivocal result, the considerations mentioned above about control sex accessory tissue weights, non AR-related changes should be taken into account and further investigations made.

Existing data to be considered

684. Existing "mechanism" *in vitro* data are assumed to be available from estrogen receptor (ER-), AR- and steroidogenesis-based assays (Level 2). Assays may also be available for interference with thyroid modalities. In practice, it is possible that data from all of these assays may not be available, so judgement will need to be used to decide which assays to perform. As noted above, there is no validated assay available for 5-alpha reductase inhibitors at present and although 5-alpha reductase is present in H295R cells used in the Steroidogenesis Assay, the assay does not include the required endpoint for this (dihydrotestosterone). Although the current *in vitro* test guidelines do not incorporate metabolic activation, published information on use of metabolic activation systems is available (Jacobs et al., 2008; 2013) as is an OECD detailed review paper (OECD, 2008a). These methods, however, have not yet been validated.

685. Existing "effects" data refer to *in vivo* effects that may come from varied sources and will depend on the type of substance (e.g. new chemicals, high production volume chemical, pesticide). Thus, available data may range from repeated dose toxicity studies (28-day, 90-day) or combined repeat dose/reproductive screening assays to chronic toxicity studies and multigeneration reproductive tests. Some studies fail to identify endocrine disruptors (EDs) that weakly affect estrogen or androgen receptors as was demonstrated on the basis of data generated in the validation process of OECD TG 407 assay with endocrine endpoints. In this validation, only moderate EDs such as nonylphenol and DDE, and strong EDs such as ethinylestradiol and flutamide (acting via ER and AR respectively), were detected. Thus, OECD TG 407 cannot be regarded as a screening assay for endocrine activity. This means that when a relatively insensitive test is positive for both endocrine-specific and apical endpoints, this should be taken as an indication that the substance is a

potential ED. Caution should be exercised, however, because endocrine endpoints may be impacted secondary to non-endocrine toxicity and *in vivo* apical endpoints can be affected by all MOA, including endocrine modalities. If data are available from single or multigeneration studies that are adequately conducted with updated guidelines that include endpoints sensitive to EDs, then there should be no reason to conduct an H assay as the higher tier test will provide stronger evidence for hazard identification/characterisation. Multigeneration studies conducted prior to the introduction of these endpoints will still provide valuable information on reproductive and endocrine organ toxicity, reproduction and development, but may not be sufficiently sensitive to endocrine active substances (EASs), in which case the H assay would provide further valuable information. A decision about whether to conduct further animal tests would, however, need to consider whether sufficient supplementary data may be provided by *in vitro* tests. Data may also be available on effects in mammalian and non-mammalian wildlife species, although caution should be used when extrapolating between taxa. A chemical causing endocrine effects in non-mammalian environmental species (fish, for example) may also have endocrine effects in mammals, but the physiological consequences of the effects are likely to be different.

686. When considering the results of the H assay, all available data should be used in order to reach a conclusion and a weight of evidence approach taken. This may include high throughput screening data, read-across data from structural analogues and quantitative structure activity relationships (QSARs). Several QSAR models for ER and AR binding/activation are now available (see Sections B.1.1.1 and B.1.1.2).

Scenarios: Positive and negative results combined with existing data

687. The scenarios (A to R) presented in Table C.3.2 represent all the possibilities of positive or negative results in combination with the presence or absence of existing data. The action taken will also depend on the regulatory environment, but the considerations given here are generally science based. Although OECD TG 441 uses rats, the well-conserved nature of AR across taxa should be a strong indication that results in this assay are relevant to other vertebrate species. Results in laboratory mammal tests are also highly relevant for environmental mammalian species. Wherever possible, the recommended "next step which could be taken" avoids unnecessary animal testing. However, sometimes conducting an animal test will be indicated and then the relevance of species, strain, exposure route and species-specific metabolism should always be considered. The sensitivity and physiological function of the hormone under investigation in the test species should also be considered. In general, lower level tests should be conducted before higher level tests in order to avoid unnecessary animal usage, unless it is apparent that a Level 5 test will be required anyway or will be needed to establish the evidence to conclude on ED properties. Information on some endocrine-related tumours may be detected more comprehensively in carcinogenicity studies (OECD TG 451/453) (Level 4); for example, detection of certain types of thyroid tumors in the absence of reproductive or developmental effects, as well as substances causing tumors in other endocrine-sensitive tissues. At Level 5, the EOGRTS (OECD TG 443) is the most sensitive reproduction assay for detecting endocrine disruption because it includes evaluation of a number of endocrine endpoints not included in the current two-generation study (OECD TG 416) adopted in 2001. It is recognised, however, that some jurisdictions may require a two-generation study. Further considerations specific to each scenario are given in the table.

688. Scenarios A to C represent positive results in the H assay in the presence of positive *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. A positive

result in AR-based assays in combination with a positive H assay is strong evidence for (anti)androgenic activity that may or may not be supported by the *in vivo* effects data. There may be sufficient evidence to conclude concern for endocrine disruption and therefore no need for further screening. Positive results in the H assay may also indicate similar (anti)androgenicity in lower vertebrates. These could be followed up with partial life cycle tests such as the Fish Sexual Development Test (FSDT) or the Larval Amphibian Growth and Development Assay (LAGDA). *In vivo* assays/tests with negative results should be interpreted with caution as they may either indicate that the tests used do not have sufficient power to detect weak effects or, alternatively, that the effects do not present a concern for endocrine disruption. The possibility of other mechanisms should also not be overlooked, e.g. positive ER-based assays and a positive result H assay may indicate (anti)estrogenic effects. Other (non-E,A,T,S) mechanisms may also be considered (e.g. involving other receptors or endocrine axes).

689. Scenarios D to F represent positive results in the H assay in the presence of negative *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. Negative results in the *in vitro* assays should be viewed with caution in case a metabolite is responsible for the positive H assay. These scenarios may also occur if enhanced metabolism or clearance of testosterone is responsible for the positive H assay. Unless the metabolic profile of the test substance is known, then one option may be to conduct these assays with an added metabolising system. If the metabolic profile is known, then a higher level *in vivo* test may be advisable. The choice of tests will depend on the available *in vivo* effects data. Positive results in the H assay may also indicate similar (anti)androgenicity in lower vertebrates. As in Scenarios A to C, *in vivo* assays/tests with negative results should be interpreted with caution as they may either indicate that the tests used do not have sufficient power to detect weak effects or, alternatively, that the effects do not present a concern for endocrine disruption.

690. Scenarios G to I represent positive results in the H assay in the presence of various combinations of missing or equivocal data. Positive results in the H assay may also indicate similar (anti)androgenicity in lower vertebrates. The next step to take in these eventualities will depend on the nature of the other available data and the jurisdiction in which it is being used. In some cases, equivocal data may be viewed as positive whilst in others it may or may not contribute to the weight of evidence. The interpretation may also depend on the MOA in question and why the data are considered equivocal, e.g. a study that is equivocal for thyroid effects may still be of value in evaluating (anti)androgenic effects. In all three scenarios, the recommended first step is to obtain reliable mechanistic (*in vitro*) data rather than proceed further with *in vivo* testing. Equivocal and missing data are alternative scenarios and two possibilities for the next step are given in most cases, but the nature of equivocal data means that decisions need to be taken on a case-by-case basis. In all cases, the role of metabolism, route of exposure and data from structural analogues should be considered before deciding on the next step.

691. Scenarios J to L represent negative results in the H assay in the presence of positive *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. The *in vitro* mechanistic data given in the table could be any of the estrogen/androgen/thyroid/ steroidogenesis (E,A,T,S) tests (e.g. the ER binding or Steroidogenesis Assay). A weak aromatase inhibitor, for example, could give Scenario J a positive result in the Steroidogenesis Assay and a positive result in the female PP assay. All three scenarios could also arise from a chemical that binds to AR but is metabolised to a non-androgenic metabolite leading to negative results in the H assay and this should be considered first when investigating the next step. Endocrine active potency may also explain differences

between *in vitro* and *in vivo* results (e.g. a chemical with weak endocrine activity may give a positive result *in vitro* but may be negative *in vivo*). Positive *in vivo* effects data may involve other E,A,T,S, non-E,A,T,S mechanisms (e.g. involving other receptors or endocrine axes), more sensitive endpoints, greater statistical power or life stages that are more sensitive to the substance than castrated/immature animals in the H assay.

692. Scenarios M to O represent negative results in the H assay in the presence of negative *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. Negative results for all tests (Scenario N) may be sufficient to enable a conclusion of no concern for endocrine disruption. This will depend on the weight of evidence and may not be possible. Where there are positive *in vivo* effects data, there could still be an androgen-related mechanism. The effects may be related to length of exposure, route of exposure or exposure at different life stages. Other E,A,T,S or non-E,A,T,S mechanisms may also be involved.

693. Scenarios P to R represent negative results in the H assay in the presence of various combinations of missing or equivocal data. As with the positive result scenarios above (see Paragraph 692), the next step to take in these eventualities will have to be decided on a case-by-case basis. However, the recommended first step is generally to obtain reliable mechanistic (*in vitro*) data rather than proceed directly to *in vivo* testing. In all cases, the role of metabolism, route of exposure and data from structural analogues should be considered before deciding on the next step.

694. In all scenarios (A to R), the next step to take to strengthen weight of evidence will depend on the existing information. Table C.3.2 is meant to provide a succinct guide and may not cover all circumstances or possibilities. The scenarios may also suggest that chemicals have simple or single MOA, when in practice they may have multiple endocrine and non-endocrine MOA. In some cases, for example, two opposite modes of simultaneous action (e.g. estrogenic and anti-estrogenic) could, depending on dose, lead to a minimisation or abolition of effects, while in others two different MOA (e.g. estrogenic and anti-androgenic) could potentially reinforce effects. Endocrine pathways interact, mixed effects are common and there are many pathways that cannot be distinguished with currently available TGs. If multiple MOA are suspected, either from the existing results or based on QSAR/read-across/integrated approaches, this should be investigated further if needed for regulatory decision making.

References

Jacobs, M. et al. (2013), "*In vitro* metabolism and bioavailability tests for endocrine active substances: What is needed next for regulatory purposes?", *ALTEX – Alternatives to Animal Experimentation*, Vol. 30/3, pp. 331-351.

Jacobs, M.N. et al. (2008), "The use of metabolising systems for *in vitro* testing of endocrine disrupters", *Current Drug Metabolism*, Vol. 9/8, pp. 796-826.

Kucheryavenko, O. et al. (2018), "Report from the BfR Expert Hearing on Practicability of Hormonal Measurements", *Archives of Toxicology* (in prep.).

Lo, S. et al. (2007), "Effects of various pesticides on human 5alpha reductase activity in prostate and LNCaP cells", *Toxicology In Vitro*, Vol. 21/3, pp. 502-508, https://doi.org/10.1016/j.tiv.2006.10.016.

OECD (2009), "Guidance document on the Weanling Hershberger Bioassay in Rats: A short-term screening assay for (anti)androgenic properties", OECD Series on Testing and Assessment, No. 115, OECD, Paris, www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2009)41&doclanguage=en.

OECD (2008a), *Detailed Review Paper on the Use of Metabolising Systems for* In Vitro *Testing of Endocrine Disrupters*, OECD Series on Testing and Assessment, No. 97, OECD Publishing, Paris, https://doi.org/10.1787/9789264085497-en.

OECD (2008b), "Report of the OECD validation of the Rodent Hershberger Bioassay: Phase 2: Testing of androgen agonists, androgen antagonists and a 5 α-Reductase inhibitor in dose response studies by multiple laboratories", OECD Series on Testing and Assessment, No. 86, OECD, Paris, www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2008)3&doclanguage=en.

OECD (2008c), "Guidance document on mammalian reproductive toxicity testing and assessment", OECD Series on Testing and Assessment, No. 43, OECD, Paris, www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/JM/MONO(2008)16&doclanguage=en.

WHO/IPCS (2002), "Global assessment of the state-of-the-science of endocrine disrupters", Damstra, T. et al. (eds.) WHO/PCS/EDC/02.2, World Health Organization, Geneva, www.who.int/ipcs/publications/new_issues/endocrine_disruptors/en.

Table C.3.2. **Hershberger Bioassay (H assay) (OECD TG 441) (including OECD GD 115 on the Weanling Hershberger Bioassay): Guidance for scenarios of combinations of results with existing data**

This table represents possible conclusions to be drawn from assay data, and a next step which could be taken if further evidence is required about possible endocrine disrupting properties and/or effects. The guidance offered is not meant to be prescriptive, but provides science-based considerations. It encourages the use of all available data and expert judgement in a weight of evidence approach. Regional and national interpretation of results and "next steps" may vary.

The conclusions are grouped into a series of scenarios (A-R), each scenario representing a different combination of assay results, existing *in vitro* data and existing *in vivo* data. The symbol "+" indicates that the data in question represent a positive result, «-" indicates a negative result, and "Eq/0" indicates that the data are either equivocal or are not available.

Existing results: * "Mechanism (*in vitro* mechanistic data)" assumes that mechanistic data are available from estrogen receptor (ER-), androgen receptor (AR-) and steroidogenesis-based assays (Level 2). Thyroid hormone receptor (TR) and other assays concerning mechanisms of thyroid disruption may be available, but they are not in common use. In practice, data from all assays may not be available and therefore this must be taken into account when deciding on the "next step". Quantitative structure activity relationship (QSAR) predictions of estrogen and androgen binding/activation may be made for some substances.

Existing results: ** "Effects (*in vivo* effects of concern)" assumes effects have been observed in other *in vivo* screens/tests which give rise to concern that the test chemical may be an endocrine disrupter. These may be repeated dose toxicity tests (e.g. OECD TG 407, TG 408 28-day and 90-day studies), reproductive tests (e.g. reproduction screening assays or two-generation studies) or read-across from chemical analogues.

Scenarios	Result of OECD TG 441 (H assay)	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
A	+	+	+	Strong evidence for androgen/anti-androgen (A/anti-A) activity with (potential for) adverse effects via androgen receptor (AR) mechanism. 5-alpha reductase inhibitor with (potential for) adverse effects.	Perform assay from upper levels, e.g. male pubertal assay (Level 4) OR Extended One-Generation Reproduction Toxicity Study (EOGRTS) or two-generation assay (Level 5).	Check pattern of change across sex tissues for possible 5-alpha reductase inhibition. If existing data are from Level 4 or 5 (or less sensitive) assays, there is sufficient information to conclude evidence of concern for endocrine disruption (the EOGRTS provides the most information; however, for endocrine disrupting chemicals [EDCs] with a carcinogenic potential, OECD TG 451-3 may be more sensitive). Consider route of exposures for H assay and existing effects data and possible implications of absorption, distribution, metabolism and excretion (ADME) characteristics of the chemical. The chemical itself may give positive *in vitro* results and *in vivo* results but may also be metabolised to a metabolite that also has positive results *in vitro* and *in vivo*. A positive result could have arisen from other (E,A,T,S or non-E,A,T,S) mechanisms. A/anti-A activity possible in lower vertebrates. Consider performing a Fish Sexual Development Test (FSDT) or Larval Amphibian Growth and Development Assay (LAGDA).
B	+	+	–	Strong evidence for A/anti-A activity via AR but effects not detected in other *in vivo* studies in intact animals. 5-alpha reductase inhibitor with (potential for) adverse effects but effects not detected in other *in vivo* studies in intact animals.	Perform assay from Level 4 (e.g. male pubertal assay) OR Level 5 (e.g. EOGRTS or two-generation) assay.	Check pattern of change across sex tissues for possible 5-alpha reductase inhibition. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from a less sensitive assay, a higher level test may be required. Consider route of exposures for H assay and existing effects data and possible implications of ADME characteristics of the chemical. Check data on chemical analogues. A positive result could have arisen from other (E,A,T,S or non-E,A,T,S) mechanisms. A/anti-A activity possible in lower vertebrates. Consider performing an FSDT or LAGDA.
C	+	+	Eq/0	Strong evidence for A/anti-A activity via AR, but no or equivocal data from other *in vivo* studies. 5-alpha reductase inhibitor with (potential for) adverse effects but no or equivocal data from other *in vivo* studies.	Perform assay from Levels 4 or 5 (e.g. EOGRTS or two-generation) assay.	Check pattern of change across sex tissues for possible 5-alpha reductase inhibition. Check data on chemical analogues. Consider route of exposures for H assay and existing effects data and possible implications of ADME characteristics of the chemical. Depending on route/kinetic and existing data considerations, may perform assay from Levels 4 or 5. A positive result could have arisen from other (E,A,T,S or non-E,A,T,S) mechanisms. A/anti-A activity possible in lower vertebrates. Consider performing an FSDT or LAGDA. Equivocal results may indicate chemical has multiple modes of action (MOA).

Scenarios	Result of OECD TG 441 (H assay)	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
D	+	–	+	Strong evidence for A/anti-A activity. Acts via AR mechanism, but requires metabolic activation. 5-alpha reductase inhibitor but requires metabolic activation. Acts via non-AR mechanism and may or may not require metabolic activation.	Perform AR transactivation assay or binding assay with added metabolising system.	If existing data are from Level 4 or 5 (or less sensitive) assays, there is sufficient information to conclude evidence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). Check pattern of change across sex tissues for possible 5-alpha reductase inhibition. Further mechanistic studies may help determine MOA. A positive result could have arisen from other (E,A,T,S or non-E,A,T,S) mechanisms, e.g. hypothalamic/pituitary/gonadal (HPG) axis. A/anti-A activity possible in lower vertebrates. Consider performing an FSDT or LAGDA.
E	+	–	–	Weak evidence for A/anti-A activity via AR but requires metabolic activation. 5-alpha reductase inhibitor but requires metabolic activation. Chemical requires metabolic activation and metabolite has weak activity. Weak A/anti-A activity/5-Aapha reductase inhibition does not result in adverse effects. Acts via non-AR mechanism.	Perform AR transactivation assay or binding assay with added metabolising system OR Perform assay from Levels 4 or 5.	Check pattern of change across sex tissues for possible 5-alpha reductase inhibition. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from a less sensitive assay, a higher level test may be required. Consider route of exposures for H assay and existing effects data and possible implications of ADME characteristics of the chemical. Check data on chemical analogues. A positive result could have arisen from other (E,A,T,S or non-E,A,T,S) mechanisms, e.g. HPG axis or liver enzyme induction. A/anti-A activity possible in lower vertebrates. Consider performing an FSDT or LAGDA.
F	+	–	Eq/0	Weak evidence for A/anti-A activity via AR but requires metabolic activation. 5-alpha reductase inhibitor but requires metabolic activation. Requires metabolic activation and metabolite has weak/equivocal activity. Acts via non-AR mechanism.	Perform AR transactivation assay or binding assay with added metabolising system OR Perform assay from Levels 4 or 5.	Check pattern of change across sex tissues for possible 5-alpha reductase inhibition. Check data on chemical analogues. Further mechanistic studies may help determine MOA. Level 4 or 5 studies will provide hazard data. A positive result could have arisen from other (E,A,T,S or non-E,A,T,S) mechanisms, e.g. HPG axis or liver enzyme induction. A/anti-A activity possible in lower vertebrates. Consider performing an FSDT or LAGDA. Equivocal results may indicate chemical has multiple MOA.

Scenarios	Result of OECD TG 441 (H assay)	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
G	+	Eq/0	+	Moderate or strong evidence for A/anti-A activity via AR. May require metabolic activation. 5-alpha reductase inhibitor. May require metabolic activation. Has potential for adverse effects via AR mechanism or 5-alpha reductase inhibition. May act via non-AR mechanism and may or may not require metabolic activation.	For the "0" scenario, perform AR transactivation assay or binding assay. For the "Eq" scenario, perform AR transactivation assay or binding assay with added metabolising system.	Check pattern of change across sex tissues for possible 5-alpha reductase inhibition. If existing data are from Level 4 or 5 (or less sensitive) assays, there is sufficient information to conclude evidence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). Check data on chemical analogues. Further mechanistic studies may help determine MOA. A positive result could have arisen from other (E,A,T,S or non-E,A,T,S) mechanisms, e.g. HPG axis. A/anti-A activity possible in lower vertebrates. Consider performing an FSDT or LAGDA. Equivocal results may indicate chemical has multiple MOA.
H	+	Eq/0	–	Weak evidence for A/anti-A activity. May act via AR, metabolic activation is required. 5-alpha reductase inhibitor with (potential for) adverse effects but effects not detected in other *in vivo* studies in intact animals. A/anti-A activity/5-alpha reductase does not result in adverse effects.	For the "0" scenario, perform AR transactivation assay or binding assay. For the "Eq" scenario, perform ER transactivation assay or binding assay with added metabolising system.	Check pattern of change across sex tissues for possible 5-alpha reductase inhibition. If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from a less sensitive assay, a higher level test may be required. Consider route of exposures for H assay and existing effects data and possible implications of ADME characteristics of the chemical. Check data on chemical analogues. Further mechanistic studies may help determine MOA. A/anti-A activity possible in lower vertebrates. Consider performing an FSDT or LAGDA. Equivocal results may indicate chemical has multiple MOA.
I	+	Eq/0	Eq/0	A/anti-A activity of unknown potency. May act via AR, metabolic activation is required. 5-alpha reductase inhibitor of unknown potency. Unknown potential for adverse effects.	For the "0" scenario, perform AR transactivation assay or binding assay. For the "Eq" scenario, perform AR transactivation assay or binding assay with added metabolising system, or Level 4 or 5 assay if existing data indicate this is needed.	Check pattern of change across sex tissues for possible 5-alpha reductase inhibition. Check data on chemical analogues. Further mechanistic studies may help determine MOA. A/anti-A activity possible in lower vertebrates. Consider performing an FSDT or LAGDA. Equivocal results may indicate chemical has multiple MOA.

Scenarios	Result of OECD TG 441 (H assay)	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
J	–	+	+	No evidence for A/anti-A activity via AR or 5-alpha reductase inhibition *in vivo*. Route of exposure, metabolic differences or potency explain differences between H assay and existing *in vitro*/*in vivo* studies. Effects seen in existing studies are via non-AR/5-alpha reductase mechanism.	Perform AR transactivation assay or binding assay with added metabolising system.	If existing data are from an adequate Level 5 (or less sensitive) assay, there may be sufficient information to conclude concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). Consider route of exposure for H assay and possible implications of ADME characteristics of the chemical. Effects seen in existing studies may be in a more sensitive life stage. Check data on chemical analogues. Further mechanistic studies may help determine MOA.
K	–	+	–	No evidence for A/anti-A activity via AR or 5-alpha reductase inhibition *in vivo*. Metabolic differences or potency explain *in vitro*/*in vivo* differences.	Perform AR transactivation assay or binding assay with added metabolising system.	If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from a less sensitive assay, a higher level test may be required. Metabolic deactivation of chemical may occur *in vivo* so that potential *in vitro* activity is not realised. Consider possible routes of exposure, implications of metabolism.
L	–	+	Eq/0	No evidence for A/anti-A activity via AR or 5-alpha reductase inhibition *in vivo*. Metabolic differences or potency explain *in vitro*/*in vivo* difference. Unknown potential for adverse effects.	Perform AR transactivation assay or binding assay with added metabolising system OR Perform assay from Levels 4 or 5.	Metabolic deactivation of chemical may occur *in vivo* so that potential *in vitro* activity is not realised. Consider possible routes of exposure, implications of metabolism. Equivocal results may indicate chemical has multiple MOA.
M	–	–	+	No evidence for A/anti-A activity via AR or 5-alpha reductase inhibition in H assay or *in vitro*. Metabolic differences or route of exposure explain *in vitro*/*in vivo* existing differences. Effects seen in existing studies are via non-AR or non-endocrine mechanism.	Perform *in vitro* assays with added metabolising system.	Effects seen in existing studies may be in a more sensitive life stage. Check data on chemical analogues. Further mechanistic studies may help determine MOA.
N	–	–	–	No evidence for A/anti-A activity via AR or 5-alpha reductase inhibition *in vivo* or *in vitro*. No evidence of adverse effects.	Possibly no need for further testing. If there is uncertainty, may perform assay from Levels 4 or 5.	If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). Check data on chemical analogues.

Scenarios	Result of OECD TG 441 (H assay)	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
O	–	–	Eq/0	No evidence for A/anti-A activity via AR or 5-alpha reductase inhibition *in vivo* or *in vitro*. Unknown potential for adverse effects via other non-AR mechanisms.	Perform assay from Levels 4 or 5.	Consider route of exposure for H assay and possible implications for ADME characteristics of the chemical in follow-up assay.
P	–	Eq/0	+	No evidence for A/anti-A activity via AR or 5-alpha reductase inhibition *in vivo*. Unknown potential for adverse effects via other mechanisms.	For the "0" scenario, perform *in vitro* E,A,T,S assays, otherwise Eq result available.	Consider route of exposure for H assay and possible implications for differences from existing assay. Effects seen in existing studies may be in a more sensitive life stage. Check data on chemical analogues. Further mechanistic studies may help determine MOA. Equivocal results may indicate chemical has multiple MOA.
Q	–	Eq/0	–	No evidence for A/anti-A activity via AR or 5-alpha reductase inhibition *in vivo*. No evidence of adverse effects.	For the "0" scenario, perform *in vitro* E,A,T,S assays, otherwise Eq result available.	If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from a less sensitive assay, a higher level test may be required. Check data on chemical analogues. Further mechanistic studies may help determine MOA.
R	–	Eq/0	Eq/0	No evidence for A/anti-A activity via AR or 5-alpha reductase inhibition activity *in vivo*.	For the "0" scenario, perform *in vitro* E,A,T,S assays , otherwise Eq result available OR Perform Level 5 assay.	Consider route of exposure for H assay and possible implications for differences from existing assay. Check data on chemical analogues. Further mechanistic studies may help determine MOA. Equivocal results may indicate chemical has multiple MOA.

C.3.3. Repeated Dose 28-Day Oral Toxicity Study in Rodents (OECD TG 407)

Status: Assay validated by the OECD.

695. Modalities detected: (anti)estrogen, (anti)androgen, thyroid, steroidogenesis.

696. Endpoints: Mandatory: Weight of adrenals, testes, epididymides, prostate + seminal vesicles with coagulating glands. Histopathologic changes in testes, epididymides, prostate + seminal vesicles with coagulating glands, ovary, uterus/cervix, vagina, thyroid gland and adrenals.

697. Optional: Weight of uterus, ovaries, thyroid. Estrous cyclicity. Histopathologic changes in mammary glands and pituitary. Circulating levels of T3, T4, TSH.

Background to the assay

698. This assay determines the general toxicity of chemicals in rodents after 28 days of oral dosing (e.g. effects on liver, kidneys, heart, lungs); it also provides information on effects on the nervous, immune and reproductive systems. This is the primary purpose of this assay. The preferred species is the rat. Route of administration is oral, by gavage, via the diet or in drinking water. The original OECD TG 407 was adopted in 1981. A revised version was adopted in 1995, to obtain additional information in particular on neurotoxicity and immunotoxicity. It then underwent a validation study in the rat (OECD, 2006) where more parameters suitable for the detection of endocrine disruptors (EDs) were included and the test guideline (TG) was updated in October 2008 (most recent version). Following the validation study, many of the parameters were included in the updated guideline, as either mandatory or optional endpoints. It is important that the collection of endocrine endpoints does not interfere with the primary purpose (e.g. collection of blood for hormones should ideally be carried out at a comparable time of day in case of diurnal variations but blood collection for clinical chemistry should take precedence).

699. Experience with of serum hormone determinations in Levels 4 and 5 rodent assays has revealed that their detection/measurement in rodent studies can be challenging. A recent workshop was “Practicability of Hormonal Measurements” was organised by the BfR (Germany) and the finding from this workshop will be published (Kucheryavenko et al., 2018). The OECD Expert Group on Reproductive and Developmental Toxicity recommends that to demonstrate proficiency for thyroid hormones measurement, a laboratory should be able to show results from a separate study using a positive control substance. Laboratories may also submit their calibration curves, standard curves, as well as data on the levels of quantification and detection. This group is also establishing a historical control database with thyroid toxicant positive controls.

700. OECD TG 407 is considered to be an apical assay (i.e. it contains endpoints that may be changed by a number of different modes of action [MOA] and may not be specific to endocrine active substances [EASs]). The animals are young adults with intact hypothalamus-pituitary-gonadal/thyroid axes and therefore are a relevant model for human

health, although the sensitivity of the assay for EASs is less than that of the UT and H assays. The validation of the assay for endocrine endpoints showed that this assay is relatively insensitive and would only detect chemicals that were moderate and strong EDs for (anti)estrogenicity and (anti)androgenicity (e.g. ethinylestradiol and flutamide). However, it did detect EDs that were weak and strong modulators of thyroid hormone-related effects (e.g. propylthiouracil, T4 and methyl testosterone). It may also detect steroidogenesis inhibition, although only one (potent) chemical was used in the validation study (CGS 18320B) (OECD, 2006). Endocrine modalities other than estrogen/androgen/thyroid/steroidogenesis (E,A,T,S) may also be detected, although these have not been validated.

When/why the assay may be used

701. This assay is likely to be used as a preliminary study for longer term studies (e.g. 90-day studies or carcinogenicity studies, where the endocrine endpoints give additional information on the potential of the chemical to interact with the endocrine system). This assay is also necessary as a standard information requirement in certain chemical legislations (e.g. REACH for chemicals manufactured or imported in quantities of ten tonnes or more). It may also be used for chemicals where chronic exposure scenarios are not anticipated. Depending on the number of doses used, the assay may be used for hazard assessment (when one or two doses are used) or for hazard characterisation if a more detailed dose response curve is available. It should be noted that as this assay is not primarily designed to detect endocrine disruption, a higher degree of systemic toxicity is typically induced than is the case with the other Level 3 and 4 assays. The possibly confounding effect of systemic toxicity on endocrine endpoints therefore needs to be considered.

702. In order to provide information relevant for assessing whether or not a chemical may fulfil the WHO/IPCS (2002) definition of an ED, the study design has to be sufficiently robust to demonstrate the presence or absence of effects. In the dose selection, the investigator should also consider and ensure that data generated are adequate to fulfil the regulatory requirement across OECD countries as appropriate (e.g. hazard and risk assessment and labelling, ED assessment, etc.). The top dose or concentration should be sufficiently high to give clear systemic (i.e. non endocrine-specific) toxicity in order to ensure that a wide range of exposures (high to low) is tested. However, endocrine effects observed solely in the presence of clear systemic toxicity should be interpreted with caution and may be disregarded when sufficiently justified to be caused by secondary effects which are unlikely to be due to endocrine activity. The reason for this advice is a concern that some endocrine active substance (EAS) sensitive assays are being run at doses/concentrations of EASs that are too low to trigger direct impacts on the endocrine system. This guidance document is not the place to address this issue directly, but it should be considered when EAS-sensitive TGs are revised in the future. In addition, the number and spacing of dose/concentration levels should also be adequate to fulfil the objectives of the study (e.g. to demonstrate dose response relationships if this is required).

Introduction to the table of scenarios

703. Table C.3.3 gives guidance on a further step to take in the event of a positive (+) or negative (-) result and in the presence of positive (+), negative (-) or equivocal/absent (Eq/0) existing results. “Existing results” are subdivided into “mechanism” and “effects” data (third and fourth columns). The table is divided horizontally into a series of scenarios that represent all the combinations of these events.

704. The results of OECD TG 407 are given in the second column. As OECD TG 407 is not a screening test where a yes/no (qualitative) answer is obtained, criteria for positive results for the endocrine endpoints are not given in the TG. Results for the endpoints would be considered both individually and as a whole. It is not possible to provide guidance on all endpoints individually and therefore the endpoints have been pragmatically divided into "apical" and "indicators of hormonal activity". Both groups have similar biological importance, although the "indicators of hormonal activity" in the mammalian assays are serum hormones and are generally, but not always, more variable than "apical endpoints".

705. "Apical" endpoints are weights of testes, epididymides, prostate (+ seminal vesicles with coagulating glands), ovary, uterus, histopathologic changes in testes, epididymides, prostate, seminal vesicles, coagulating glands, ovary, uterus/cervix, vagina, thyroid and estrous cyclicity. "Indicators of hormonal activity" are hormones (T3, T4 and TSH).

706. Three possible outcomes for a positive result are therefore envisaged in Table C.3.3:

1. indicators of hormonal activity and apical endpoints positive
2. indicators of hormonal activity positive and apical endpoints negative
3. indicators of hormonal activity negative and apical endpoints positive.

707. A positive result for apical endpoints could be significant reductions in reproductive organ weights, accompanied by treatment-related histopathologic changes. A positive result for indicators of hormonal activity could be biologically significant changes in thyroid hormone profiles. The indicators of hormonal activity are optional endpoints for this TG and therefore they may not be measured. Alternatively, other endpoints not specified in the guideline (e.g. reproductive hormones), may be measured and if positive would contribute to the overall assessment of a positive result. The apical endpoints for the detection of effects on male and female reproductive organs tended to be less sensitive than the indicators of hormonal activity in the validation of OECD TG 407 and therefore changes are more likely to be indicative of an ED, although the results in entirety should be considered rather than single isolated changes. This was not true for the thyroid though, where changes in thyroid histopathology were always as sensitive, or more sensitive, than changes in thyroid hormone/TSH levels. The guidance on histopathologic changes in endocrine tests (OECD, 2009) may be helpful in interpretation. A positive result for indicators of hormonal activity alone should be considered with caution, although it is possible that these endpoints may have detected weak effects that were not detected by the apical endpoints in this study but may then be detected in longer term studies.

708. A negative result for OECD TG 407 is taken to be the absence of changes in both endocrine-relevant indicators of hormonal activity and apical endpoints. In the absence of other pertinent lines of evidence, negative results in this test alone cannot be taken as evidence that the chemical is not an ED. Further studies will be required as confirmation.

709. Equivocal results for the guideline are not considered in Table C.3.3, partly for brevity but also because equivocal results are by nature uncertain. A decision must eventually be reached about whether the endocrine endpoints tend to be positive or negative or whether the result must be put to one side and the test repeated (using the same or a different test guideline). Factors which may have interfered with the result (e.g. composition of the diet used, environmental influences), should be considered. Apparent equivocal results may arise because of the low sensitivity of the assay for (anti-)estrogens/androgens.

Existing data to be considered

710. Existing "mechanism" *in vitro* data are assumed to be available from estrogen receptor (ER; ER binding and ER STTA), androgen receptor (AR; AR binding and AR STTA) and steroidogenesis-based assays (Level 2). Assays may also be available for interference with thyroid modalities. In practice, it is possible that data from all of these assays may not be available, so judgement will need to be used to decide which assays to perform. Although the current *in vitro* test guidelines do not incorporate metabolic activation, published information on use of metabolic activation systems is available in Jacobs et al. (2008; 2013) and OECD (2008a). These methods, however, have not yet been validated.

711. Existing "effects" data refer to *in vivo* effects that may come from UT or H assays where a non-physiological animal model is used. In these cases, it should be remembered that these assays are specifically designed to be sensitive to EASs, compared to OECD TG 407. Other data such as repeat dose oral toxicity studies, reproduction/developmental toxicity screen tests may be available, although it is unlikely that OECD TG 407 will be performed if higher tier data are already available as OECD TG 407 offers no advantage over these assays. Data may also be available on effects in mammalian and non-mammalian vertebrate species, although caution should be used when extrapolating between taxa. A chemical causing endocrine effects in non-mammalian environmental species (fish, for example) may also have endocrine effects in mammals, but the physiological consequences of the effects are likely to be different.

712. When considering the results of the OECD TG 407 assay, all available data should be used in order to reach a conclusion and take a weight of evidence approach. This may include high throughput screening data, read-across data from structural analogues and quantitative structure activity relationship (QSAR). Several QSAR models for ER and AR binding/activation are now available (see Sections B.1.1.1 and B.1.1.2).

Scenarios: Positive and negative results combined with existing data

713. The scenarios (A to R) presented in Table C.3.3 represent all the possibilities of positive or negative results in combination with the presence or absence of existing data. The action taken will also depend on the regulatory environment, but the considerations given here are generally science based. Although the OECD TG 407 assay uses rodents, the well-conserved nature of the hormonal pathways across taxa indicate that results on endocrine endpoints in this assay may be relevant to other vertebrate species. Effects in laboratory mammal tests are also highly relevant for environmental mammalian species. Wherever possible, the recommended "next step which could be taken" avoids unnecessary animal testing. However, sometimes conducting an animal test will be indicated and then the relevance of species, strain and exposure route should always be considered. The sensitivity and physiological function of the hormone under investigation in the test species should also be considered. In general, lower level tests should be conducted before higher level tests in order to avoid unnecessary animal usage, unless it is apparent that a Level 5 test will be required anyway or will be needed to establish the evidence to conclude on ED properties. Information on some endocrine-related tumours may be detected more comprehensively in carcinogenicity studies (OECD TG 451/453) (Level 4); for example, detection of certain types of thyroid tumors in the absence of reproductive or developmental effects, as well as substances causing tumors in other endocrine-sensitive tissues. At Level 5, the Extended One-Generation Reproduction Toxicity Study (EOGRTS – OECD TG 443)

is the most sensitive reproduction assay for detecting endocrine disruption because it includes evaluation of a number of endocrine endpoints not included in the two-generation study (OECD TG 416) adopted in 2001. It is recognised, however, that some jurisdictions may require a two-generation study. Further considerations, specific to each scenario are given in the table.

714. Scenarios A to C represent positive results in the OECD TG 407 assay in the presence of positive *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. Each positive OECD TG 407 result scenario is divided into the three possible outcomes given above. A positive result in the *in vitro* assays in combination with a positive OECD TG 407 assay is moderate or strong evidence for E,A,T,S-mediated activity that may or may not be supported by the *in vivo* effects data. In the absence of robust upper-level data, the next step may be to conduct an upper-level test. In the presence of robust *in vivo* data, there may be sufficient evidence to conclude concern for endocrine disruption and therefore no need for further testing. Positive results in the OECD TG 407 assay may also indicate the potential for endocrine mediated effects in lower vertebrates. These could be followed up with partial life cycle tests such as the Fish Sexual Development Test (FSDT), the Larval Amphibian Growth and Development Assay (LAGDA) or the Medaka Extended One-Generation Reproduction Test (MEOGRT) if the evidence is strong enough. *In vivo* assays/tests with negative results should be interpreted with caution as they may either indicate that the tests used do not have sufficient power to detect weak effects or, alternatively, that the effects do not present a concern for endocrine disruption. The possibility of other (non-E,A,T,S) mechanisms should also not be overlooked (e.g. involving other receptors or endocrine axes).

715. Scenarios D to F represent positive results in the OECD TG 407 assay in the presence of negative *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. Each positive OECD TG 407 result scenario is divided into the three possible outcomes given above. Negative results in the *in vitro* assays should be viewed with caution in case a metabolite is responsible for the positive OECD TG 407 assay. Unless the metabolic profile of the test substance is known, one option may be to conduct these assays with an added metabolising system. If the metabolic profile is known, then a higher level *in vivo* test may be advisable. The choice of tests will depend on the available *in vivo* effects data. Positive results in the OECD TG 407 assay may also indicate the potential for endocrine mediated effects in lower vertebrates. As in scenarios A to C, *in vivo* assays/tests with negative results should be interpreted with caution as they may either indicate that the tests used do not have sufficient power to detect weak effects or, alternatively, that the effects do not present a concern for endocrine disruption.

716. Scenarios G to I represent positive results in the OECD TG 407 assay in the presence of various combinations of missing or equivocal data. Positive results in the OECD TG 407 assay may also indicate the potential for endocrine mediated effects in lower vertebrates. Each positive OECD TG 407 result scenario is divided into the three possible outcomes given above. The next step to take in these eventualities will depend on the nature of the other available data and the jurisdiction in which it is being used. In some cases, equivocal data may be viewed as positive whilst in others it may or may not contribute to the weight of evidence. The interpretation may also depend on the MOA in question and why the data are considered equivocal, e.g. a study that is equivocal for thyroid effects may still be of value in evaluating (anti)androgenic effects. In all three scenarios, the recommended first step is to obtain reliable mechanistic (*in vitro*) data rather than proceed further with *in vivo* testing. Equivocal and missing data are alternative scenarios and two possibilities for the next step are given in most cases, but the nature of equivocal data means

that decisions need to be taken on a case-by-case basis. In all cases, the role of metabolism, route of exposure and data from structural analogues should be considered before deciding on the next step.

717. Scenarios J to L represent negative results in the OECD TG 407 assay in the presence of positive *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. As a negative result for OECD TG 407 is taken to be negative findings for both indicators of hormonal activity and apical endpoints then (unlike the situation with positive outcomes), there is only one possible negative outcome. Negative outcomes in OECD TG 407 should be viewed with caution because of the power of the assay to detect (anti)estrogens and androgens. All three scenarios could also arise from a chemical that is positive in *in vitro* assays, but is metabolised to a non-active metabolite leading to negative results in the OECD TG 407 assay. This should be considered first when investigating the next step. Endocrine active potency may also explain differences between *in vitro* and *in vivo* results (e.g. a chemical with weak endocrine activity may give a positive result *in vitro* but may be negative *in vivo*). Positive *in vivo* effects data may involve other E,A,T,S, non-E,A,T,S mechanisms (e.g. involving other receptors or endocrine axes), more sensitive endpoints, greater statistical power or life stages that are more sensitive to the substance than the young adult animals in OECD TG 407.

718. Scenarios M to O represent negative results in the OECD TG 407 assay in the presence of negative *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. Negative results for all tests (Scenario N) may be sufficient to enable a conclusion of no concern for endocrine disruption. This will depend on the weight of evidence and may not be possible. Where there are positive *in vivo* effects data, there could still be an E,A,T,S-related mechanism, the effects may be related to length of exposure, route of exposure or exposure at different life stages. Other E,A,T,S or non-E,A,T,S mechanisms may also be involved.

719. Scenarios P to R represent negative results in the OECD TG 407 assay in the presence of various combinations of missing or equivocal data. As with the positive result scenarios above (see Paragraph 712) the next step to take in these eventualities will have to be decided on a case-by-case basis. However, the recommended first step is generally to obtain reliable mechanistic (*in vitro*) data rather than proceed further with *in vivo* testing. In all cases, the role of metabolism, route of exposure and data from structural analogues should be considered before deciding on the next step.

720. In all scenarios (A to R), the next step to take to strengthen weight of evidence will depend on the existing information. Table C.3.3 is meant to provide a succinct guide and may not cover all circumstances or possibilities. The scenarios may also suggest that chemicals have simple or single MOA, when in practice they may have multiple endocrine and non-endocrine MOA. In some cases, for example, two opposite modes of simultaneous action (e.g. estrogenic and anti-estrogenic) could, depending on dose, lead to a minimisation or abolition of effects, while in others two different MOA (e.g. estrogenic and anti-androgenic) could potentially reinforce effects. Endocrine pathways interact, mixed effects are common and there are many pathways that cannot be distinguished with currently available TGs. If multiple MOA are suspected, either from the existing results or based on QSAR/read-across/integrated approaches, this should be investigated further if needed for regulatory decision making.

References

Jacobs, M. et al. (2013), "*In vitro* metabolism and bioavailability tests for endocrine active substances: What is needed next for regulatory purposes?", *ALTEX – Alternatives to Animal Experimentation*, Vol. 30/3, pp. 331-351.

Jacobs, M.N. et al. (2008), "The use of metabolising systems for *in vitro* testing of endocrine disrupters", *Current Drug Metabolism*, Vol. 9/8, pp. 796-826.

Kucheryavenko, O. et al. (2018), "Report from the BfR Expert Hearing on Practicability of Hormonal Measurements", *Archives of Toxicology* (in prep.).

OECD (2009), "Guidance document for histologic evaluation of endocrine and reproductive tests in rodents", OECD Series on Testing and Assessment, No. 106, OECD, Paris, www.oecd.org/env/ehs/testing/43411534.pdf.

OECD (2008a), *Detailed Review Paper on the Use of Metabolising Systems for* In Vitro *Testing of Endocrine Disrupters*, OECD Series on Testing and Assessment, No. 97, OECD Publishing, Paris, https://doi.org/10.1787/9789264085497-en.

OECD (2008b), "Report of the OECD validation of the Rodent Hershberger Bioassay: Phase 2: Testing of androgen agonists, androgen antagonists and a 5 α-Reductase inhibitor in dose response studies by multiple laboratories", OECD Series on Testing and Assessment, No. 86, OECD, Paris, www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2008)3&doclanguage=en.

OECD (2008c), "Guidance document on mammalian reproductive toxicity testing and assessment", OECD Series on Testing and Assessment, No. 43, OECD, Paris, www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/JM/MONO(2008)16&doclanguage=en.

OECD (2006), "Detailed review paper on thyroid hormone disruption assays", OECD Series on Testing and Assessment, No. 57, OECD, Paris, www.oecd.org/env/ehs/testing/37235405.pdf.

WHO/IPCS (2002), "Global assessment of the state-of-the-science of endocrine disrupters", Damstra, T. et al. (eds.) WHO/PCS/EDC/02.2, World Health Organization, Geneva, www.who.int/ipcs/publications/new_issues/endocrine_disruptors/en.

Table C.3.3. **Repeated Dose 28-Day Oral Toxicity Study in Rodents (OECD TG 407): Guidance for scenarios of combinations of results with existing data**

This table represents possible conclusions to be drawn from assay data, and a next step which could be taken if further evidence is required about possible endocrine disrupting properties and/or effects. The guidance offered is not meant to be prescriptive, but provides science-based considerations. It encourages the use of all available data and expert judgement in a weight of evidence approach. Regional and national interpretation of results and "next steps" may vary.

The conclusions are grouped into a series of scenarios (A-R), each scenario representing a different combination of assay results, existing *in vitro* data and existing *in vivo* data. The symbol "+" indicates that the data in question represent a positive result, «-" indicates a negative result, and "Eq/0" indicates that the data are either equivocal or are not available.

Existing results * "Mechanism (*in vitro* mechanistic data)" assumes that mechanistic data are available from estrogen receptor (ER-), androgen receptor (AR-) and steroidogenesis-based assays (Level 2). Thyroid hormone receptor (TR) and other assays concerning mechanisms of thyroid disruption may be available, but they are not in common use. In practice, data from all assays may not be available and therefore this must be taken into account when deciding on the "next step". Quantitative structure activity relationship (QSAR) predictions of estrogen and androgen binding/activation may be made for some substances.

Existing results ** "Effects (*in vivo* effects of concern)" assumes effects have been observed in other *in vivo* screens/tests which give rise to concern that the test chemical may be an endocrine disrupter. These may be other repeated dose toxicity tests, Uterotrophic Bioassays and Hershberger Bioassays or read-across from chemical analogues.

*** *Note:* three possible outcomes for a positive result are given:

1. indicators of hormonal activity and apical endpoints positive

2. indicators of hormonal activity positive and apical endpoints negative

3. indicators of hormonal activity negative and apical endpoints positive.

"Apical endpoints" are weights of testes, epididymides, prostate (+ seminal vesicles with coagulating glands), ovary, uterus, histopathologic changes in testes, epididymides, prostate, seminal vesicles, coagulating glands, ovary, uterus/cervix, vagina, thyroid and estrous cyclicity.

"Indicators of hormonal activity" are hormones (T3, T4 and TSH).

Scenarios	Result of OECD TG 407 (rodent 28-day) assay	Existing results		Possible conclusions: 1) Indicators of hormonal activity and apical endpoints positive 2) Indicators of hormonal activity positive and apical endpoints negative 3) Indicators of hormonal activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
A	+ ***	+	+	1) Moderate or strong (anti)-E,A,T,S activity. Increased evidence of (anti)-E,A,T,S activity. 2) Possible evidence of (anti)-E,A,T,S activity, apical endpoints may be less sensitive or unaffected. Increased evidence of (anti)-E,A,T,S activity. 3) Moderate or strong (anti)-E,A,T,S activity, indicators of hormonal activity may be less sensitive or unaffected. Increased evidence of (anti)-E,A,T,S activity.	Perform assay from Level 5 (e.g. Extended One-Generation Reproduction Toxicity Study [EOGRTS,) or two-generation) assay.	If existing data are from a Level 5 assay, there may be sufficient information to conclude evidence of concern for endocrine disruption (the EOGRTS provides the most information; however, for endocrine disrupting chemicals [EDCs] with a carcinogenic potential, OECD TG 451-3 may be more sensitive). Effects on indicators of hormonal activity alone may be indicative of changes not detected by apical endpoints. Effects on apical endpoints alone may indicate E,A,T,S modalities or other mechanisms. Consider route of exposures for effects data and possible implications of absorption, distribution, metabolism and excretion (ADME) characteristics of the chemical. Endocrine activity possible in lower vertebrates. Consider performing a Fish Sexual Development Test (FSDT), a Larval Amphibian Growth and Development Assay (LAGDA) or a Medaka Extended One-Generation Reproduction Test (MEOGRT).
B	+	+	–	1) Moderate or strong (anti)-E,A,T,S activity. Increased evidence of (anti)-E,A,T,S activity. 2) Possible evidence of (anti)-E,A,T,S activity, apical endpoints may be less sensitive or unaffected. Increased evidence of (anti)-E,A,T,S activity. 3) Moderate or strong (anti)-E,A,T,S activity, indicators of hormonal activity may be less sensitive or unaffected. Increased evidence of (anti)-E,A,T,S activity.	Perform assay from Level 5 (e.g. EOGRTS or two-generation) assay.	If existing data are from an adequate Level 5 assay, question why there are differences. Consider route of exposures and possible implications of ADME characteristics of the chemical. If existing data are from a less sensitive assay, a higher level test may be required. Effects on indicators of hormonal activity alone may be indicative of changes not detected by apical endpoints. Effects on apical endpoints alone may indicate E,A,T,S modalities or other mechanisms. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT.
C	+	+	Eq/0	1) Moderate or strong (anti)-E,A,T,S activity. Increased evidence of (anti)-E,A,T,S activity. 2) Possible evidence of (anti)-E,A,T,S activity, apical endpoints may be less sensitive or unaffected. Increased evidence of (anti)-E,A,T,S activity. 3) Moderate or strong (anti)-E,A,T,S activity, indicators of hormonal activity may be less sensitive or unaffected. Increased evidence of (anti)-E,A,T,S activity.	Perform assay from Level 5 (e.g. EOGRTS or two-generation) assay.	Check data on chemical analogues. Effects on indicators of hormonal activity alone may be indicative of changes not detected by apical endpoints. Effects on apical endpoints alone may indicate E,A,T,S modalities or other mechanisms. Consider route of exposure for OECD TG 407 and follow-up assay. Possible implications of ADME characteristics of the chemical. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT. Equivocal results may indicate chemical has multiple modes of action (MOA).

Scenarios	Result of OECD TG 407 (rodent 28-day) assay	Existing results		Possible conclusions: 1) Indicators of hormonal activity and apical endpoints positive 2) Indicators of hormonal activity positive and apical endpoints negative 3) Indicators of hormonal activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
D	+	–	+	1) Moderate or strong (anti)-E,A,T,S activity. Acts via non-ER,AR,TR, S mechanism or requires metabolic activation for activity. 2) Possible evidence of (anti)-E,A,T,S activity, apical endpoints may be less sensitive or unaffected. Acts via non-ER, AR, TR, S mechanism or requires metabolic activation for activity. 3) Moderate or strong (anti)-E,A,T,S activity, indicators of hormonal activity may be less sensitive or unaffected. Acts via non-ER, AR, TR, S mechanism or requires metabolic activation for activity.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system.	If existing data are from an adequate Level 5 assay, there is sufficient information to conclude evidence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). Effects on indicators of hormonal activity alone may be indicative of changes not detected by apical endpoints. Effects on apical endpoints alone may indicate E,A,T,S modalities or other mechanisms. Consider route of exposures for effects data and possible implications of ADME characteristics of the chemical. Further mechanistic studies may help determine MOA. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT.
E	+	–	–	1) Moderate (anti)-E,A,T,S activity. Acts via non-ER, AR, TR, S mechanism or requires metabolic activation for activity. Route of exposure may account for the differences between OECD TG 407 and existing data. 2) Possible evidence of (anti)-E,A,T,S activity, apical endpoints may be less sensitive or unaffected. Acts via non-ER, AR, TR, S mechanism or requires metabolic activation for activity. Weak activity does not result in adverse effects. 3) Moderate (anti)-E,A,T,S activity, indicators of hormonal activity may be less sensitive or unaffected. Acts via non-ER, AR, TR, S mechanism or requires metabolic activation for activity. Weak activity does not result in adverse effects.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system. OR Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	If existing data are from an adequate Level 5 assay, question why there are differences. Consider route of exposures and possible implications of ADME characteristics of the chemical. If existing data are from a less sensitive assay, a higher level test may be required. Effects on indicators of hormonal activity alone may be indicative of changes not detected by apical endpoints. Effects on apical endpoints alone may indicate other mechanisms. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT.

Scenarios	Result of OECD TG 407 (rodent 28-day) assay	Existing results		Possible conclusions: 1) Indicators of hormonal activity and apical endpoints positive 2) Indicators of hormonal activity positive and apical endpoints negative 3) Indicators of hormonal activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
F	+	–	Eq/0	1) Moderate (anti)-E,A,T,S activity. Acts via non-ER, AR, TR, S mechanism or requires metabolic activation for activity. 2) Possible evidence of (anti)-E,A,T,S activity, apical endpoints may be less sensitive or unaffected. Acts via non-ER,AR,TR, S mechanism or requires metabolic activation for activity. 3) Moderate (anti)-E,A,T,S activity, indicators of hormonal activity may be less sensitive or unaffected. Acts via non-ER, AR, TR, S mechanism or requires metabolic activation for activity.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system OR Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	Effects on indicators of hormonal activity alone may be indicative of changes not detected by apical endpoints. Effects on apical endpoints alone may indicate other mechanisms. Check data on chemical analogues. Further mechanistic studies may help determine MOA. Level 5 studies will provide hazard data. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT. Equivocal results may indicate chemical has multiple MOA.
G	+	Eq/0	+	1) Moderate or strong (anti)-E,A,T,S activity. Increased evidence of (anti)-E,A,T,S activity. May act via ER, AR, TR, S mechanism (metabolic activation needed). 2) Possible evidence of (anti)-E,A,T,S activity, apical endpoints may be less sensitive or unaffected. Increased evidence of (anti)-E,A,T,S activity. May act via ER, AR, TR, S mechanism (metabolic activation needed). 3) Moderate or strong (anti)-E,A,T,S activity, indicators of hormonal activity may be less sensitive or unaffected. Increased evidence of (anti)-E,A,T,S activity. May act via ER, AR, TR, S mechanism (metabolic activation needed).	Perform *in vitro* ER, AR, TR, S assays (for the "0" scenario), otherwise Eq result available) OR Perform *in vitro* ER, AR, TR, S assays with added metabolising system.	If existing data are from a Level 5 assay, there may be sufficient information to conclude evidence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). Effects on indicators of hormonal activity alone may be indicative of changes not detected by apical endpoints. Effects on apical endpoints alone may indicate E,A,T,S modalities or other mechanisms. Check data on chemical analogues. Further mechanistic studies may help determine MOA. Consider route of exposures for effects data and possible implications of ADME characteristics of the chemical. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT. Equivocal results may indicate chemical has multiple MOA.

Scenarios	Result of OECD TG 407 (rodent 28-day) assay	Existing results		Possible conclusions: 1) Indicators of hormonal activity and apical endpoints positive 2) Indicators of hormonal activity positive and apical endpoints negative 3) Indicators of hormonal activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
H	+	Eq/0	–	1) Moderate (anti)-E,A,T,S activity. Acts via unknown mechanismor may require metabolic activation for activity. Route of exposure may account for the differences between OECD TG 407 and existing data. 2) Possible evidence of (anti)-E,A,T,S activity, apical endpoints may be less sensitive or unaffected. Acts via unknown mechanism. Weak activity does not result in adverse effects. 3) Moderate (anti)-E,A,T,S activity, indicators of hormonal activity may be less sensitive or unaffected. Acts via unknown mechanism. Weak activity does not result in adverse effects.	For the "0" scenario, perform *in vitro* ER, AR, TR, S assays, maybe with added metabolising system (otherwise Eq result available).	If existing data are from an adequate Level 5 assay, question why there are differences. Consider route of exposures and possible implications of ADME characteristics of the chemical. If existing data are from a less sensitive assay, a higher level test may be required. Effects on indicators of hormonal activity alone may be indicative of subtle changes not detected by apical endpoints. Effects on apical endpoints alone may indicate E,A,T,S modalities or other mechanisms. Check data on chemical analogues. Further mechanistic studies may help determine MOA. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT. Equivocal results may indicate chemical has multiple MOA.
I	+	Eq/0	Eq/0	1) Moderate or strong (anti)-E,A,T,S activity. Acts via unknown mechanism. Unknown potential for adverse effects. 2) Possible evidence of (anti)-E,A,T,S activity, apical endpoints may be less sensitive or unaffected. Acts via unknown mechanism. Unknown potential for adverse effects. 3) Moderate or strong (anti)-E,A,T,S activity, indicators of hormonal activity may be less sensitive or unaffected. Acts via unknown mechanism. Unknown potential for adverse effects. There may be a need for metabolic activation.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system OR Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	Effects on indicators of hormonal activity alone may be indicative of changes not detected by apical endpoints. Effects on apical endpoints alone may indicate E,A,T,S modalities or other mechanisms. Check data on chemical analogues. Further mechanistic studies may help determine MOA. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT. Equivocal results may indicate chemical has multiple MOA.
J	–	+	+	No evidence for moderate or strong (anti)-E,A,T,S activity in OECD TG 407. Weak (anti)-E,A,S activity not detected by this assay. Metabolism or potency explains the difference from existing *in vitro* and *in vivo* data.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system OR Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	If existing data are from an adequate Level 5 assay, question why there are differences. Effects seen in existing studies may be in a more sensitive life stage. Consider route of exposures and possible implications of ADME characteristics of the chemical. Further mechanistic studies may help determine MOA.

Scenarios	Result of OECD TG 407 (rodent 28-day) assay	Existing results		Possible conclusions: 1) Indicators of hormonal activity and apical endpoints positive 2) Indicators of hormonal activity positive and apical endpoints negative 3) Indicators of hormonal activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
K	–	+	–	No evidence for moderate or strong (anti)-E,A,T,S activity in OECD TG 407. Weak (anti)-E,A,S activity not detected by this assay. Metabolism or potency explains *in vitro*/*in vivo* differences.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system OR Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from a less sensitive assay, a higher level test may be required. Further mechanistic studies with metabolism may help determine MOA.
L	–	+	Eq/0	No evidence for moderate or strong (anti)-E,A,T,S activity in OECD TG 407. Weak (anti)-E,A,S activity not detected by this assay. Metabolism or potency explains *in vitro*/*in vivo* differences. Unknown potential for adverse effects.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system OR Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	Metabolic deactivation of chemical may occur *in vivo* so that possible *in vitro* activity is not realised. Consider possible routes of exposure implications of metabolism. Equivocal results may indicate chemical has multiple MOA.
M	–	–	+	No evidence for moderate or strong (anti)-E,A,T,S activity in OECD TG 407. Weak (anti)-E,A,S activity not detected by this assay.	Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	If existing data are from an adequate Level 5 assay, question why there are differences. Effects seen in existing studies may be in a more sensitive life stage. Consider route of exposures and possible implications of ADME characteristics of the chemical.
N	–	–	–	No evidence for moderate or strong (anti)-E,A,T,S activity in OECD TG 407. Weak (anti)-E,A,S activity not detected by this assay. No evidence for (anti)-E,A,T,S activity *in vitro*. No evidence of adverse effects.	Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive).
O	–	–	Eq/0	No evidence for moderate or strong (anti)-E,A,T,S activity in OECD TG 407. Weak (anti)-E,A,S activity not detected by this assay. No evidence for (anti)-E,A,T,S activity *in vitro*. Unknown potential for adverse effects.	Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	Consider route of exposures and possible implications for ADME characteristics of the chemical in follow-up assay.
P	–	Eq/0	+	No evidence for moderate or strong (anti)-E,A,T,S activity in OECD TG 407. Weak (anti)-E,A,S activity not detected by this assay. Potential for adverse effects via unknown mechanism.	Perform *in vitro* ER, AR,TR, S assays with added metabolising system.	Consider route of exposure for OECD TG 407 assay and possible implications for differences from existing assay. Effects seen in existing studies may be in a more sensitive life stage. Equivocal results may indicate chemical has multiple MOA.

Scenarios	Result of OECD TG 407 (rodent 28-day) assay	Existing results		Possible conclusions: 1) Indicators of hormonal activity and apical endpoints positive 2) Indicators of hormonal activity positive and apical endpoints negative 3) Indicators of hormonal activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
Q	–	Eq/0	–	No evidence for moderate or strong (anti)-E,A,T,S activity in OECD TG 407. Weak (anti)-E,A,S activity not detected by this assay. No evidence of adverse effects.	Perform *in vitro* ER, AR,TR, S assays with added metabolising system	If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). Further mechanistic studies may strengthen weight of evidence.
R	–	Eq/0	Eq/0	No evidence for moderate or strong (anti)-E,A,T,S activity in OECD TG 407. Weak (anti)-E,A,S activity not detected by this assay.	Perform *in vitro* ER, AR, TR, S assays, with added metabolising system, otherwise Eq result available.	Further mechanistic studies may strengthen weight of evidence. Check data on chemical analogues. Equivocal results may indicate chemical has multiple MOA.

C.3.4. Repeated Dose 90-Day Oral Toxicity Study in Rodents (OECD TG 408)

Status: Assay validated by the OECD.

721. Modalities detected: (anti)estrogen, (anti)androgen, thyroid, steroidogenesis.

722. Endpoints: Weight of thyroid gland, adrenals, testes, epididymides, uterus, ovaries, prostate + seminal vesicles with coagulating glands. Histopathologic changes in pituitary, thyroid gland, gonads, uterus, accessory sex organs, male and female mammary gland, testes and adrenals. Serum total T4, T3, and TSH..

723. Optional:. Estrous cyclicity. Circulating levels of, testosterone, oestradiol, follicle stimulating hormone (FSH), luteinising hormone (LH). Enumeration of cauda epididymis sperm reserves. Sperm morphology, sperm motility.

Background to the assay

724. This assay determines the general toxicity of chemicals in rodents after 90 days of oral dosing (by gavage, via the diet or in drinking water). The rat is the preferred species. It provides information on major toxic effects and target organ toxicity likely to arise from the post-weaning period until well into adulthood. OECD TG 408 was adopted in September 1998 and was updated in 2017 to add endocrine disrupter relevant endpoints intended to improve the detection of endocrine activity of test chemicals and mirrors updates to OECD TG 407 (Repeated Dose 28-Day Oral Toxicity Study in Rodents). In the updated version, an emphasis was placed on including additional thyroid parameters that could inform, alone or in combination with other information, on the potential of test chemicals to perturb the thyroid pathway. The update mirrored that of OECD TG 407 and therefore a comparison can be made with validation of the OECD TG 407 (28-Day Oral Toxicity Study) for endocrine endpoints where substances that were moderate and strong endocrine disruptors (EDs) for (anti)estrogenicity and (anti)androgenicity (e.g. ethinylestradiol and flutamide) and weak and strong modulators of thyroid hormone-related effects (e.g. propylthiouracil, T4 and methyl testosterone) were detected (OECD, 2006). Steroidogenesis inhibition was also detected, although only one (potent) chemical was used in the validation study (CGS 18320B). OECD TG 408 is likely to be more sensitive than OECD TG 407 because of the extended dosing period and the larger number of animals per group (ten male and ten female per group compared with five in OECD TG 407).

725. Experience with of serum hormone determinations in Levels 4 and 5 rodent assays has revealed that their detection/measurement in rodent studies can be challenging. A recent workshop on “Practicability of Hormonal Measurements” was organised by the BfR (Germany) and the finding from this workshop will be published (Kucheryavenko et al., 2018). The OECD Expert Group on Reproductive and Developmental Toxicity recommends that to demonstrate proficiency for thyroid hormones measurement, a laboratory should be able to show results from a separate study using a positive control substance. Laboratories may also submit their calibration curves, standard curves, as well as data on the levels of

quantification and detection. This group is also establishing a historical control database with thyroid toxicant positive controls.

When/why the assay may be used

726. This assay is likely to be used as part of a pesticide submission package and forms part of the standard information requirements in certain chemical legislations (e.g. REACH for chemicals which are manufactured or imported in quantities of 100 tonnes or more). At least three dose levels are included so that an estimate of no-adverse-effect-level can be determined and the assay used for hazard identification/characterisation. It should be noted that as this assay is not primarily designed to detect endocrine disruption, a higher degree of systemic toxicity is typically induced than is the case with the other Level 3 and 4 assays. The possibly confounding effect of systemic toxicity on endocrine endpoints therefore needs to be considered.

727. In order to provide information relevant for assessing whether or not a chemical may fulfil the WHO/IPCS (2002) definition of an endocrine disruptor (ED), the study design has to be sufficiently robust to demonstrate the presence or absence of effects. In the dose selection, the investigator should also consider and ensure that data generated are adequate to fulfil the regulatory requirement across OECD countries as appropriate (e.g. hazard and risk assessment and labelling, ED assessment, etc.). The top dose or concentration should be sufficiently high to give clear systemic (i.e. non endocrine-specific) toxicity in order to ensure that a wide range of exposures (high to low) is tested. However, endocrine effects observed solely in the presence of clear systemic toxicity should be interpreted with caution and may be disregarded when sufficiently justified to be caused by secondary effects which are unlikely to be due to endocrine activity. The reason for this advice is a concern that some endocrine active substance (EAS) sensitive assays are being run at doses/concentrations of EASs that are too low to trigger direct impacts on the endocrine system. This guidance document is not the place to address this issue directly, but it should be considered when EAS-sensitive test guidelines (TGs) are revised in the future. In addition, the number and spacing of dose/concentration levels should also be adequate to fulfil the objectives of the study (e.g. to demonstrate dose response relationships if this is required).

Introduction to the table of scenarios

728. Table C.3.4 gives guidance on a further step to take in the event of a positive (+) or negative (-) result and in the presence of positive (+), negative (-) or equivocal/absent (Eq/0) existing results. “Existing results” are subdivided into “mechanism” and “effects” data (third and fourth columns). The table is divided horizontally into a series of scenarios that represent all the combinations of these events.

729. The results of OECD TG 408 are given in the second column. As OECD TG 408 is not a screening test where a yes/no (qualitative) answer is obtained for the test as a whole, positive results would generally be assessed for individual endpoints. For the purposes of this guidance, however, a positive result is defined as a biologically significant change in any of the endocrine endpoints listed above (e.g. statistically significant reductions in reproductive organ weights). Changes in related endpoints will increase their biological significance (e.g. changes in the weights of testes and epididymides accompanied by histopathological changes). The guidance on histopathologic changes in endocrine tests

(OECD, 2009) may be helpful in interpretation. A negative result for OECD TG 408 is taken to be the absence of biologically significant changes in all endocrine endpoints.

730. In the absence of other pertinent lines of evidence, negative results in this test alone cannot be taken as evidence that the substance is not an ED. Further studies will be required as confirmation.

731. Equivocal results for the guideline are not considered in Table C.3.4, partly for brevity but also because equivocal results are by nature uncertain. A decision must eventually be reached about whether the endocrine endpoints tend to be positive or negative or whether the result must be put to one side and the test repeated (using the same or a different test guideline). Factors which may have interfered with the result (e.g. composition of the diet used, environmental influences) should be considered.

Existing data to be considered

732. Existing "mechanism" *in vitro* data are assumed to be available from estrogen receptor (ER-), androgen receptor (AR-) and steroidogenesis-based assays (Level 2). Assays may also be available for interference with thyroid modalities. In practice, it is possible that data from all of these assays may not be available, so judgement will need to be used to decide which assays to perform. Although the current *in vitro* test guidelines do not incorporate metabolic activation, published information on use of metabolic activation systems is available in Jacobs et al. (2008; 2013) and OECD (2008a). These methods, however, have not yet been validated.

733. Existing "effects" data refer to *in vivo* effects that may come from Level 3 or 4 tests in the Conceptual Framework (e.g. UT or H assays). In these cases, it should be remembered that these assays are specifically designed to be sensitive to EASs. It is unlikely that OECD TG 408 will be performed if higher tier data are already available as OECD TG 408 offers no advantage over these assays. As mentioned above, the results of the study may be interpreted as part of a battery or group of tests carried out for regulatory purposes. Data may also be available on effects in mammalian and non-mammalian wildlife species, although caution should be used when extrapolating between taxa. A chemical causing endocrine effects in non-mammalian environmental species (fish, for example) may also have endocrine effects in mammals, but the physiological consequences of the effects are likely to be different.

734. When considering the results of the OECD TG 408 assay, all available data should be used in order to reach a conclusion and a weight of evidence approach taken. This may include high throughput screening data, read-across data from structural analogues and quantitative structure activity relationship (QSAR). Several QSAR models for ER and AR binding/activation are now available (see Sections B.1.1.1 and B.1.1.2).

Scenarios: Positive and negative results combined with existing data

735. A series of scenarios (A to R) are presented in Table C.3.4 and represent all the possibilities of positive or negative results in combination with the presence or absence of existing data. The action taken will also depend on the regulatory environment, but the considerations given here are generally science based. Although the OECD TG 408 assay uses rodents, the well-conserved nature of the hormonal pathways across taxa indicate that results on endocrine endpoints in this assay may be relevant to other vertebrate species. Effects in laboratory mammal tests are also highly relevant for environmental mammalian species. Wherever possible, the recommended "next step which could be taken" avoids

unnecessary animal testing. However, sometimes conducting an animal test will be indicated and then the relevance of species, strain and exposure route should always be considered. The sensitivity and physiological function of the hormone under investigation in the test species should also be considered. In general, lower level tests should be conducted before higher level tests in order to avoid unnecessary animal usage, unless it is apparent that a Level 5 test will be required anyway or will be needed to establish the evidence to conclude on ED properties. Information on some endocrine-related tumours may be detected more comprehensively in carcinogenicity studies (OECD TG 451/453) (Level 4); for example, detection of certain types of thyroid tumors in the absence of reproductive or developmental effects, as well as substances causing tumors in other endocrine-sensitive tissues. At Level 5, the Extended One-Generation Reproduction Toxicity Study (EOGRTS – OECD TG 443) is the most sensitive reproduction assay for detecting endocrine disruption because it includes evaluation of a number of endocrine endpoints not included in the two-generation study (OECD TG 416) adopted in 2001. It is recognised, however, that some jurisdictions may require a two-generation study. Further considerations specific to each scenario are given in the table.

736. Scenarios A to C represent positive results in the OECD TG 408 assay in the presence of positive *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. A positive result in the *in vitro* assays in combination with a positive OECD TG 408 assay is moderate or strong evidence for estrogen/androgen/thyroid/steroidogenesis (E,A,T,S-) mediated activity that may or may not be supported by the *in vivo* effects data. In the absence of robust upper-level data, the next step may be to conduct an upper-level test. In the presence of robust *in vivo* data, there may be sufficient evidence to conclude concern for endocrine disruption and therefore no need for further testing. Positive results in the OECD TG 408 assay may also indicate the potential for endocrine mediated effects in lower vertebrates. These could be followed up with partial life cycle tests such as the Fish Sexual Development Test (FSDT), the Larval Amphibian Growth and Development Assay (LAGDA) or the Medaka Extended One-Generation Reproduction Test (MEOGRT) if the evidence were strong enough. *In vivo* assays/tests with negative results should be interpreted with caution as they may either indicate that the tests used do not have sufficient power to detect weak effects or alternatively that the effects do not present a concern for endocrine disruption. The possibility of other (non-E,A,T,S) mechanisms should also not be overlooked (e.g. involving other receptors or endocrine axes).

737. Scenarios D to F represent positive results in the OECD TG 408 assay in the presence of negative *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. Negative results in the *in vitro* assays should be viewed with caution in case a metabolite is responsible for the positive OECD TG 408 assay. Unless the metabolic profile of the test substance is known, one option may be to conduct these assays with an added metabolising system. If the metabolic profile is known, then a higher level *in vivo* test may be advisable. The choice of tests will depend on the available *in vivo* effects data. Positive results in the OECD TG 408 assay may also indicate the potential for endocrine mediated effects in lower vertebrates. As in Scenarios A to C, *in vivo* assays/tests with negative results should be interpreted with caution, as they may either indicate that the tests used do not have sufficient power to detect weak effects or, alternatively, that the effects do not present a concern for endocrine disruption.

738. Scenarios G to I represent positive results in the OECD TG 408 assay in the presence of various combinations of missing or equivocal data. Positive results in the OECD TG 408 assay may also indicate the potential for endocrine mediated effects in lower vertebrates. The next step to take in these eventualities will depend on the nature of the

other available data and the jurisdiction in which it is being used. In some cases, equivocal data may be viewed as positive whilst in others it may or may not contribute to the weight of evidence. The interpretation may also depend on the mode of action (MOA) in question and why the data are considered equivocal, e.g. a study that is equivocal for thyroid effects may still be of value in evaluating (anti)androgenic effects. In all three scenarios, the recommended first step is to obtain reliable mechanistic (*in vitro*) data rather than proceed further with *in vivo* testing. Equivocal and missing data are alternative scenarios and two possibilities for the next step are given in most cases, but the nature of equivocal data means that decisions need to be taken on a case-by-case basis. In all cases, the role of metabolism, route of exposure and data from structural analogues should be considered before deciding on the next step.

739. Scenarios J to L represent negative results in the OECD TG 408 assay in the presence of positive *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. Negative outcomes in OECD TG 408 should be viewed with caution because of the power of the assay to detect (anti)estrogens and androgens may be limited All three scenarios could also arise from a chemical that is positive in *in vitro* assays, but is metabolised to a non-active metabolite leading to negative results in the OECD TG 408 assay. This should be considered first when investigating the next step. Endocrine active potency may also explain differences between *in vitro* and *in vivo* results (e.g. a chemical with weak endocrine activity may give a positive result *in vitro* but may be negative *in vivo*). Positive *in vivo* effects data may involve other E,A,T,S, non-E,A,T,S mechanisms (e.g. involving other receptors or endocrine axes), more sensitive endpoints, greater statistical power or life stages that are more sensitive to the substance than the young adult exposed animals in OECD TG 408.

740. Scenarios M to O represent negative results in the OECD TG 408 assay in the presence of negative *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. Negative results for all tests (Scenario N) may be sufficient to enable a conclusion of no concern for endocrine disruption. This will depend on the weight of evidence and may not be possible. Where there are positive *in vivo* effects data, there could still be an E,A,T,S-related mechanism, the effects may be related to length of exposure, route of exposure or exposure at different life stages. Other E,A,T,S or non-E,A,T,S mechanisms may also be involved.

741. Scenarios P to R represent negative results in the OECD TG 408 assay in the presence of various combinations of missing or equivocal data. As with the positive result scenarios above, the next step to take in these eventualities will have to be decided on a case-by-case basis. However, the recommended first step is generally to obtain reliable mechanistic (*in vitro*) data rather than proceed further with *in vivo* testing. In all cases, the role of metabolism, route of exposure and data from structural analogues should be considered before deciding on the next step.

742. In all scenarios (A to R), the next step to take to strengthen weight of evidence will depend on the existing information. The table is meant to provide a succinct guide and may not cover all circumstances or possibilities. The scenarios may also suggest that chemicals have simple or single MOA, when in practice they may have multiple endocrine and non-endocrine MOA. In some cases, for example, two opposite modes of simultaneous action (e.g. estrogenic and anti-estrogenic) could, depending on dose, lead to a minimisation or abolition of effects, while in others two different MOA (e.g. estrogenic and anti-androgenic) could potentially reinforce effects. Endocrine pathways interact, mixed effects are common and there are many pathways that cannot be distinguished with currently available TGs. If multiple MOA are suspected, either from the existing results or based on QSAR/read-

across/integrated approaches, this should be investigated further if needed for regulatory decision making.

References

Jacobs, M. et al. (2013), "*In vitro* metabolism and bioavailability tests for endocrine active substances: What is needed next for regulatory purposes?", *ALTEX – Alternatives to Animal Experimentation*, Vol. 30/3, pp. 331-351.

Jacobs, M.N. et al. (2008), "The use of metabolising systems for *in vitro* testing of endocrine disrupters", *Current Drug Metabolism*, Vol. 9/8, pp. 796-826.

Kucheryavenko, O. et al. (2018), "Report from the BfR Expert Hearing on Practicability of Hormonal Measurements", *Archives of Toxicology* (in prep.).

OECD (2009), "Guidance document for histologic evaluation of endocrine and reproductive tests in rodents", OECD Series on Testing and Assessment, No. 106, OECD, Paris, www.oecd.org/env/ehs/testing/43411534.pdf.

OECD (2008a), *Detailed Review Paper on the Use of Metabolising Systems for* In Vitro *Testing of Endocrine Disrupters*, OECD Series on Testing and Assessment, No. 97, OECD Publishing, Paris, https://doi.org/10.1787/9789264085497-en.

OECD (2008b), "Report of the OECD validation of the Rodent Hershberger Bioassay: Phase 2: Testing of androgen agonists, androgen antagonists and a 5 α-Reductase inhibitor in dose response studies by multiple laboratories", OECD Series on Testing and Assessment, No. 86, OECD, Paris, www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2008)3&doclanguage=en.

OECD (2006), "Detailed review paper on thyroid hormone disruption assays", OECD Series on Testing and Assessment, No. 57, OECD, Paris, www.oecd.org/env/ehs/testing/37235405.pdf.

WHO/IPCS (2002), "Global assessment of the state-of-the-science of endocrine disrupters", Damstra, T. et al. (eds.) WHO/PCS/EDC/02.2, World Health Organization, Geneva, www.who.int/ipcs/publications/new_issues/endocrine_disruptors/en.

Table C.3.4. **Repeated Dose 90-Day Oral Toxicity Study in Rodents (OECD TG 408): Guidance for scenarios of combinations of results with existing data**

This table represents possible conclusions to be drawn from assay data, and a next step which could be taken if further evidence is required about possible endocrine disrupting properties and/or effects. The guidance offered is not meant to be prescriptive, but provides science-based considerations. It encourages the use of all available data and expert judgement in a weight of evidence approach. Regional and national interpretation of results and "next steps" may vary.

The conclusions are grouped into a series of scenarios (A-R), each scenario representing a different combination of assay results, existing *in vitro* data and existing *in vivo* data. The symbol "+" indicates that the data in question represent a positive result, «-" indicates a negative result, and "Eq/0" indicates that the data are either equivocal or are not available.

Existing results: * "Mechanism (*in vitro* mechanistic data)" assumes that mechanistic data are available from estrogen receptor (ER-), androgen receptor (AR-) and steroidogenesis-based assays (Level 2). Thyroid hormone receptor (TR) and other assays concerning mechanisms of thyroid disruption may be available, but they are not in common use. In practice, data from all assays may not be available and therefore this must be taken into account when deciding on the "next step". Quantitative structure activity relationship (QSAR) predictions of estrogen and androgen binding/activation may be made for some substances.

Existing results: ** "Effects (*in vivo* effects of concern)" assumes various information, such as data from repeat dose oral toxicity studies, reproduction/developmental toxicity screen tests, read-across from analogues, will be available.

*** *Note:* a positive result is defined as a biologically significant change in any of the endocrine endpoints.

Scenarios	Result of OECD TG 408 (rodent 90-day) assay	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
A	+ ***	+	+	(Anti)-E,A,T,S activity. Increased evidence of (anti)-E,A,T,S activity.	Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	If existing data are from a Level 5 assay, there may be sufficient information to conclude evidence of concern for endocrine disruption (the EOGRTS provides the most information; however, for endocrine disrupting chemincals [EDCs] with a carcinogenic potential, OECD TG 451-3 may be more sensitive). Consider route of exposures for effects data and possible implications of absorption, distribution, metabolism and excretion (ADME) characteristics of the chemical. Endocrine activity possible in lower vertebrates. Consider performing a Fish Sexual Development Test (FSDT), a Larval Amphibian Growth and Development Assay (LAGDA) or a Medaka Extended One-Generation Reproduction Test (MEOGRT).
B	+	+	–	(Anti)-E,A,T,S activity. Increased evidence of (anti)-E,A,T,S activity.	Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	If existing data are from an adequate Level 5 assay, question why there are differences. If existing data are from a less sensitive assay, a higher level test may be required. Consider route of exposures and possible implications of ADME characteristics of the chemical. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT.
C	+	+	Eq/0	(Anti)-E,A,T,S activity. Increased evidence of (anti)-E,A,T,S activity.	Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	Check data on chemical analogues. Consider route of exposure for OECD TG 408 and follow-up assay. Possible implications of ADME characteristics of the chemical. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT. Equivocal results may indicate chemical has multiple modes of action (MOA).
D	+	–	+	(Anti)-E,A,T,S activity. Acts via non-estrogen receptor (ER),androgen receptor (AR), thyroid hormone receptor (TR), steroidogenesis (S) mechanism or requires metabolic activation for activity.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system.	If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude evidence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). Consider route of exposures for effects data and possible implications of ADME characteristics of the chemical. Further mechanistic studies may help determine MOA. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT.
E	+	–	–	(Anti)-E,A,T,S activity. Acts via non-ER, AR, TR, S mechanism or requires metabolic activation for activity. Route of exposure may account for the differences between OECD TG 408 and existing data.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system OR Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	If existing data are from an adequate Level 5 assay, question why there are differences. If existing data are from a less sensitive assay, a higher level test may be required. Consider route of exposures and possible implications of ADME characteristics of the chemical. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT.

Scenarios	Result of OECD TG 408 (rodent 90-day) assay	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
F	+	–	Eq/0	(Anti)-E,A,T,S activity. Acts via non-ER, AR, TR, S mechanism or requires metabolic activation for activity.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system OR Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	Check data on chemical analogues. Further mechanistic studies may help determine MOA. Level 5 studies will provide hazard data. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT. Equivocal results may indicate chemical has multiple MOA.
G	+	Eq/0	+	(Anti)-E,A,T,S activity. May act via ER, AR, TR, S mechanism (metabolic activation needed).	Perform *in vitro* ER, AR, TR, S assays for the "0" scenario, otherwise Eq result available OR Perform *in vitro* ER, AR, TR, S assays with added metabolising system.	If existing data are from a Level 5 assay, there may be sufficient information to conclude evidence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). Check data on chemical analogues. Further mechanistic studies may help determine MOA. Consider route of exposures for effects data and possible implications of ADME characteristics of the chemical. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT. Equivocal results may indicate chemical has multiple MOA.
H	+	Eq/0	–	(Anti)-E,A,T,S activity. Acts via unknown mechanism or may require metabolic activation for activity. Route of exposure may account for the differences between OECD TG 408 and existing data.	For the "0" scenario, perform *in vitro* ER, AR, TR, S assays with added metabolising system (otherwise Eq result available).	If existing data are from an adequate Level 5 assay, question why there are differences. Consider route of exposures and possible implications of ADME characteristics of the chemical. If existing data are from a less sensitive assay, a higher level test may be required. Check data on chemical analogues. Further mechanistic studies may help determine MOA. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT. Equivocal results may indicate chemical has multiple MOA.
I	+	Eq/0	Eq/0	(Anti)-E,A,T,S activity. Acts via unknown mechanism. Unknown potential for adverse effects.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system OR Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	Check data on chemical analogues. Further mechanistic studies may help determine MOA. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT. Equivocal results may indicate chemical has multiple MOA.

Scenarios	Result of OECD TG 408 (rodent 90-day) assay	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
J	–	+	+	No evidence for (anti)-E,A,T,S activity in OECD TG 408. Weak (anti)-E,A,S activity may not be detected by this assay. Metabolism or potency explains the difference from existing *in vitro* and *in vivo* data.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system OR Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	If existing data are from an adequate Level 5 assay, question why there are differences. Effects seen in existing studies may be in a more sensitive life stage. Consider route of exposures and possible implications of ADME characteristics of the chemical. Further mechanistic studies may help determine MOA.
K	–	+	–	No evidence for (anti)-E,A,T,S activity in OECD TG 408. Weak (anti)-E,A,S activity may not be detected by this assay. Metabolism or potency explains *in vitro*/*in vivo* differences.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system OR Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from a less sensitive assay, a higher level test may be required. Further mechanistic studies with metabolism may help determine MOA.
L	–	+	Eq/0	No evidence for (anti)-E,A,T,S activity in OECD TG 408. Weak (anti)-E,A,S activity may not be detected by this assay. Metabolism or potency explains *in vitro*/*in vivo* differences. Unknown potential for adverse effects.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system OR Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	Metabolic deactivation of chemical may occur *in vivo* so that possible *in vitro* activity is not realised. Consider possible routes of exposure implications of metabolism. Equivocal results may indicate chemical has multiple MOA.
M	–	–	+	No evidence for (anti)-E,A,T,S activity in OECD TG 408. Weak (anti)-E,A,S activity may not be detected by this assay. Effects seen in existing studies are via non-E,A,T,S mechanism.	Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	If existing data are from an adequate Level 5 assay, question why there are differences. Effects seen in existing studies may be in a more sensitive life stage. Consider route of exposures and possible implications of ADME characteristics of the chemical.
N	–	–	–	No evidence for (anti)-E,A,T,S activity in OECD TG 408. Weak (anti)-E,A,S activity may not be detected by this assay. No evidence for (anti)-E,A,T,S activity *in vitro*. No evidence of adverse effects.	Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive).
O	–	–	Eq/0	No evidence for (anti)-E,A,T,S activity in OECD TG 408. Weak (anti)-E,A,S activity may not be detected by this assay. No evidence for (anti)-E,A,TS activity *in vitro*. Unknown potential for adverse effects.	Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	Consider route of exposures and possible implications for ADME characteristics of the chemical in follow-up assay.

Scenarios	Result of OECD TG 408 (rodent 90-day) assay	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
P	–	Eq/0	+	No evidence for (anti)-E,A,T,S activity in OECD TG 408. Weak (anti)-E,A,S activity may not be detected by this assay. Potential for adverse effects via unknown mechanism.	Perform *in vitro* ER, AR,TR, S assays with added metabolising system.	Consider route of exposure for OECD TG 408 assay and possible implications for differences from existing assay. Effects seen in existing studies may be in a more sensitive life stage. Further mechanistic studies may strengthen weight of evidence. Equivocal results may indicate chemical has multiple MOA.
Q	–	Eq/0	–	No evidence for (anti)-E,A,T,S activity in OECD TG 408. Weak (anti)-E,A,S activity may not be detected by this assay. No evidence of adverse effects.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system.	If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). Further mechanistic studies may strengthen weight of evidence.
R	–	Eq/0	Eq/0	No evidence for (anti)-E,A,T,S activity in OECD TG 408. Weak (anti)-E,A,S activity may not be detected by this assay.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system, otherwise Eq result available.	Further mechanistic studies may strengthen weight of evidence. Check data on chemical analogues. Equivocal results may indicate chemical has multiple MOA.

C.3.5. Combined Chronic Toxicity/Carcinogenicity Studies (OECD TG 451-3)

Status: Assay validated by the OECD.

743. Modalities detected: (anti)estrogen, (anti)androgen, thyroid, steroidogenesis.

744. Endpoints: Weight of adrenals, epididymides, ovaries, testes, thyroid, uterus (chronic toxicity studies). Pathological and histopathologic changes in adrenals, cervix, coagulating gland, epididymides, mammary glands, ovaries, pituitary, prostate, seminal vesicles, testes, thyroid gland, uterus.

Background to the assay

745. These assays determine the general toxicity (OECD TG 452 and TG 453) and carcinogenicity (OECD TG 451 and TG 453) of chemicals in laboratory animals after exposure for a period lasting most of the lifespan. Route of administration may be oral, dermal or inhalation, although oral via diet is the most common. OECD TG 453 was revised in September 2009 and replaced OECD TG 451 (older studies may have used OECD TG 451). General toxicity studies usually have a duration of 12 months whilst carcinogenicity studies usually have a duration of 18 or 24 months depending on the species tested. They provide information on major toxic effects, target organ toxicity and carcinogenicity. Although they have not been validated for the detection of endocrine disruptors (EDs), they contain many endpoints that are suitable for the determination of endocrine effects. Organ weights are not always included in the carcinogenicity phases of these studies as neoplastic changes may confound them, but they are generally determined at 12 months. A comparison can be made with validation of the OECD TG 407 (28-Day Oral Toxicity Study) for endocrine endpoints (OECD, 2006) where substances that were moderate and strong EDs for (anti)estrogenicity and (anti)androgenicity (e.g. ethinylestradiol and flutamide) and weak and strong modulators of thyroid hormone-related effects (e.g. propylthiouracil, T4 and methyl testosterone) were detected. Steroidogenesis inhibition was also detected although only one (potent) chemical was used in the validation study (CGS 18320B). OECD TG 453 and TG 452 are likely to be more sensitive than OECD TG 407 because of the extended dosing period and the larger number of animals per group (20 or 50 rodents per sex per group for chronic or carcinogenicity studies respectively compared with 5 in OECD TG 407). OECD TG 453 and TG 452, however, do not contain some sensitive endpoints (e.g. thyroid hormones, estrous cyclicity) that may be included in OECD TG 407.

When/why the assay may be used

746. These assays are likely to be used as part of a pesticide submission package and form part of the standard information requirements in certain chemical legislations (e.g. REACH for chemicals which are manufactured or imported in quantities of 1 000 tonnes or more). At least three dose levels are included so that an estimate of no-adverse-effect-levels or point of departure for benchmark doses can be determined. The assays are used for hazard identification/characterisation. It should be noted that as these

assays are not primarily designed to detect endocrine disruption, a higher degree of systemic toxicity is typically induced than is the case with the other Level 3 and 4 assays. The possibly confounding effect of systemic toxicity on endocrine endpoints therefore needs to be considered.

747. In order to provide information relevant for assessing whether or not a chemical may fulfil the WHO/IPCS (2002) definition of an ED, the study design has to be sufficiently robust to demonstrate the presence or absence of effects. In the dose selection, the investigator should also consider and ensure that data generated are adequate to fulfil the regulatory requirement across OECD countries as appropriate (e.g. hazard and risk assessment and labelling, ED assessment, etc.). The top dose or concentration should be sufficiently high to give clear systemic (i.e. non endocrine-specific) toxicity in order to ensure that a wide range of exposures (high to low) is tested. However, endocrine effects observed solely in the presence of clear systemic toxicity should be interpreted with caution and may be disregarded when sufficiently justified to be caused by secondary effects which are unlikely to be due to endocrine activity. The reason for this advice is a concern that some endocrine active substance (EAS) sensitive assays are being run at doses/concentrations of EASs that are too low to trigger direct impacts on the endocrine system. This guidance document is not the place to address this issue directly, but it should be considered when EAS-sensitive test guidelines (TGs) are revised in the future. In addition, the number and spacing of dose/concentration levels should also be adequate to fulfil the objectives of the study (e.g. to demonstrate dose response relationships if this is required).

Introduction to the table of scenarios

748. Table C.3.5 gives guidance on a further step to take in the event of a positive (+) or negative (-) result and in the presence of positive (+), negative (-) or equivocal/absent (Eq/0) existing results. “Existing results” are subdivided into “mechanism” and “effects” data (third and fourth columns). The table is divided horizontally into a series of scenarios that represent all the combinations of these events.

749. The results of OECD TG 451-3 are given in the second column. As they are not screening tests where a yes/no (qualitative) answer is obtained for the test as a whole, positive results would generally be assessed for individual endpoints. For the purposes of this guidance, however, a positive result is defined as a biologically significant change in any of the endocrine endpoints listed above (e.g. statistically significant reductions in reproductive organ weights). Changes in related endpoints will increase their biological significance (e.g. changes in the weights of testes and epididymides accompanied by histopathological/neoplastic changes). The guidance on histopathologic changes in endocrine tests (OECD, 2009) may be helpful in interpretation. A negative result for OECD TG 451-3 is taken to be the absence of biologically significant changes in all endocrine endpoints.

750. In the absence of other pertinent lines of evidence, negative results in this test alone cannot be taken as evidence that the substance is not an ED. Further studies will be required as confirmation.

751. Equivocal results for the guideline are not considered in Table C.3.5, partly for brevity but also because equivocal results are by nature uncertain. A decision must eventually be reached about whether the endocrine endpoints tend to be positive or negative, whether the result must be put to one side, or whether further testing should be

carried out. Factors which may have interfered with the result (e.g. composition of the diet used, environmental influences) should be considered.

Existing data to be considered

752. Existing "mechanism" *in vitro* data are assumed to be available from estrogen receptor (ER-), androgen receptor (AR-) and steroidogenesis-based assays (Level 2). Assays may also be available for interference with thyroid modalities. In practice, it is possible that data from all of these assays may not be available, so judgement will need to be used to decide which assays to perform. Although the current *in vitro* test guidelines do not incorporate metabolic activation, published information on use of metabolic activation systems is available in Jacobs et al. (2008; 2013) and OECD (2008). These methods, however, have not yet been validated.

753. Existing "effects" data refer to *in vivo* effects that may come from Level 3 or 4 tests in the Conceptual Framework (e.g. UT or H assays); or other sub-chronic repeat dosing studies (e.g. OECD TG 407 [28-day] or OECD TG 408 [90-day]). In these cases, it should be remembered that these assays are specifically designed to be sensitive to EASs. As mentioned above, the results of these studies may be interpreted as part of a battery or group of tests carried out for regulatory purposes. Data may also be available on effects in mammalian and non-mammalian wildlife species, although caution should be used when extrapolating between taxa. A chemical causing endocrine effects in in non-mammalian environmental species (fish, for example) may also have endocrine effects in mammals, but the physiological consequences of the effects are likely to be different.

754. When considering the results of OECD TG 451-3 assays, all available data should be used in order to reach a conclusion and a weight of evidence approach taken. This may include high throughput screening data, read-across data from structural analogues and quantitative structure activity relationships (QSARs). Several QSAR models for ER and AR binding/activation are now available (see Sections B.1.1.1 and B.1.1.2).

Scenarios: Positive and negative results combined with existing data

755. A series of scenarios (A to R) are presented in Table C.3.5 and represent all the possibilities of positive or negative results in combination with the presence or absence of existing data. The action taken will also depend on the regulatory environment, but the considerations given here are generally science based. Although OECD TG 451-3 assays primarily use rodents, the well-conserved nature of the hormonal pathways across taxa indicate that results on endocrine endpoints in these assays may be relevant to other vertebrate species. Effects in laboratory mammal tests are also highly relevant for environmental mammalian species. Wherever possible, the recommended "next step which could be taken" avoids unnecessary animal testing. However, sometimes conducting an animal test will be indicated and then the relevance of species, strain and exposure route should always be considered. The sensitivity and physiological function of the hormone under investigation in the test species should also be considered. In general, lower level tests should be conducted before higher level tests in order to avoid unnecessary animal usage, unless it is apparent that a Level 5 test will be required anyway or will be needed to establish the evidence to conclude on ED properties. At Level 5, the Extended One-Generation Reproduction Toxicity Study (EOGRTS – OECD TG 443) is the most sensitive reproduction assay for detecting endocrine disruption because it includes evaluation of a number of endocrine endpoints not included in the two-generation study (OECD TG 416)

adopted in 2001. It is recognised, however, that some jurisdictions may require a two-generation study. Further considerations specific to each scenario are given in the table.

756. Scenarios A to C represent positive results in the OECD TG 451-3 assays in the presence of positive *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. A positive result in the *in vitro* assays in combination with a positive OECD TG 451-3 assay is moderate or strong evidence for estrogen/androgen/thyroid/steroidogenesis (E,A,T,S-) mediated activity that may or may not be supported by the *in vivo* effects data. In the absence of robust upper-level data, the next step may be to conduct an upper-level test. In the presence of robust *in vivo* data, there may be sufficient evidence to conclude concern for endocrine disruption and therefore no need for further testing. Positive results in the OECD TG 451-3 assays also indicate the potential for endocrine mediated effects in lower vertebrates. These could be followed up with partial life cycle tests such as the Fish Sexual Development Test, a Larval Amphibian Growth and Development Assay or a Medaka Extended One-Generation Reproduction Test if the evidence were strong enough. *In vivo* assays/tests with negative results should be interpreted with caution as they may either indicate that the tests used do not have sufficient power to detect weak effects or, alternatively, that the effects do not present a concern for endocrine disruption. The possibility of other (non-E,A,T,S) mechanisms should also not be overlooked (e.g. involving other receptors or endocrine axes).

757. Scenarios D to F represent positive results in the OECD TG 451-3 assays in the presence of negative *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. Negative results in the *in vitro* assays should be viewed with caution in case a metabolite is responsible for the positive OECD TG 451-3 result. Unless the metabolic profile of the test substance is known, one option may be to conduct these assays with an added metabolising system. If the metabolic profile is known, then a higher level *in vivo* test may be advisable. The choice of tests will depend on the available *in vivo* effects data. Positive results in the OECD TG 451-3 assays may also indicate the potential for endocrine mediated effects in lower vertebrates. As in Scenarios A to C, *in vivo* assays/tests with negative results should be interpreted with caution as they may either indicate that the tests used do not have sufficient power to detect weak effects or, alternatively, that the effects do not present a concern for endocrine disruption.

758. Scenarios G to I represent positive results in the OECD TG 451-3 assays in the presence of various combinations of missing or equivocal data. Positive results in the OECD TG 451-3 assays also indicate the potential for endocrine mediated effects in lower vertebrates. The next step to take in these eventualities will depend on the nature of the other available data and the jurisdiction in which it is being used. In some cases, equivocal data may be viewed as positive whilst in others it may or may not contribute to the weight of evidence. The interpretation may also depend on the mode of action (MOA) in question and why the data are considered equivocal, e.g. a study that is equivocal for thyroid effects may still be of value in evaluating (anti)androgenic effects. In all three scenarios, the recommended first step is to obtain reliable mechanistic (*in vitro*) data rather than proceed further with *in vivo* testing. Equivocal and missing data are alternative scenarios and two possibilities for the next step are given in most cases, but the nature of equivocal data means that decisions need to be taken on a case-by-case basis. In all cases, the role of metabolism, route of exposure and data from structural analogues should be considered before deciding on the next step.

759. Scenarios J to L represent negative results in the OECD TG 451-3 assays in the presence of positive *in vitro* mechanistic data and positive, negative or equivocal *in vivo*

effects data. Negative outcomes in OECD TG 451-3 should be viewed with caution because of the power of the assay to detect (anti)estrogens and androgens may be limited All three scenarios could also arise from a chemical that is positive in *in vitro* assays but is metabolised to a non-active metabolite leading to negative results in the OECD TG 451-3 assays. This should be considered when investigating the next step. Endocrine-active potency may also explain differences between *in vitro* and *in vivo* results (e.g. a chemical with weak endocrine activity may give a positive result *in vitro* but may be negative *in vivo*). Positive *in vivo* effects data may involve other E,A,T,S, non-E,A,T,S mechanisms (e.g. involving other receptors or endocrine axes), more sensitive endpoints, greater statistical power or life stages that are more sensitive to the substance than the young adult exposed animals in OECD TG 451-3.

760. Scenarios M to O represent negative results in the OECD TG 451-3 assays in the presence of negative *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. Negative results for all tests (Scenario N) may be sufficient to enable a conclusion of no concern for endocrine disruption. This will depend on the weight of evidence and may not be possible. Where there are positive *in vivo* effects data, there could still be an E,A,T,S-related mechanism, the effects may be related to length of exposure, route of exposure or exposure at different life stages. Other E,A,T,S or non-E,A,T,S mechanisms may also be involved.

761. Scenarios P to R represent negative results in the OECD TG 451-3 assays in the presence of various combinations of missing or equivocal data. As with the positive result scenarios above (see Paragraph 760), the next step to take in these eventualities will have to be decided on a case-by-case basis. However, the recommended first step is generally to obtain reliable mechanistic (*in vitro*) data rather than proceed further with *in vivo* testing. In all cases, the role of metabolism, route of exposure and data from structural analogues should be considered before deciding on the next step.

762. In all scenarios (A to R), the next step to take to strengthen weight of evidence will depend on the existing information. The table is meant to provide a succinct guide and may not cover all circumstances or possibilities. The scenarios may also suggest that chemicals have simple or single MOA, when in practice they may have multiple endocrine and non-endocrine MOA. In some cases, for example, two opposite modes of simultaneous action (e.g. estrogenic and anti-estrogenic) could, depending on dose, lead to a minimisation or abolition of effects, while in others two different MOA (e.g. estrogenic and anti-androgenic) could potentially reinforce effects. Endocrine pathways interact, mixed effects are common and there are many pathways that cannot be distinguished with currently available TGs. If multiple MOA are suspected, either from the existing results or based on QSAR/read-across/integrated approaches, this should be investigated further if needed for regulatory decision making.

References

Jacobs, M. et al. (2013), "*In vitro* metabolism and bioavailability tests for endocrine active substances: What is needed next for regulatory purposes?", *ALTEX – Alternatives to Animal Experimentation*, Vol. 30/3, pp. 331-351.

Jacobs, M.N. et al. (2008), "The use of metabolising systems for *in vitro* testing of endocrine disrupters", *Current Drug Metabolism*, Vol. 9/8, pp. 796-826.

OECD (2009), "Guidance document for histologic evaluation of endocrine and reproductive tests in rodents", OECD Series on Testing and Assessment, No. 106, OECD, Paris, www.oecd.org/env/ehs/testing/43411534.pdf.

OECD (2008), *Detailed Review Paper on the Use of Metabolising Systems for* In Vitro *Testing of Endocrine Disrupters*, OECD Series on Testing and Assessment, No. 97, OECD Publishing, Paris, https://doi.org/10.1787/9789264085497-en.

OECD (2006), "Report of the validation of the updated test guideline 407: Repeat dose 28-day oral toxicity study in laboratory rats", OECD Series on Testing and Assessment, No. 59, OECD, Paris, www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2006)26&doclanguage=en.

WHO/IPCS (2002), "Global assessment of the state-of-the-science of endocrine disrupters", Damstra, T. et al. (eds.) WHO/PCS/EDC/02.2, World Health Organization, Geneva, www.who.int/ipcs/publications/new_issues/endocrine_disruptors/en.

Table C.3.5. **Combined Chronic Toxicity/Carcinogenicity Studies (OECD TG 451-3): Guidance for scenarios of combinations of results with existing data**

This table represents possible conclusions to be drawn from assay data, and a next step which could be taken if further evidence is required about possible endocrine disrupting properties and/or effects. The guidance offered is not meant to be prescriptive, but provides science-based considerations. It encourages the use of all available data and expert judgement in a weight of evidence approach. Regional and national interpretation of results and "next steps" may vary.

The conclusions are grouped into a series of scenarios (A-R), each scenario representing a different combination of assay results, existing *in vitro* data and existing *in vivo* data. The symbol "+" indicates that the data in question represent a positive result, «-" indicates a negative result, and "Eq/0" indicates that the data are either equivocal or are not available.

Existing results: * "Mechanism (*in vitro* mechanistic data)" assumes that mechanistic data are available from estrogen receptor (ER-), androgen receptor (AR-) and steroidogenesis-based assays (Level 2). Thyroid hormone receptor (TR) and other assays concerning mechanisms of thyroid disruption may be available, but they are not in common use. In practice, data from all assays may not be available and therefore this must be taken into account when deciding on the "next step". Quantitative structure activity relationship (QSAR) predictions of estrogen and androgen binding/activation may be made for some substances.

Existing results: ** "Effects (*in vivo* effects of concern)" assumes various information, such as data from repeat dose oral toxicity studies, reproduction/developmental toxicity screen tests, read-across from analogues, will be available.

*** *Note:* a positive result is defined as a biologically significant change in any of the endocrine endpoints.

Scenarios	Result of OECD TG 451-3 (carc/1-year) assay	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
A	+ ***	+	+	(Anti)-E,A,T,S activity. Increased evidence of (anti)-E,A,T,S activity.	Perform assay from Level 5 (e.g. EOGRTS or two-gerenation assay).	If existing data are from a Level 5 assay, there may be sufficient information to conclude evidence of concern for endocrine disruption (the EOGRTS provides the most information; however, for endocrine disrupting chemicals [EDCs] with a carcinogenic potential, OECD TG 451-3 may be more sensitive). Consider route of exposures for effects data and possible implications of absorption, distribution, metabolism and excretion (ADME) characteristics of the chemical. Endocrine activity possible in lower vertebrates. Consider performing a Fish Sexual Development Test (FSDT), a Larval Amphibian Growth and Development Assay (LAGDA) or a Medaka Extended One-Generation Reproduction Test (MEOGRT).
B	+	+	–	(Anti)-E,A,T,S activity. Increased evidence of (anti)-E,A,T,S activity.	Perform assay from Level 5 (e.g. EOGRTS or two-gerenation assay).	If existing data are from an adequate Level 5 study, question why there are differences. Consider route of exposures and possible implications of ADME characteristics of the chemical. If existing data are from a less sensitive assay, a higher level test may be required. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT.
C	+	+	Eq/0	(Anti)-E,A,T,S activity. Increased evidence of (anti)-E,A,T,S activity.	Perform assay from Level 5 (e.g. EOGRTS or two-gerenation assay).	Check data on chemical analogues. Consider route of exposure for OECD TG 452/453 and follow-up assay. Possible implications of ADME characteristics of the chemical. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT. Equivocal results may indicate chemical has multiple modes of action (MOA).
D	+	–	+	(Anti)-E,A,T,S activity. Acts via non-estrogen receptor (ER), androgen receptor (AR), thyroid hormone receptor (TR), steroidogenesis (S) mechanism or requires metabolic activation for activity.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system.	If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude evidence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). Consider route of exposures for effects data and possible implications of ADME characteristics of the chemical. Further mechanistic studies may help determine MOA. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT.
E	+	–	–	(Anti)-E,A,T,S activity. Acts via non-ER, AR, TR, S mechanism or requires metabolic activation for activity. Route of exposure may account for the differences between OECD TG 451-3 and existing data	Perform *in vitro* ER, AR, TR, S assays with added metabolising system OR Perform assay from Level 5 (e.g. EOGRTS or two-gerenation assay).	If existing data are from an adequate Level 5 assay, question why there are differences. Consider route of exposures and possible implications of ADME characteristics of the chemical. If existing data are from a less sensitive assay, a higher level test may be required. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT.

Scenarios	Result of OECD TG 451-3 (carc/1-year) assay	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
F	+	–	Eq/0	(Anti)-E,A,T,S activity. Acts via non-ER, AR, TR, S mechanism or requires metabolic activation for activity.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system OR Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	Check data on chemical analogues. Further mechanistic studies may help determine MOA. Upper-level studies will provide more hazard data. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT. Equivocal results may indicate chemical has multiple MOA.
G	+	Eq/0	+	(Anti)-E,A,T,S activity. May act via ER, AR, TR, S mechanism (metabolic activation needed).	Perform *in vitro* ER, AR, TR, S assays (for the "0" scenario, otherwise Eq result available) OR Perform *in vitro* ER, AR, TR, S assays with added metabolising system.	If existing data are from a Level 5 assay, there may be sufficient information to conclude evidence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). Check data on chemical analogues. Further mechanistic studies may help determine MOA. Consider route of exposures for effects data and possible implications of ADME characteristics of the chemical. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT. Equivocal results may indicate chemical has multiple MOA.
H	+	Eq/0	–	(Anti)-E,A,T,S activity. Acts via unknown mechanism or may require metabolic activation for activity. Route of exposure may account for the differences between OECD TG 451-3 and existing data.	For the "0" scenario, perform *in vitro* ER, AR, TR, S assays with added metabolising system (otherwise Eq result available).	If existing data are from an adequate Level 5 assay, question why there are differences. Consider route of exposures and possible implications of ADME characteristics of the chemical. If existing data are from a less sensitive assay, a higher level test may be required. Check data on chemical analogues. Further mechanistic studies may help determine MOA. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT. Equivocal results may indicate chemical has multiple MOA.
I	+	Eq/0	Eq/0	(Anti)-E,A,T,S activity. Acts via unknown mechanism. Unknown potential for adverse effects.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system OR Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	Check data on chemical analogues. Further mechanistic studies may help determine MOA. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT. Equivocal results may indicate chemical has multiple MOA.
J	–	+	+	No evidence for (anti)-E,A,T,S activity in OECD TG 451-3. Effects on reproduction will not be detected by these assays. Metabolism or potency explains the difference from existing *in vitro* and *in vivo* data.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system OR Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	If existing data are from an adequate Level 5 assay, question why there are differences. Effects seen in existing studies may be in a more sensitive life stage. Consider route of exposures and possible implications of ADME characteristics of the chemical. Further mechanistic studies may help determine MOA.

Scenarios	Result of OECD TG 451-3 (carc/1 year) assay	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
K	–	+	–	No evidence for (anti)-E,A,T,S activity in OECD TG 451-3. Effects on reproduction will not be detected by these assays. Metabolism or potency explains *in vitro*/*in vivo* differences.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system OR Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from a less sensitive assay, a higher level test may be required. Further mechanistic studies with metabolism may help determine MOA.
L	–	+	Eq/0	No evidence for (anti)-E,A,T,S activity in OECD TG 451-3. Effects on reproduction will not be detected by these assays. Metabolism or potency explains *in vitro*/*in vivo* differences. Unknown potential for adverse effects.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system OR Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	Metabolic deactivation of chemical may occur *in vivo* so that possible *in vitro* activity is not realised. Consider possible routes of exposure implications of metabolism. Equivocal results may indicate chemical has multiple MOA.
M	–	–	+	No evidence for (anti)-E,A,T,S activity in OECD TG 451-3. Effects on reproduction will not be detected by these assays.. Effects seen in existing studies are via non-E,A,T,S mechanism.	Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	If existing data are from an adequate Level 5 assay, question why there are differences. Effects seen in existing studies may be in a more sensitive life stage. Consider route of exposures and possible implications of ADME characteristics of the chemical.
N	–	–	–	No evidence for (anti)-E,A,T,S activity in OECD TG 451-3. Weak (anti)-E,A,S activity may not be detected by this assay. No evidence for (anti)-E,A,T,S activity *in vitro*. No evidence of adverse effects.	Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive).
O	–	–	Eq/0	No evidence for (anti)-E,A,T,S activity in OECD TG 451-3. Effects on reproduction will not be detected by these assays. No evidence for (anti)-E,A,T,S activity *in vitro*. Unknown potential for adverse effects.	Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	Consider route of exposures and possible implications for ADME characteristics of the chemical in follow-up assay.

Scenarios	Result of OECD TG 451-3 (carc/1 year) assay	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
P	–	Eq/0	+	No evidence for (anti)-E,A,T,S activity in OECD TG 451-3. Effects on reproduction will not be detected by these assays. Potential for adverse effects via unknown mechanism.	Perform *in vitro* ER, AR,TR, S assays with added metabolising system.	Consider route of exposure for OECD TG 452/453 and possible implications for differences from existing assay. Effects seen in existing studies may be in a more sensitive life stage. Equivocal results may indicate chemical has multiple MOA.
Q	–	Eq/0	–	No evidence for (anti)-E,A,T,S activity in OECD TG 451-3. Effects on reproduction will not be detected by these assays. No evidence of adverse effects.	Perform *in vitro* ER, AR,TR, S assays with added metabolising system.	If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). Further mechanistic studies may strengthen weight of evidence.
R	–	Eq/0	Eq/0	No evidence for (anti)-E,A,T,S activity in OECD TG 451-3. Effects on reproduction will not be detected by these assays.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system, otherwise Eq result available.	Further mechanistic studies may strengthen weight of evidence. Check data on chemical analogues. Equivocal results may indicate chemical has multiple MOA.

C.3.6. Reproduction/Developmental Toxicity Screening Test (OECD TG 421) and Combined Repeated Dose Toxicity Study with the Reproduction/Developmental Toxicity Screening Test (OECD TG 422)

Status: Assay validated by the OECD.

763. Modalities detected: (anti)estrogen, (anti)androgen, steroidogenesis thyroid.

764. Endpoints: Estrous cyclicity, time to mating, male fertility, female fertility, dystocia, gestation length, number of implantations and corpora lutea, number of live births and post-implantation loss, litter size, sex ratio, litter/pup weight, pup survival index, anogenital distance (PND 0-4) and nipple retention (PND 13) in pups.

Weights of: (parents only) testes, epididymides, prostate and seminal vesicles with coagulating glands (OECD TG 421 and TG 422), plus adrenals (OECD TG 422 only). Optional organ weights could include levator ani plus bulbocavernosus muscle complex, Cowper's glands and glans penis in males, and paired ovaries and uterus (including cervix). Thyroid gland weight could be measured (after fixation) in all adults and PND 13 pups (OECD TG 421 and TG 422).

Vaginal smears at necropsy to determine stage of the estrous cycle and allow correlation with histopathology of ovaries.

Histopathologic changes in: testis, epididymides, ovaries (parents only); thyroid (parents and pups "when necessary") (OECD TG 421 and TG 422). Also adrenals, uterus and cervix, prostate, seminal vesicles plus coagulating glands, in parents only (OECD TG 422 only). Mammary glands and pituitary are optional (OECD TG 422).

Mandatory measurement of serum thyroid hormone (T4) in parental males and pups at PND 13. T4 in dams and PND 4 pups "if relevant". TSH (and other hormones) optional "if relevant".

Background to the assay

765. These assays are designed to provide limited information about the effects of a chemical on the male and female reproductive systems including gonadal function, mating, conception, gestation, development of the conceptus and parturition. The assays are designed for use with the rat. The recommended route of administration is oral, usually via gavage although administration may be via diet or drinking water. Although the titles of the test guidelines (TGs) imply that they are screening tests, they are not screens as given in the definition in Section A, but are apical assays. The TGs have similar experimental schedules, but OECD TG 422 includes a more detailed assessment of repeated dose toxicity and thus more endpoints. The studies were not designed to detect endocrine active substances (EASs), but they have endpoints relevant for the assessment of possible endocrine disruption and provide data on adverse effects related to reproduction and development. The developing fetus is a life stage that may be particularly sensitive to EASs.

The scope of these assays is much smaller than OECD TG 416 and TG 443 (Two-generation Assay and Extended One-Generation Reproductive Toxicity Study), e.g. duration of premating exposure is much shorter, group sizes are generally half and only a relatively short period of postnatal development (13 days) is included. OECD TG 421 was originally adopted in 1995 and TG 422 in 1996. Following a feasibility study (OECD, 2015), both assays were updated in July 2016 to include some endocrine-relevant endpoints as the exposure periods cover some of the sensitive periods during development (pre- or early postnatal periods). It should be noted that if the assays were conducted before 2016, they are unlikely to include these extra endpoints.

766. Anogenital distance (AGD) and nipple retention are included in OECD TG 421 and TG 422 as sensitive endpoints of endocrine effects; however, their utility as apical endpoints or as biological indicators of endocrine action may require further experience in their use. Increased nipple retention and reduced AGD in male offspring are hallmarks of anti-androgenicity. Nevertheless, "retained nipples/areolae" as a qualitative endpoint may have high biological variability (e.g. Melching-Kollmuss et al. [2017]) and alteration of AGD can occur via other modes of action (MOA) (e.g. Miyagawa et al. [2011]; Seifert et al. [2009]). However, current OECD guidance on these endpoints can be found in OECD GD 43 and GD 151 and it is clear that these should be considered as apical endpoints. With regard to AGD, OECD GD 43 (OECD, 2008b) states, "A statistically significant change in [anogenital distance] that cannot be explained by the size of the animal indicates effects of the exposure and should be used for setting the [no observed adverse effect level (NOAEL)]". With regard to nipple retention, OECD GD 151 (OECD, 2013) states "a statistically significant change in nipple retention should be evaluated similarly to an effect on anogenital distance as both endpoints indicate an adverse effect of exposure and should be considered in setting a NOAEL".

767. The feasibility report on OECD TG 421 and TG 422 (OECD, 2015) indicated that the sensitivity for detecting effects based on qualitative nipple retention (i.e. the number of males with or without nipples) was quite low irrespective of the number of litters included. However, nipple retention is a sensitive endpoint if measured quantitatively (i.e. if the number of nipples from 0-12 is recorded). This endpoint of quantitative nipple retention in the male pups was therefore included in these study updates.

768. Both original (1996) and revised (2017) OECD TG 421 and TG 422 contain endpoints that are suitable for the determination of endocrine effects although the TGs as revised in 2017 are more comprehensive. In addition to reproduction/development, OECD TG 421 and TG 422 may both provide information about potential endocrine effects on male reproductive organs. Information about effects on the thyroid are included in TG 422 and TG 421 as revised in 2017. Female reproductive organs are also examined, but detection of endocrine effects in these organs may be obscured because of pregnancy. Male animals are dosed for a total period of 28 days. A comparison can be made with OECD TG 407 (28-day oral toxicity study) where validation studies (OECD, 2006) demonstrated that substances that were moderate and strong endocrine disruptors (EDs) for (anti)estrogenicity and (anti)androgenicity (e.g. ethinylestradiol and flutamide) and weak and strong modulators of thyroid hormone-related effects (e.g. propylthiouracil, T4 and methyl testosterone) were detected. Steroidogenesis inhibition was also detected, although only one (potent) chemical was used in the validation study (CGS 18320B).

769. The 2017 revised assays contain many endocrine-sensitive endpoints that are also included in the Conceptual Framework (CF) Level 5 assays – OECD TG 416 and TG 443. Many of these endpoints indicate (anti)estrogenicity, (anti)androgenicity and thyroid

hormone disruption and may help to discern MOA. Information on mode of action (MOA) from *in vitro* estrogen/androgen/thyroid/steroidogenesis (E,A,T,S) assays or *in vivo* lower tier tests such as a Uterotrophic Bioassay (UT) and a Hershberger Bioassay (H) is also helpful. The original (1996) TGs, however, only contained apical endpoints, therefore in assays conducted prior to 2016 it would be difficult to discern MOA from these tests alone.

Thyroid hormone analysis and interpretation in OECD TG 421/422

770. The revised TGs require the determination of serum thyroid hormones in adults and pups. Detection/measurement of thyroid hormones (T4, T3, TSH) in pups can be challenging. In order to provide assurance that thyroid hormone measurements are reliable, laboratories should be able to demonstrate that they are proficient with the assay. Only such a demonstration would enable regulatory authorities to interpret and use the data. A recent workshop on "Practicability of Hormonal Measurements" was organised by the BfR (Germany) and the finding from this workshop will be published (Kucheryavenko et al., 2018). The OECD Expert Group on Reproductive and Developmental Toxicity recommends that to demonstrate proficiency for thyroid hormone measurement, a laboratory should be able to show results from a separate study using a positive control substance. Laboratories may also submit their calibration curves, standard curves, as well as data on the levels of quantification and detection. This group is also establishing a historical control database with thyroid toxicant positive controls.

771. Within OECD TG 421/422, T4 should be measured in adult males and in pups at PND 13, while in pups at PND 4 and in females T4 would be measured "if relevant". Triggers for further measurement may be changes in the initial endpoints, to provide clarity on the time course of thyroid changes. This may also prompt measurement of TSH and possibly T3.

When/why the assay may be used

772. These assays are frequently used for initial hazard assessments for chemicals, as part of the Screening Information Data Set (SIDS) for the assessment of chemicals for which there is little information or for dose setting for more extensive reproduction/developmental assays. These assays are also likely to be used as part of a pesticide submission package and forms part of the standard information requirements in certain chemical legislations (e.g. REACH in the European Union). At least three dose levels are included so that an estimate of no-adverse-effect-level can be determined and the assay used for hazard identification/characterisation.

773. In order to provide information relevant for assessing whether or not a chemical may fulfil the WHO/IPCS (2002) definition of an ED, the study design has to be sufficiently robust to demonstrate the presence or absence of effects. In the dose selection, the investigator should also consider and ensure that data generated are adequate to fulfil the regulatory requirement across OECD countries as appropriate (e.g. hazard and risk assessment and labelling, ED assessment, etc.). The top dose or concentration should be sufficiently high to give clear systemic (i.e. non endocrine-specific) toxicity in order to ensure that a wide range of exposures (high to low) is tested. However, endocrine effects observed solely in the presence of clear systemic toxicity should be interpreted with caution and may be disregarded when sufficiently justified to be caused by secondary effects which are unlikely to be due to endocrine activity. The reason for this advice is a concern that some EAS-sensitive assays are being run at doses/concentrations of EASs that are too low to trigger direct impacts on the endocrine system. This guidance document is not the place

to address this issue directly, but it should be considered when EAS-sensitive TGs are revised in the future. In addition, the number and spacing of dose/concentration levels should also be adequate to fulfil the objectives of the study (e.g. to demonstrate dose response relationships if this is required).

Introduction to the table of scenarios

774. Table C.3.6 gives guidance on a further step to take in the event of a positive (+) or negative (-) result and in the presence of positive (+), negative (-) or equivocal/absent (Eq/0) existing results. "Existing results" are subdivided into "mechanism" and "effects" data (third and fouth columns). The table is divided horizontally into a series of scenarios that represent all the combinations of these events.

775. The results of OECD TG 421/422 are given in the second column. As these are not tests where a yes/no (qualitative) answer is obtained for the test as a whole, positive results would generally be assessed for individual endpoints and as a whole. It is not possible to provide guidance on all endpoints individually, therefore the endpoints have been pragmatically divided into "apical" and "indicators of hormonal activity". The terminology used has been chosen to be consistent between both the non-mammalian wildlife and mammalian tests. Both groups have similar biological importance, although the "indicators of hormonal activity" in the mammalian assays are serum hormones and are generally, but not always, more variable than "apical endpoints".

776. For this guideline "apical" endpoints are reproductive and developmental parameters (including anogenital distance, presence of nipples, genital abnormalities), estrous cyclicity, weights and histopathologic changes in testes, epididymides, prostate, seminal vesicles (with coagulating glands), ovary, uterus, thyroid. "Indicators of hormonal activity" are hormones (T4, TSH).

777. Three possible outcomes for a positive result are therefore envisaged in Table C.3.6:

1. indicators of hormonal activity and apical endpoints positive
2. indicators of hormonal activity positive and apical endpoints negative
3. indicators of hormonal activity negative and apical endpoints positive.

778. A positive result for apical endpoints could be biologically significant changes in pup AGD, accompanied by treatment-related histopathologic changes in parental reproductive organs. A positive result for indicators of hormonal activity could be biologically significant changes in hormone profiles. A positive result for indicators of hormonal activity alone should be considered with caution, although it is possible that these endpoints may have detected weak effects that were not detected by the apical endpoints.

779. A negative result for OECD TG 421/422 is taken to be the absence of biologically significant changes in all endocrine endpoints.

780. In the absence of other pertinent lines of evidence, negative results in this test alone cannot be taken as firm evidence that the substance is not an ED. Further studies may be required as confirmation.

781. Equivocal results for the guideline are not considered in the table, partly for brevity but also because equivocal results are by nature uncertain. A decision must eventually be reached about whether the endocrine endpoints tend to be positive or negative or whether the result must be put to one side and the test repeated (using the same or a different test

guideline). Factors which may have interfered with the result (e.g. composition of the diet used, environmental influences) should be considered.

Existing data to be considered

782. Existing "mechanism" *in vitro* data are assumed to be available from estrogen receptor (ER-), androgen receptor (AR-) and steroidogenesis-based assays (Level 2). Assays may also be available for interference with thyroid modalities. In practice, it is possible that data from all of these assays may not be available, therefore judgement will need to be used to decide which assays to perform. Although the current *in vitro* test guidelines do not incorporate metabolic activation, published information on use of metabolic activation systems is available in Jacobs et al. (2008; 2013) and OECD (2008a). These methods, however, have not yet been validated.

783. Existing "effects" data refer to *in vivo* effects that may come from Level 3 or 4 tests in the Conceptual Framework (e.g. UT or H assays). In these cases, it should be remembered that these assays are specifically designed to be sensitive to EASs. Given the usage of these assays for general chemical testing, it is possible that an OECD TG 407 (28-day test) is available. It is unlikely that OECD TG 421/422 will be performed if higher tier reproduction/developmental toxicity data are already available, as they offer no advantage over these assays. The results of the study may also be interpreted as part of a battery or group of tests carried out for regulatory purposes. Data may also be available on effects in mammalian and non-mammalian wildlife species, although caution should be used when extrapolating between taxa. A chemical causing endocrine effects in non-mammalian environmental species (fish, for example) may also have endocrine effects in mammals, but the physiological consequences of the effects are likely to be different.

784. When considering the results of OECD TG 421/422, all available data should be used in order to reach a conclusion and a weight of evidence approach taken. This may include high throughput screening data, read-across data from structural analogues and quantitative structure activity relationship (QSAR). Several QSAR models for ER and AR binding/activation are now available (see Sections B.1.1.1 and B.1.1.2).

Scenarios: Positive and negative results combined with existing data

785. A series of scenarios (A to R) are presented in Table C.3.6 and represent all the possibilities of positive or negative results in combination with the presence or absence of existing data. The action taken will also depend on the regulatory environment, but the considerations given here are generally science based. Although rats are the preferred species for TG 421/422, the well-conserved nature of the hormonal pathways across taxa should be a strong indication that results in this assay may be relevant to other vertebrate species. Effects in laboratory mammal tests are also highly relevant for environmental mammalian species. Wherever possible, the recommended "next step which could be taken" avoids unnecessary animal testing. However, sometimes conducting an animal test will be indicated and then the relevance of species, strain, exposure route and species-specific metabolism should always be considered. The sensitivity and physiological function of the hormone under investigation in the test species should also be considered. In general, lower level tests should be conducted before higher level tests in order to avoid unnecessary animal usage, unless it is apparent that a Level 5 test will be required anyway or will be needed to establish the evidence to conclude on ED properties. Information on some endocrine-related tumours may be detected more comprehensively in

carcinogenicity studies (OECD TG 451/453) (Level 4); for example, detection of certain types of thyroid tumors in the absence of reproductive or developmental effects, as well as substances causing tumors in other endocrine-sensitive tissues. At Level 5, the Extended One-Generation Reproduction Toxicity Study (EOGRTS – OECD TG 443) is the most sensitive reproduction assay for detecting endocrine disruption because it includes evaluation of a number of endocrine endpoints not included in the two-generation study (OECD TG 416) adopted in 2001. It is recognised, however, that some jurisdictions may require a two-generation study. Further considerations specific to each scenario are given in the table.

786. Scenarios A to C represent positive results in OECD TG 421/422 in the presence of positive *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. Each positive result scenario is divided into the three possible outcomes given above. A positive result in the *in vitro* assays in combination with a positive OECD TG 421/422 is strong evidence of adverse effects on reproduction/development and/or endocrine organs via E,A,T,S mechanisms. Effects on only apical endpoints or only indicators of hormonal activity may assist with interpretation. Positive results in the OECD TG 421/422 assay may also indicate the potential for endocrine mediated effects in lower vertebrates. These could be followed up with partial life cycle tests such as the Fish Sexual Development Test (FSDT), the Larval Amphibian Growth and Development Assay (LAGDA) or the Medaka Extended One-Generation Reproduction Test (MEOGRT) if the evidence is strong enough. *In vivo* assays/tests with negative results should be interpreted with caution as they may either indicate that the tests used do not have sufficient power to detect weak effects or, alternatively, that the effects do not present a concern for endocrine disruption. The possibility of other (non-E,A,T,S) mechanisms should also not be overlooked (e.g. involving other receptors or endocrine axes).

787. Scenarios D to F represent positive results in OECD TG 421/422 in the presence of negative *in vitro* mechanistic data and positive, negative or equivocal in vivo effects data. Each positive result scenario is divided into the three possible outcomes given above. A positive result is strong evidence of adverse effects on reproduction/development and/or endocrine organs. Effects on the different apical endpoints or indicators of hormonal activity may assist with interpretation. Negative results in the *in vitro* assays should be viewed with caution in case a metabolite is responsible for the positive OECD TG 421/422. If the metabolic profile of the test substance is not known, performing the *in vitro* assays with addition of a metabolising system may help to understand mechanism. Positive results in the OECD TG 421/422 assay may also indicate the potential for endocrine mediated effects in lower vertebrates.

788. Scenarios G to I represent positive results in OECD TG 421/422 in the presence of various combinations of missing or equivocal data. Positive results in the OECD TG 421/422 assay may also indicate the potential for endocrine mediated effects in lower vertebrates. Each positive result scenario is divided into the three possible outcomes given above. The next step to take in these eventualities will depend on the nature of the other available data and the jurisdiction in which it is being used. In some cases, equivocal data may be viewed as positive whilst in others it may or may not contribute to the weight of evidence. The interpretation may also depend on the MOA in question and why the data are considered equivocal, e.g. a study that is equivocal for thyroid effects may still be of value in evaluating (anti)androgenic effects. In all three scenarios, the recommended first step is to obtain reliable mechanistic (*in vitro*) data rather than proceed further with *in vivo* testing. Equivocal and missing data are alternative scenarios and two possibilities for the next step are given in most cases, but the nature of equivocal data means that decisions

need to be taken on a case-by-case basis. In all cases, the role of metabolism, route of exposure and data from structural analogues should be considered before deciding on the next step.

789. Scenarios J to L represent negative results in OECD TG 421/422 in the presence of positive *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. A negative result is taken to be negative findings for both indicators of hormonal activity and apical endpoints (unlike the situation with positive outcomes), therefore there is only one possible negative outcome. All three scenarios could fit a chemical that is positive in *in vitro* assays but is metabolised to a non-active metabolite, leading to negative results in OECD TG 421/422. This possibility may be investigated to help understand mechanism. Endocrine active potency may also explain differences between *in vitro* and *in vivo* results (e.g. a a chemical with weak endocrine activity may give a positive result *in vitro* but may be negative *in vivo*). Positive *in vivo* effects data may involve E,A,T,S or non-E,A,T,S mechanisms (e.g. involving other receptors or endocrine axes), more sensitive endpoints, or greater statistical power, but knowledge of absorption, distribution, metabolism and excretion (ADME) may help to explain differences from the OECD TG 421/422 data.

790. Scenarios M to O represent negative results in the OECD TG 421/422 in the presence of negative *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. Negative results for all tests (Scenario N) indicate an absence of concern for reproductive toxicity via an endocrine disruption mechanism. If the study is not considered to be robust, supplemental testing could be considered. Positive *in vivo* effects data may involve E,A,T,S or non-E,A,T,S mechanisms (e.g. involving other receptors or endocrine axes), but knowledge of ADME may help to explain differences from the OECD TG 421/422 data.

791. Scenarios P to R represent negative results in OECD TG 421/422 in the presence of various combinations of missing or equivocal data. As with the positive result scenarios, the next step to take in these eventualities will have to be decided on a case-by-case basis. In all cases, the role of metabolism, route of exposure and data from structural analogues should be considered before deciding on the next step.

792. In all scenarios (A to R), the next step to take to strengthen weight of evidence will depend on the existing information. Table C.3.6 is meant to provide a succinct guide and may not cover all circumstances or possibilities. The scenarios may also suggest that chemicals have simple or single MOA, when in practice they may have multiple endocrine and non-endocrine MOA. In some cases, for example, two opposite modes of simultaneous action (e.g. estrogenic and anti-estrogenic) could, depending on dose, lead to a minimisation or abolition of effects, while in others two different MOA (e.g. estrogenic and anti-androgenic) could potentially reinforce effects. Endocrine pathways interact, mixed effects are common and there are many pathways that cannot be distinguished with currently available TGs. If multiple MOA are suspected, either from the existing results or based on QSAR/read-across/integrated approaches, this should be investigated further if needed for regulatory decision making.

References

Jacobs, M. et al. (2013), "*In vitro* metabolism and bioavailability tests for endocrine active substances: What is needed next for regulatory purposes?", *ALTEX – Alternatives to Animal Experimentation*, Vol. 30/3, pp. 331-351.

Jacobs, M.N. et al. (2008), "The use of metabolising systems for *in vitro* testing of endocrine disrupters", *Current Drug Metabolism*, Vol. 9/8, pp. 796-826.

Kucheryavenko, O. et al. (2018), "Report from the BfR Expert Hearing on Practicability of Hormonal Measurements", *Archives of Toxicology* (in prep.).

Melching-Kollmuss, S. et al. (2017), "Comparing effect levels of regulatory studies with endpoints derived in targeted anti-androgenic studies: Example prochloraz", *Archives of Toxicology*, Vol. 91/1, pp. 143-162, https://doi.org/10.1007/s00204-016-1678-y.

Miyagawa, S. et al. (2011), "The role of sonic hedgehog-Gli2 pathway in the masculinization of external genitalia", *Endocrinology*, Vol. 152/7, pp. 2894-2903, https://doi.org/10.1210/en.2011-0263.

OECD (2015), "Feasibility study for minor enhancements of TG 421/422 with ED relevant endpoints", OECD Series on Testing and Assessment, No. 217, OECD, Paris, www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2015)24&doclanguage=en.

OECD (2013), "Guidance document supporting of OECD test guideline on the extended one-generation reproductive toxicity study", OECD Series on Testing and Assessment, No. 151, OECD, Paris, www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/JM/MONO(2013)10&doclanguage=en.

OECD (2008a), *Detailed Review Paper on the Use of Metabolising Systems for* In Vitro *Testing of Endocrine Disrupters*, OECD Series on Testing and Assessment, No. 97, OECD Publishing, Paris, https://doi.org/10.1787/9789264085497-en.

OECD (2008b), "Guidance document on mammalian reproductive toxicity testing and assessment", OECD Series on Testing and Assessment, No. 43, OECD, Paris, www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/JM/MONO(2008)16&doclanguage=en.

OECD (2006), "Report of the validation of the updated test guideline 407: Repeat dose 28-day oral toxicity study in laboratory rats", OECD Series on Testing and Assessment, No. 59, OECD, Paris, www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2006)26&doclanguage=en.

Seifert, A.W. et al. (2009), "Multiphasic and tissue-specific roles of sonic hedgehog in cloacal septation and external genitalia development", *Development*, Vol. 136/23, pp. 3949-3957, http://dx.doi.org/10.1242/dev.042291.

WHO/IPCS (2002), "Global assessment of the state-of-the-science of endocrine disrupters", Damstra, T. et al. (eds.) WHO/PCS/EDC/02.2, World Health Organization, Geneva, www.who.int/ipcs/publications/new_issues/endocrine_disruptors/en.

Table C.3.6. **Reproduction/Developmental Toxicity Screening Test (OECD TG 421) and Combined Repeated Dose Toxicity Study with the Reproduction/Developmental Toxicity Screening Test (OECD TG 422): Guidance for scenarios of combinations of results with existing data**

This table represents possible conclusions to be drawn from assay data, and a next step which could be taken if further evidence is required about possible endocrine disrupting properties and/or effects. The guidance offered is not meant to be prescriptive, but provides science-based considerations. It encourages the use of all available data and expert judgement in a weight of evidence approach. Regional and national interpretation of results and "next steps" may vary.

The conclusions are grouped into a series of scenarios (A-R), each scenario representing a different combination of assay results, existing *in vitro* data and existing *in vivo* data. The symbol "+" indicates that the data in question represent a positive result, «-" indicates a negative result, and "Eq/0" indicates that the data are either equivocal or are not available.

Existing results: * "Mechanism (*in vitro* mechanistic data)" assumes that mechanistic data are available from estrogen receptor (ER-), androgen receptor (AR-) and steroidogenesis-based assays (Level 2). Thyroid hormone receptor (TR) and other assays concerning mechanisms of thyroid disruption may be available, but they are not in common use. In practice, data from all assays may not be available and therefore this must be taken into account when deciding on the "next step". Quantitative structure activity relationship (QSAR) predictions of estrogen and androgen binding/activation may be made for some substances.

Existing results: ** "Effects (*in vivo* effects of concern)" assumes effects have been observed in other *in vivo* screens/tests which give rise to concern that the test chemical may be an endocrine disrupter. These may be other repeated dose toxicity tests, UT and H assays or read-across from chemical analogues.

*** *Note:* three possible outcomes for a positive result are given:

1. indicators of hormonal activity and apical endpoints positive
2. indicators of hormonal activity positive and apical endpoints negative
3. indicators of hormonal activity negative and apical endpoints positive.

"Apical endpoints" are reproductive and developmental parameters (including anogenital distance, presence of nipples, genital abnormalities), estrous cyclicity, weights and histopathologic changes in testes, epididymides, prostate, seminal vesicles (with coagulating glands), ovary, uterus, thyroid.

"Indicators of hormonal activity" are hormones (T4, TSH).

Scenarios	Result of OECD TG 421/422	Existing results		Possible conclusions: 1) Indicators of hormonal activity and apical endpoints positive 2) Indicators of hormonal activity positive and apical endpoints negative 3) Indicators of hormonal activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
A	+ ***	+	+	1) Evidence of adverse effects on endocrine/apical endpoints via estrogen/androgen/thyroid/ steroidogenesis (E,A,T,S) mechanism. 2) Evidence of effects on hormonal endpoints via E,A,T,S mechanism. Apical endpoints may be less sensitive or unaffected. 3) Evidence of adverse effects via E,A,T,S mechanism. Indicators of hormonal activity may be less sensitive or unaffected.	Perform assay from Level 5 (e.g. Extended One-Generation Reproduction Toxicity Study [EOGRTS] or two-generation assay).	May be sufficient information to conclude evidence of concern for reproductive toxicity via endocrine disruption mechanism. Note that the EOGRTS provides the most information on endocrine disruption; however, for endocrine disrupting chemicals (EDCs) with a carcinogenic potential, OECD TG 451-3 may be more sensitive. Consider potency of effects for existing results and whether E,A,T,S mechanism is credible for reproductive/developmental effects or whether there may be non-endocrine mechanisms. Consider route of exposures for effects data and possible implications of absorption, distribution, metabolism and excretion (ADME) characteristics of the chemical. Endocrine activity possible in lower vertebrates. Consider performing a Fish Sexual Development Test (FSDT), Larval Amphibian Growth and Development Assay (LAGDA) or Medaka Extended One-Generation Reproduction Test (MEOGRT).
B	+	+	–	1) Evidence of adverse effects on endocrine/apical endpoints via E,A,T,S mechanism. 2) Evidence of effects on hormonal endpoints via E,A,T,S mechanism. Apical endpoints may be less sensitive or unaffected. 3) Evidence of adverse effects via E,A,T,S mechanism. Indicators of hormonal activity may be less sensitive or unaffected.	Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	May be sufficient information to conclude evidence of concern for reproductive toxicity via endocrine disruption mechanism. Note that the EOGRTS provides the most information on endocrine disruption; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive. Consider potency of effects for existing results and whether E,A,T,S mechanism is credible for reproductive/developmental effects or whether there may be non-endocrine mechanisms. Consider route of exposures for effects data and possible implications of ADME characteristics of the chemical. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT.

Scenarios	Result of OECD TG 421/422	Existing results		Possible conclusions: 1) Indicators of hormonal activity and apical endpoints positive 2) Indicators of hormonal activity positive and apical endpoints negative 3) Indicators of hormonal activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
C	+	+	Eq/0	1) Evidence of adverse effects on endocrine/apical endpoints via E,A,T,S mechanism. 2) Evidence of effects on hormonal endpoints via E,A,T,S mechanism. Apical endpoints may be less sensitive or unaffected. 3) Evidence of adverse effects via E,A,T,S mechanism. Indicators of hormonal activity may be less sensitive or unaffected.	Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	Note that the EOGRTS provides the most information on endocrine disruption; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive. Consider potency of effects for existing results and whether E,A,T,S mechanism is credible for reproductive/developmental effects or whether there may be non-endocrine mechanisms. Consider route of exposures for effects data and possible implications of ADME characteristics of the chemical. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT. Equivocal results may indicate chemical has multiple modes of action (MOA).
D	+	–	+	1) Evidence of adverse effects on endocrine/apical endpoints but not via E,A,T,S mechanism or requires metabolic activation for activity. 2) Evidence of effects on hormonal endpoints, possibly requires metabolic activation for activity. Apical endpoints may be less sensitive or unaffected. 3) Evidence of adverse effects on apical endpoints but not via E,A,T,S mechanism or requires metabolic activation for activity. Indicators of hormonal activity may be less sensitive or unaffected.	Perform assay from Level 5 (e.g. EOGRTS or two-generation assay). To further discern mechanism, could perform *in vitro* estrogen receptor (ER-), androgen receptor (AR-), thyroid hormone receptor (TR), steroidogenesis (S) assays with added metabolising system.	May be sufficient information to conclude evidence of concern for reproductive toxicity via endocrine disruption mechanism. Note that the EOGRTS provides the most information on endocrine disruption; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive. Consider potency of effects for existing results and whether endocrine disruption mechanism is credible for reproductive/developmental effects or whether there may be non-endocrine mechanisms. Consider route of exposures for effects data and possible implications of ADME characteristics of the chemical. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT.
E	+	–	–	1) Evidence of adverse effects on apical endpoints but not via E,A,T,S mechanism or requires metabolic activation for activity. 2) Evidence of effects on hormonal endpoints, possibly requires metabolic activation for activity. Apical endpoints may be less sensitive or unaffected. 3) Evidence of adverse effects on apical endpoints but not via E,A,T,S mechanism or requires metabolic activation for activity. Indicators of hormonal activity may be less sensitive or unaffected.	Perform assay from Level 5 (e.g. EOGRTS or two-generation assay). To further discern mechanism, could perform *in vitro* ER, AR, TR, S assays with added metabolising system.	May be sufficient information to conclude evidence of concern for reproductive toxicity via endocrine disruption mechanism. Note that the EOGRTS provides the most information on endocrine disruption; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive. Consider potency of effects for existing results and whether endocrine disruption mechanism is credible for reproductive/developmental effects or whether there may be non-endocrine mechanisms. Consider route of exposures for effects data and possible implications of ADME characteristics of the chemical.

Scenarios	Result of OECD TG 421/422	Existing results		Possible conclusions: 1) Indicators of hormonal activity and apical endpoints positive 2) Indicators of hormonal activity positive and apical endpoints negative 3) Indicators of hormonal activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
F	+	–	Eq/0	1) Evidence of adverse effects on apical endpoints via non-E,A,T,S/non-endocrine mechanism or requires metabolic activation for activity. 2) Evidence of effects on hormonal endpoints, possibly requires metabolic activation for activity. Apical endpoints may be less sensitive or unaffected. 3) Evidence of adverse effects on apical endpoints, via non-E,A,T,S/non-endocrine disruption mechanism or requires metabolic activation for activity. Indicators of hormonal activity may be less sensitive or unaffected.	Perform assay from Level 5 (e.g. EOGRTS or two-generation assay). To further discern mechanism, could perform *in vitro* ER, AR, TR, S assays with added metabolising system.	May be sufficient information to conclude evidence of concern for reproductive toxicity via endocrine disruption mechanism. Note that the EOGRTS provides the most information on endocrine disruption; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive. Consider potency of effects for existing results and whether endocrine disruption mechanism is credible for reproductive/developmental effects or whether there may be non-endocrine mechanisms. Consider route of exposures for effects data and possible implications of ADME characteristics of the chemical. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT. Equivocal results may indicate chemical has multiple MOA.
G	+	Eq/0	+	1) Evidence of adverse effects on apical endpoints, may act via E,A,T,S mechanism and may require metabolic activation for activity. 2) Evidence of effects on hormonal endpoints, may act via E,A,T,S mechanism and may require metabolic activation for activity. Apical endpoints may be less sensitive or unaffected. 3) Evidence of adverse effects on apical endpoints, may act via E,A,T,S mechanism and may require metabolic activation for activity. Indicators of hormonal activity may be less sensitive or unaffected.	Perform assay from Level 5 (e.g. EOGRTS or two-generation assay). To further discern mechanism, could perform *in vitro* ER, AR, TR, S assays with added metabolising system.	May be sufficient information to conclude evidence of concern for reproductive toxicity via endocrine disruption mechanism. Note that the EOGRTS provides the most information on endocrine disruption; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive. Consider potency of effects for existing results and whether endocrine disruption mechanism is credible for reproductive/developmental effects or whether there may be non-endocrine mechanisms. Consider route of exposures for effects data and possible implications of ADME characteristics of the chemical. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT. Equivocal results may indicate chemical has multiple MOA.

Scenarios	Result of OECD TG 421/422	Existing results		Possible conclusions: 1) Indicators of hormonal activity and apical endpoints positive 2) Indicators of hormonal activity positive and apical endpoints negative 3) Indicators of hormonal activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
H	+	Eq/0	–	1) Evidence of adverse effects on apical endpoints via non-E,A,T,S/non-endocrine disruption mechanism or requires metabolic activation for activity. 2) Evidence of effects on hormonal endpoints, possibly requires metabolic activation for activity. 3) Evidence of adverse effects on apical endpoints via non-E,A,T,S/non-endocrine disruption mechanism or requires metabolic activation for activity. Indicators of hormonal activity may be less sensitive or unaffected.	Perform assay from Level 5 (e.g. EOGRTS or two-generation assay). To further discern mechanism, could perform *in vitro* ER, AR, TR, S assays with added metabolising system.	May be sufficient information to conclude evidence of concern for reproductive toxicity via endocrine disruption mechanism. Note that the EOGRTS provides the most information on endocrine disruption; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive. Consider potency of effects for existing results and whether endocrine disruption mechanism is credible for reproductive/developmental effects or whether there may be non-endocrine mechanisms. Consider route of exposures for effects data and possible implications of ADME characteristics of the chemical. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT. Equivocal results may indicate chemical has multiple MOA.
I	+	Eq/0	Eq/0	1) Evidence of adverse effects on apical endpoints via unknown mechanism. 2) Evidence of effects on hormonal endpoints via unknown mechanism. Apical endpoints may be less sensitive or unaffected. 3) Evidence of adverse effects on apical endpoints via unknown mechanism. Hormonal endpoints may be less sensitive or unaffected.	Perform assay from Level 5 (e.g. EOGRTS or two-generation assay). To further discern mechanism could perform *in vitro* ER, AR, TR, S assays with added metabolising system.	May be sufficient information to conclude evidence of concern for reproductive toxicity via endocrine disruption mechanism. Note that the EOGRTS provides the most information on endocrine disruption; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive. Consider existing results and whether endocrine disruption mechanism is credible for reproductive/developmental effects or whether there may be non-endocrine mechanisms. Consider route of exposures for effects data and possible implications of ADME characteristics of the chemical. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT. Equivocal results may indicate chemical has multiple MOA.
J	–	+	+	No evidence of adverse effects in OECD TG 421/422. Effects seen in existing (lower level) studies do not lead to adverse outcome in Level 5 assay. Metabolism or potency explains the difference from existing *in vitro* and *in vivo* data.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system OR Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	Consider route of exposures and possible implications for ADME characteristics of the chemical. Effects seen in existing studies may be in a more sensitive life stage. Further mechanistic studies with metabolism may help determine MOA.

Scenarios	Result of OECD TG 421/422	Existing results		Possible conclusions: 1) Indicators of hormonal activity and apical endpoints positive 2) Indicators of hormonal activity positive and apical endpoints negative 3) Indicators of hormonal activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
K	–	+	–	No evidence of adverse effects in OECD TG 421/422. Metabolism or potency explains *in vitro*/*in vivo* differences.	To further discern mechanism, could perform *in vitro* ER, AR, TR, S assays with added metabolising system.	There may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). Metabolic deactivation of chemical may occur *in vivo* so that possible *in vitro* activity is not realised. Further mechanistic studies with metabolism may help determine MOA. Equivocal results may indicate chemical has multiple MOA.
L	–	+	Eq/0	No evidence of adverse effects in OECD TG 421/422. Metabolism or potency explains *in vitro*/*in vivo* differences. Effects seen in existing (lower level) studies do not lead to adverse outcome in this assay.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system OR Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	Metabolic deactivation of chemical may occur *in vivo* so that possible *in vitro* activity is not realised. Further mechanistic studies with metabolism may help determine MOA. Consider route of exposures and possible implications for ADME characteristics of the chemical. Equivocal results may indicate chemical has multiple MOA.
M	–	–	+	No evidence of adverse effects in OECD TG 421/422. Effects seen in existing (lower level) studies do not lead to adverse outcome in this assay.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system OR Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	If existing data are from an adequate Level 5 assay, question why there are differences. Effects seen in existing studies may be in a more sensitive life stage. Consider route of exposures and possible implications of ADME characteristics of the chemical.
N	–	–	–	No evidence of adverse effects in OECD TG 421/422.	Consider existing data, there may be no need for further testing.	There may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive).
O	–	–	Eq/0	No evidence of adverse effects in OECD TG 421/422. No evidence for (anti)-E,A,T,S activity *in vitro*.	Consider existing data, there may be no need for further testing. To further discern mechanism, could perform *in vitro* ER, AR, TR, S assays with added metabolising system.	There may be sufficient information to conclude absence of concern for endocrine disruption. Further mechanistic studies may strengthen weight of evidence. Consider route of exposures and possible implications for ADME characteristics of the chemical. Check data on chemical analogues.

Scenarios	Result of OECD TG 421/422	Existing results		Possible conclusions: 1) Indicators of hormonal activity and apical endpoints positive 2) Indicators of hormonal activity positive and apical endpoints negative 3) Indicators of hormonal activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
P	–	Eq/0	+	No evidence of adverse effects in OECD TG 421/422. Effects seen in existing (lower level) studies do not lead to adverse outcome in this assay. Effects seen in existing studies are via unknown mechanism.	Perform *in vitro* ER, AR,TR, S assays.	Further mechanistic studies may strengthen weight of evidence. Consider route of exposures and possible implications for ADME characteristics of the chemical. Effects seen in existing studies may be in a more sensitive life stage. Check data on chemical analogues. Equivocal results may indicate chemical has multiple MOA.
Q	–	Eq/0	–	No evidence of adverse effects in OECD TG 421/422.	To further discern mechanism, could perform *in vitro* ER, AR, TR, S assays with added metabolising system.	There may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). Further mechanistic studies may strengthen weight of evidence. Check data on chemical analogues.
R	–	Eq/0	Eq/0	No evidence of adverse effects in OECD TG 421/422.	To further discern mechanism, could perform *in vitro* ER, AR, TR, S assays with added metabolising system.	Further mechanistic studies may strengthen weight of evidence. Check data on chemical analogues. Equivocal results may indicate chemical has multiple MOA.

C.3.7. Prenatal Developmental Toxicity Study (OECD TG 414)

Status: Assay validated by the OECD.

793. Modalities detected: (anti)estrogen, (anti)androgen, thyroid.

794. Endpoints: Number of implantations and corpora lutea; post-implantation loss, litter size, sex ratio, litter/fetal weight; fetal anogenital distance; fetal external, soft tissue and skeletal changes. Observations of external fetal sex (determined by gross examination) and internal (gonadal sex) in all fetuses. Indications of incomplete testicular descent/cryptorchidism in male fetuses.

Hormones: T4, T3 and TSH in dams. Other hormones "if relevant".

Background to the assay

795. The OECD Prenatal Developmental Toxicity Study is an apical assay designed to provide general information concerning the effects of prenatal exposure to a chemical on the pregnant test animal and on the developing organism. This may include assessment of maternal effects as well as death, structural abnormalities or altered growth in the foetus. The primary purpose of this study is to provide data on adverse effects related to development. The current version of the guideline was adopted in January 2001. Previous versions of this test guideline (TG) had a less extensive exposure period and fewer endpoints. The study was not designed to detect endocrine active substances (EASs), but has some endpoints relevant for the assessment of possible endocrine disruption and is currently being updated to include more. Following a feasibility study (OECD, 2015), this assay was updated in July 2018 to include some endocrine-relevant endpoints as the exposure periods cover some of the sensitive periods during development (prenatal period). It should be noted that if an assay was conducted before 2018, it is unlikely to include these extra endpoints.

796. Test substance is administered from implantation throughout pregnancy. The rat is the preferred rodent species and the rabbit the preferred non-rodent species. Route of administration is typically via oral gavage, but other routes may be used. The exposure of the fetus (which may be a sensitive life stage for endocrine disruption effects) means that some endocrine effects on development may be detected in this assay. Anogenital distance (AGD), appearance of external genitalia and sex ratio are examples of apical endpoints that may be affected by via estrogen- or androgen-mediated activity. T4 and TSH may be affected by disturbance of the thyroid hormonal system. Most endpoints are apical and therefore it may be difficult to discern mode of action (MOA) from this test alone. Information on potential endocrine mode of action may need to be obtained from *in vitro* estrogen/androgen/thyroid/steroidogenesis (E,A,T,S) assays or *in vivo* lower tier tests such as a Uterotrophic Bioassay (UT) and a Hershberger Bioassay (H), for example. The small number of endocrine-sensitive endpoints means that an absence of effect in this assay alone cannot lead to a conclusion that a substance does not have endocrine disrupting effects. Additional data from more comprehensive assays may be required.

797. AGD is included in OECD TG 414 as a sensitive endpoint of endocrine effects; however, its utility as an apical endpoint or as a biological indicator of endocrine action may require further experience in its use. Reduced anogenital distance in male offspring is a hallmark of anti-androgenicity. Nevertheless, alteration of AGD can occur via other MOA (e.g. Miyagawa et al. [2011]; Seifert et al. [2009]). However current OECD guidance on AGD can be found in OECD GD 43 and GD 151 and it is clear that it should be considered as an apical endpoint. OECD GD 43 (OECD, 2008b) states, "A statistically significant change in [anogenital distance] that cannot be explained by the size of the animal indicates effects of the exposure and should be used for setting the [no observed adverse effect level]".

Thyroid hormone analysis and interpretation in OECD TG 414

798. The revised TG 414 requires the determination of serum thyroid hormones in dams. Detection/measurement of thyroid hormones (T4, TSH) in some rodent studies can be challenging. In order to provide assurance that thyroid hormone measurements are reliable, laboratories should be able to demonstrate that they are proficient with the assay. Only such a demonstration would enable regulatory authorities to interpret and use the data. A recent workshop on "Practicability of Hormonal Measurements" was organised by the BfR (Germany) and the finding from this workshop will be published (Kucheryavenko et al., 2018). The OECD Expert Group on Reproductive and Developmental Toxicity recommends that to demonstrate proficiency for thyroid hormones measurement, a laboratory should be able to show results from a separate study using a positive control substance. Laboratories may also submit their calibration curves, standard curves, as well as data on the levels of quantification and detection. This group is also establishing a historical control database with thyroid toxicant positive controls.

When/why the assay may be used

799. This assay forms part of the package of studies required for registration of pesticides in many jurisdictions. It forms part of the standard information requirements in certain chemical legislations (e.g. REACH for chemicals which are manufactured or imported in quantities of 1 000 tonnes or more). It may also be carried out for high production volume chemicals of high concern. It is likely to have at least three dose levels and therefore may be used for hazard identification/characterisation.

800. In order to provide information relevant for assessing whether or not a chemical may fulfil the WHO/IPCS (2002) definition of an endocrine disruptor (ED), the study design has to be sufficiently robust to demonstrate the presence or absence of effects. In the dose selection, the investigator should also consider and ensure that data generated are adequate to fulfil the regulatory requirement across OECD countries as appropriate (e.g. hazard and risk assessment and labelling, ED assessment, etc.). The top dose or concentration should be sufficiently high to give clear systemic (i.e. non endocrine-specific) toxicity in order to ensure that a wide range of exposures (high to low) is tested. However, endocrine effects observed solely in the presence of clear systemic toxicity should be interpreted with caution and may be disregarded when sufficiently justified to be caused by secondary effects which are unlikely to be due to endocrine activity. The reason for this advice is a concern that some endocrine active substance (EAS) sensitive assays are being run at doses/concentrations of EASs that are too low to trigger direct impacts on the endocrine system. This guidance document is not the place to address this issue directly, but it should be considered when EAS-sensitive TGs are revised in the future. In addition, the number

and spacing of dose/concentration levels should also be adequate to fulfil the objectives of the study (e.g. to demonstrate dose response relationships if this is required).

Introduction to the table of scenarios

801. Table C.3.7 gives guidance on a further step to take in the event of a positive (+) or negative (-) result and in the presence of positive (+), negative (-) or equivocal/absent (Eq/0) existing results. "Existing results" are subdivided into "mechanism" and "effects" data (third and fourth columns). The table is divided horizontally into a series of scenarios that represent all the combinations of these events.

802. The results of OECD TG 414 are given in the second column. This assay is not a screening test where a yes/no (qualitative) answer is obtained. Results for the endpoints would be considered both individually and as a whole. It is not possible to provide guidance on all endpoints individually, therefore the endpoints have been pragmatically divided into "apical" and "indicators of hormonal activity". The terminology used has been chosen to be consistent between both the non-mammalian wildlife and mammalian tests. Both groups of endpoints have similar biological importance, although the "indicators of hormonal activity" in the mammalian assays are serum hormones and are generally, but not always, more variable than "apical endpoints".

803. For this guideline "apical" endpoints are developmental parameters (including anogenital distance, genital abnormalities and sex ratio). "Indicators of hormonal activity" are hormones (including T4, TSH).

804. Three possible outcomes for a positive result are therefore envisaged in Table C.3.7:

1. indicators of hormonal activity and apical endpoints positive
2. indicators of hormonal activity positive and apical endpoints negative
3. indicators of hormonal activity negative and apical endpoints positive.

805. A positive result for apical endpoints could be biologically significant changes in fetal anogenital distance, accompanied changes in sex ratio. A positive result for indicators of hormonal activity could be biologically significant changes in hormone profiles. A positive result for indicators of hormonal activity alone should be considered with caution, although it is possible that these endpoints may have detected weak effects that were not detected by the apical endpoints.

806. A negative result for OECD TG 414 is taken to be the absence of biologically significant changes in all of the endocrine endpoints measured in this study. Studies conducted to current standards are considered to be more predictive for absence of reproductive and developmental effects.

807. In the absence of other pertinent lines of evidence, negative results in this test alone cannot be taken as evidence that the substance is not an ED. Further studies may be required as confirmation.

808. Equivocal results for the guideline are not considered in Table C.3.7, partly for brevity but also because equivocal results are by nature uncertain. A decision must eventually be reached about whether the endocrine endpoints tend to be positive or negative or whether the result must be put to one side and the test repeated or supplemented by a different test.

Existing data to be considered

809. Existing "mechanism" *in vitro* data are assumed to be available from estrogen receptor (ER-), androgen receptor (AR-) and steroidogenesis-based assays (Level 2). Assays may also be available for interference with thyroid modalities. In practice, it is possible that data from all of these assays may not be available, so judgement will need to be used to decide which assays to perform. Although the current *in vitro* test guidelines do not incorporate metabolic activation, published information on use of metabolic activation systems is available in Jacobs et al. (2008; 2013) and OECD (2008a). These methods, however, have not yet been validated.

810. Existing "effects" data refer to *in vivo* effects that may come from lower level assays, e.g. UT or H Assays (Level 3); Peripubertal (PP) Assays or OECD TG 407 assays (Level 4), or there may be longer term studies (e.g. in the case of pesticide registration packages where 90-day and carcinogenicity studies may be available). Data may also be available on effects in mammalian and non-mammalian wildlife species, although caution should be used when extrapolating between taxa. A chemical causing endocrine effects in non-mammalian environmental species (fish, for example) may also have endocrine effects in mammals, but the physiological consequences of the effects are likely to be different.

811. When considering the results of OECD TG 414, all available data should be used in order to reach a conclusion and a weight of evidence approach taken. This may include high throughput screening data, read-across data from structural analogues and quantitative structure activity relationship (QSAR). Several QSAR models for ER and AR binding/activation are now available (see Sections B.1.1.1 and B.1.1.2).

Scenarios: Positive and negative results combined with existing data

812. The scenarios (A to R) presented in Table C.3.7 represent all the possibilities of positive or negative results in combination with the presence or absence of existing data. The action taken will also depend on the regulatory environment, but the considerations given here are generally science based. Although rats and rabbits are the preferred species for OECD TG 414, the well-conserved nature of the hormonal pathways across taxa should be an indication that results in this assay may be relevant to other vertebrate species. Effects in laboratory mammal tests are also highly relevant for environmental mammalian species. Wherever possible, the recommended "next step which could be taken" avoids unnecessary animal testing. However, sometimes conducting an animal test will be indicated and then the relevance of species, strain, exposure route and species-specific metabolism should always be considered. The sensitivity and physiological function of the hormone under investigation in the test species should also be considered. In general, lower level tests should be conducted before higher level tests in order to avoid unnecessary animal usage, unless it is apparent that a Level 5 test will be required anyway or will be needed to establish the evidence to conclude on ED properties. Information on some endocrine-related tumours may be detected more comprehensively in carcinogenicity studies (OECD TG 451/453) (Level 4); for example, detection of certain types of thyroid tumors in the absence of reproductive or developmental effects, as well as substances causing tumors in other endocrine-sensitive tissues. At Level 5, the Extended One-Generation Reproduction Toxicity Study (EOGRTS – OECD TG 443) is the most sensitive reproduction assay for detecting endocrine disruption because it includes evaluation of a number of endocrine endpoints not included in the two-generation study (OECD TG 416) adopted in 2001. Further considerations specific to each scenario are given in the table.

813. Scenarios A to C represent positive results in OECD TG 414 in the presence of positive *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. Each positive result scenario is divided into the three possible outcomes given above. A positive result in the *in vitro* assays in combination with a positive OECD TG 414 assay is evidence of adverse effects on development and/or endocrine endpoints via E,A,T,S mechanisms. Effects on only apical endpoints or only indicators of hormonal activity may assist with interpretation. In the absence of robust upper-level data, the next step may be to conduct an upper-level test. In the presence of robust *in vivo* data, there may be sufficient evidence to conclude concern for endocrine disruption and therefore no need for further testing. Positive results in the OECD TG 414 assay may also indicate the potential for endocrine mediated effects in lower vertebrates. These could be followed up with partial life cycle tests such as the Fish Sexual Development Test (FSDT), Larval Amphibian Growth and Development Assay (LAGDA) or Medaka Extended One-Generation Reproduction Test (MEOGRT) if the evidence is strong enough. *In vivo* assays/tests with negative results should be interpreted with caution as they may either indicate that the tests used do not have sufficient power to detect weak effects or, alternatively, that the effects do not present a concern for endocrine disruption. The possibility of other (non-E,A,T,S) mechanisms should also not be overlooked (e.g. involving other receptors or endocrine axes).

814. Scenarios D to F represent positive results in OECD TG 414 in the presence of negative *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. Each positive result scenario is divided into the three possible outcomes given above. A positive result in OECD TG 414 is evidence of adverse effects on development. Effects on only apical endpoints or only indicators of hormonal activity may assist with interpretation. Negative results in the *in vitro* assays should be viewed with caution in case a metabolite is responsible for the positive OECD TG 414 study. If the metabolic profile of the test substance is not known, then performing the *in vitro* assays with addition of a metabolising system may help to understand mechanism. The choice of tests will depend on the available *in vivo* effects data. Positive results in the OECD TG 414 assay may also indicate the potential for endocrine mediated effects in lower vertebrates. As in Scenarios A to C, *in vivo* assays/tests with negative results should be interpreted with caution as they may either indicate that the tests used do not have sufficient power to detect weak effects or, alternatively, that the effects do not present a concern for endocrine disruption.

815. Scenarios G to I represent positive results in OECD TG 414 in the presence of various combinations of missing or equivocal data. The next step to take in these eventualities will depend on the nature of the other available data and the jurisdiction in which it is being used. In some cases, equivocal data may be viewed as positive whilst in others it may or may not contribute to the weight of evidence. The interpretation may also depend on the MOA in question and why the data are considered equivocal. In all three scenarios, the recommended first step is to obtain reliable mechanistic (*in vitro*) data rather than proceed further with *in vivo* testing. Positive results in the OECD TG 414 assay may also indicate the potential for endocrine mediated effects in lower vertebrates. Equivocal and missing data are alternative scenarios and two possibilities for the next step are given in most cases, but the nature of equivocal data means that decisions need to be taken on a case-by-case basis. In all cases, the role of metabolism, route of exposure and data from structural analogues should be considered before deciding on the next step.

816. Scenarios J to L represent negative results in OECD TG 414 in the presence of positive *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. A negative result is taken to be negative findings for both indicators of hormonal activity and apical endpoints (unlike the situation with positive outcomes), therefore there is only

one possible negative outcome. In all scenarios, the small number of endocrine-sensitive endpoints in OECD TG 414 means that an absence of effect in this assay alone cannot lead to a conclusion that a substance does not have endocrine disrupting effects. Additional data from more comprehensive assays are required. All three scenarios could fit a chemical that is positive in *in vitro* assays but is metabolised to a non-active metabolite, leading to negative results in OECD TG 414. This possibility may be investigated to help understand mechanism. Endocrine active potency may also explain differences between *in vitro* and *in vivo* results (e.g. a chemical with weak endocrine activity may give a positive result *in vitro* but may be negative *in vivo*). Positive *in vivo* effects data may involve E,A,T,S or non-E,A,T,S mechanisms (e.g. involving other receptors or endocrine axes), more sensitive endpoints, greater statistical power, but knowledge of absorption, distribution, metabolism and excretion (ADME) may help to explain differences from the OECD TG 414 data.

817. Scenarios M to O represent negative results in OECD TG 414 in the presence of negative *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. In all scenarios, the small number of endocrine-sensitive endpoints in OECD TG 414 means that an absence of effect in this assay alone cannot lead to a conclusion that a substance does not have endocrine disrupting effects. Positive *in vivo* effects data may involve E,A,T,S or non-E,A,T,S mechanisms (e.g. involving other receptors or endocrine axes), but knowledge of ADME may help to explain differences from the OECD TG 414 data.

818. Scenarios P to R represent negative results in OECD TG 414 in the presence of various combinations of missing or equivocal data. As with the positive result scenarios above (see Paragraph 787), the next step to take in these eventualities will have to be decided on a case-by-case basis. In all cases, the role of metabolism, route of exposure and data from structural analogues should be considered before deciding on the next step.

819. In all scenarios (A to R), the next step to take to strengthen weight of evidence will depend on the existing information. Table C.3.6 is meant to provide a succinct guide and may not cover all circumstances or possibilities. The scenarios may also suggest that chemicals have simple or single MOA, when in practice they may have multiple endocrine and non-endocrine MOA. In some cases, for example, two opposite modes of simultaneous action (e.g. estrogenic and anti-estrogenic) could, depending on dose, lead to a minimisation or abolition of effects, while in others two different MOA (e.g. estrogenic and anti-androgenic) could potentially reinforce effects. Endocrine pathways interact, mixed effects are common and there are many pathways that cannot be distinguished with currently available TGs. If multiple MOA are suspected, either from the existing results or based on QSAR/read-across/integrated approaches, this should be investigated further if needed for regulatory decision making.

References

Jacobs, M. et al. (2013), "*In vitro* metabolism and bioavailability tests for endocrine active substances: What is needed next for regulatory purposes?", *ALTEX – Alternatives to Animal Experimentation*, Vol. 30/3, pp. 331-351.

Jacobs, M.N. et al. (2008), "The use of metabolising systems for *in vitro* testing of endocrine disrupters", *Current Drug Metabolism*, Vol. 9/8, pp. 796-826.

Kucheryavenko, O. et al. (2018), "Report from the BfR Expert Hearing on Practicability of Hormonal Measurements", *Archives of Toxicology* (in prep.).

Miyagawa, S. et al. (2011), "The role of sonic hedgehog-Gli2 pathway in the masculinization of external genitalia", *Endocrinology*, Vol. 152/7, pp. 2894-2903, https://doi.org/10.1210/en.2011-0263.

OECD (2015), "Feasibility study for minor enhancements of TG 421/422 with ED relevant endpoints", Series on Testing and Assessment, No. 217, OECD, Paris, www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2015)24&doclanguage=en.

OECD (2008a), *Detailed Review Paper on the Use of Metabolising Systems for* In Vitro *Testing of Endocrine Disrupters*, OECD Series on Testing and Assessment, No. 97, OECD Publishing, Paris, https://doi.org/10.1787/9789264085497-en.

OECD (2008b), "Guidance document on mammalian reproductive toxicity testing and assessment", OECD Series on Testing and Assessment, No. 43, OECD, Paris, www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/JM/MONO(2008)16&doclanguage=en.

Seifert, A.W. et al. (2009), "Multiphasic and tissue-specific roles of sonic hedgehog in cloacal septation and external genitalia development", *Development*, Vol. 136/23, pp. 3949-3957, http://dx.doi.org/10.1242/dev.042291.

WHO/IPCS (2002), "Global assessment of the state-of-the-science of endocrine disrupters", Damstra, T. et al. (eds.) WHO/PCS/EDC/02.2, World Health Organization, Geneva, www.who.int/ipcs/publications/new_issues/endocrine_disruptors/en.

Table C.3.7. **Prenatal Developmental Toxicity Study (OECD TG 414): Guidance for scenarios of combinations of results with existing data**

This table represents possible conclusions to be drawn from assay data, and a next step which could be taken if further evidence is required about possible endocrine disrupting properties and/or effects. The guidance offered is not meant to be prescriptive, but provides science-based considerations. It encourages the use of all available data and expert judgement in a weight of evidence approach. Regional and national interpretation of results and "next steps" may vary.

The conclusions are grouped into a series of scenarios (A-R), each scenario representing a different combination of assay results, existing *in vitro* data and existing *in vivo* data. The symbol "+" indicates that the data in question represent a positive result, «-" indicates a negative result, and "Eq/0" indicates that the data are either equivocal or are not available.

Existing results: * "Mechanism (*in vitro* mechanistic data)" assumes that mechanistic data are available from estrogen receptor (ER-), androgen receptor (AR-) and steroidogenesis-based assays (Level 2). Thyroid hormone receptor (TR) and other assays concerning mechanisms of thyroid disruption may be available, but they are not in common use. In practice, data from all assays may not be available and therefore this must be taken into account when deciding on the "next step". Quantitative structure activity relationship (QSAR) predictions of estrogen and androgen binding/activation may be made for some substances.

Existing results: ** "Effects (*in vivo* effects of concern)" assumes effects have been observed in other *in vivo* screens/tests which give rise to concern that the test chemical may be an endocrine disrupter. These may be other repeated dose toxicity tests, the Uterotrophic Bioassay (UT) and Hershberger Bioassay (H), Peripubertal (PP) Assays or read-across from chemical analogues.

*** *Note:* three possible outcomes for a positive result are given:

1. indicators of hormonal activity and apical endpoints positive
2. indicators of hormonal activity positive and apical endpoints negative
3. indicators of hormonal activity negative and apical endpoints positive.

"Apical endpoints" are developmental parameters (number of implantations and corpora lutea; post implantation loss, litter size, sex ratio, litter/fetal weight; fetal anogenital distance; fetal external, soft tissue and skeletal changes).

"Indicators of hormonal activity" are hormones (T4, TSH, optional testosterone).

Scenarios	Result of OECD TG 414 (dev tox)	Existing results		Possible conclusions: 1) Indicators of hormonal activity and apical endpoints positive 2) Indicators of hormonal activity positive and apical endpoints negative 3) Indicators of hormonal activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
A	+ ***	+	+	1) Evidence of adverse effects on endocrine/ apical endpoints via estrogen/androgen/ thyroid/steroidogenesis (E,A,T,S) mechanism. 2) Evidence of effects on hormonal endpoints via E,A,T,S mechanism. Apical endpoints may be less sensitive or unaffected. 3) Evidence of adverse effects via E,A,T,S mechanism. Indicators of hormonal activity may be less sensitive or unaffected.	Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	May be sufficient information to conclude evidence of concern for deveopmental toxicity via endocrine disruption mechanism. Note that the EOGRTS rovides the most information on endocrine disruption. Effects on apical endpoints may indicate E,A,T,S modalities or other mechanisms. Consider potency of effects for existing results and whether E,A,T,S mechanism is credible for reproductive/developmental effects or whether there may be non-endocrine mechanisms. Consider route of exposures for effects data and possible implications of absorption, distribution, metabolism and excretion (ADME) characteristics of the chemical. Endocrine activity possible in lower vertebrates. Consider performing a Fish Sexual Development Test (FSDT), Larval Amphibian Growth and Development Assay (LAGDA) or Medaka Extended One-Generation Reproduction Test (MEOGRT).
B	+	+	–	1) Evidence of adverse effects on endocrine/ apical endpoints via E,A,T,S mechanism. 2) Evidence of effects on hormonal endpoints via E,A,T,S mechanism. Apical endpoints may be less sensitive or unaffected. 3) Evidence of adverse effects via E,A,T,S mechanism. Indicators of hormonal activity may be less sensitive or unaffected.	Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	Note that the EOGRTS provides the most information on endocrine disruption; however, for endocrine disrupting chemicals (EDCs) with a carcinogenic potential, OECD TG 451-3 may be more sensitive. Effects on apical endpoints may indicate E,A,T,S modalities or other mechanisms. Consider potency of effects for existing results and whether E,A,T,S mechanism is credible for reproductive/developmental effects or whether there may be non-endocrine mechanisms. Consider route of exposures for effects data and possible implications of ADME characteristics of the chemical. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT.

Scenarios	Result of OECD TG 414 (Dev tox)	Existing results		Possible conclusions: 1) Indicators of hormonal activity and apical endpoints positive 2) Indicators of hormonal activity positive and apical endpoints negative 3) Indicators of hormonal activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
C	+	+	Eq/0	1) Evidence of adverse effects on endocrine/ apical endpoints via E,A,T,S mechanism. 2) Evidence of effects on hormonal endpoints via E,A,T,S mechanism. Apical endpoints may be less sensitive or unaffected. 3) Evidence of adverse effects via E,A,T,S mechanism. Indicators of hormonal activity may be less sensitive or unaffected.	Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	Note that the EOGRTS provides the most information on endocrine disruption; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive. Effects on apical endpoints may indicate E,A,T,S modalities or other mechanisms. Consider potency of effects for existing results and whether E,A,T,S mechanism is credible for reproductive/developmental effects or whether there may be non-endocrine mechanisms. Equivocal results may indicate chemical has multiple modes of action (MOA). Consider route of exposures for effects data and possible implications of ADME characteristics of the chemical. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT.
D	+	–	+	1) Evidence of adverse effects on endocrine/apical endpoints but not via E,A,T,S mechanism or requires metabolic activation for activity. 2) Evidence of effects on hormonal endpoints, possibly requires metabolic activation for activity. Apical endpoints may be less sensitive or unaffected. 3) Evidence of adverse effects on apical endpoints but not via E,A,T,S mechanism or requires metabolic activation for activity. Indicators of hormonal activity may be less sensitive or unaffected.	Perform assay from Level 5 (e.g. EOGRTS or two-generation assay). To further discern mechanism, could perform *in vitro* estrogen receptor (ER-), androgen receptor (AR-), thyroid hormone receptor (TR), steroidogenesis (S) assays with added metabolising system.	May be sufficient information to conclude evidence of concern for reproductive toxicity via endocrine disruption mechanism. Note that the EOGRTS provides the most information on endocrine disruption; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive. Effects on apical endpoints may indicate E,A,T,S modalities or other mechanisms. Consider potency of effects for existing results and whether endocrine disruption mechanism is credible for reproductive/developmental effects or whether there may be non-endocrine mechanisms. Consider route of exposures for effects data and possible implications of ADME characteristics of the chemical. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT.

Scenarios	Result of OECD TG 414 (Dev tox)	Existing results		Possible conclusions: 1) Indicators of hormonal activity and apical endpoints positive 2) Indicators of hormonal activity positive and apical endpoints negative 3) Indicators of hormonal activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
E	+	–	–	1) Evidence of adverse effects on apical endpoints but not via E,A,T,S mechanism or requires metabolic activation for activity. 2) Evidence of effects on hormonal endpoints, possibly requires metabolic activation for activity. Apical endpoints may be less sensitive or unaffected. 3) Evidence of adverse effects on apical endpoints but not via E,A,T,S mechanism or requires metabolic activation for activity. Indicators of hormonal activity may be less sensitive or unaffected.	Perform assay from Level 5 (e.g. EOGRTS or two-generation assay). To further discern mechanism, could perform *in vitro* ER, AR, TR, S assays with added metabolising system.	Note that the EOGRTS provides the most information on endocrine disruption; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive. Effects on apical endpoints may indicate E,A,T,S modalities or other mechanisms. Consider potency of effects for existing results and whether endocrine disruption mechanism is credible for reproductive/developmental effects or whether there may be non-endocrine mechanisms. Consider route of exposures for effects data and possible implications of ADME characteristics of the chemical. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT.
F	+	–	Eq/0	1) Evidence of adverse effects on apical endpoints via non-E,A,T,S/non-endocrine mechanism or requires metabolic activation for activity. 2) Evidence of effects on hormonal endpoints, possibly requires metabolic activation for activity. Apical endpoints may be less sensitive or unaffected. 3) Evidence of adverse effects on apical endpoints, via non-E,A,T,S/non-endocrine disruption mechanism or requires metabolic activation for activity. Indicators of hormonal activity may be less sensitive or unaffected.	Perform assay from Level 5 (e.g. EOGRTS or two-generation assay). To further discern mechanism, could perform *in vitro* ER, AR, TR, S assays with added metabolising system.	Note that the EOGRTS provides the most information on endocrine disruption; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive. Effects on apical endpoints may indicate E,A,T,S modalities or other mechanisms. Equivocal results may indicate chemical has multiple MOA. Consider potency of effects for existing results and whether endocrine disruption mechanism is credible for reproductive/developmental effects or whether there may be non-endocrine mechanisms. Consider route of exposures for effects data and possible implications of ADME characteristics of the chemical. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT.

Scenarios	Result of OECD TG 414 (Dev tox)	Existing results		Possible conclusions: 1) Indicators of hormonal activity and apical endpoints positive 2) Indicators of hormonal activity positive and apical endpoints negative 3) Indicators of hormonal activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
G	+	Eq/0	+	1) Evidence of adverse effects on apical endpoints, may act via E,A,T,S mechanism and may require metabolic activation for activity. 2) Evidence of effects on hormonal endpoints, may act via E,A,T,S mechanism and may require metabolic activation for activity. Apical endpoints may be less sensitive or unaffected. 3) Evidence of adverse effects on apical endpoints, may act via E,A,T,S mechanism and may require metabolic activation for activity. Indicators of hormonal activity may be less sensitive or unaffected.	Perform assay from Level 5 (e.g. EOGRTS or two-generation assay). To further discern mechanism, could perform *in vitro* ER, AR, TR, S assays with added metabolising system.	Note that the EOGRTS provides the most information on endocrine disruption; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive. Effects on apical endpoints may indicate E,A,T,S modalities or other mechanisms. Equivocal results may indicate chemical has multiple MOA. Consider potency of effects for existing results and whether endocrine disruption mechanism is credible for reproductive/developmental effects or whether there may be non-endocrine mechanisms. Consider route of exposures for effects data and possible implications of ADME characteristics of the chemical. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT
H	+	Eq/0	–	1) Evidence of adverse effects on apical endpoints via non-E,A,T,S/non-endocrine disruption mechanism or requires metabolic activation for activity. 2) Evidence of effects on hormonal endpoints, possibly requires metabolic activation for activity. 3) Evidence of adverse effects on apical endpoints via non-E,A,T,S/non-endocrine disruption mechanism or requires metabolic activation for activity. Indicators of hormonal activity may be less sensitive or unaffected.	Perform assay from Level 5 (e.g. EOGRTS or two-generation assay). To further discern mechanism, could perform *in vitro* ER, AR, TR, S assays with added metabolising system.	Note that the EOGRTS provides the most information on endocrine disruption; however; for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive. Effects on apical endpoints may indicate E,A,T,S modalities or other mechanisms. Consider potency of effects for existing results and whether endocrine disruption mechanism is credible for reproductive/developmental effects or whether there may be non-endocrine mechanisms. Consider route of exposures for effects data and possible implications of ADME characteristics of the chemical. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT
I	+	Eq/0	Eq/0	1) Evidence of adverse effects on apical endpoints via unknown mechanism. 2) Evidence of effects on hormonal endpoints via unknown mechanism. Apical endpoints may be less sensitive or unaffected. 3) Evidence of adverse effects on apical endpoints via unknown mechanism. Hormonal endpoints may be less sensitive or unaffected.	Perform assay from Level 5 (e.g. EOGRTS or two-generation assay). To further discern mechanism, could perform *in vitro* ER, AR, TR, S assays with added metabolising system.	Consider existing results and whether endocrine disruption mechanism is credible for reproductive/developmental effects or whether there may be non-endocrine mechanisms. Equivocal results may indicate chemical has multiple MOA. Consider route of exposures for effects data and possible implications of ADME characteristics of the chemical. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT.

Scenarios	Result of OECD TG 414 (Dev tox)	Existing results		Possible conclusions: 1) Indicators of hormonal activity and apical endpoints positive 2) Indicators of hormonal activity positive and apical endpoints negative 3) Indicators of hormonal activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
J	–	+	+	No evidence of adverse effects in OECD TG 414. Effects seen in existing (lower level) studies do not lead to adverse outcome. Metabolism or potency explains the difference from existing *in vitro* and *in vivo* data.	Consider supplemental testing, depending on existing data. To further discern mechanism, could perform *in vitro* ER, AR, TR, S assays with added metabolising system.	Note that the EOGRTS provides the most information on endocrine disruption; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive. Consider route of exposures and possible implications for ADME characteristics of the chemical with existing studies. Further mechanistic studies with metabolism may help determine MOA.
K	–	+	–	No evidence of adverse effects in OECD TG 414. Metabolism or potency explains *in vitro*/*in vivo* differences.	Consider supplemental testing, depending on existing data. To further discern mechanism, could perform *in vitro* ER, AR, TR, S assays with added metabolising system.	If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption. Further mechanistic studies with metabolism may help determine MOA.
L	–	+	Eq/0	No evidence of adverse effects in OECD TG 414. Metabolism or potency explains *in vitro*/*in vivo* differences. Effects seen in existing (lower level) studies do not lead to adverse outcome.	Consider supplemental testing, depending on existing data. To further discern mechanism, could perform *in vitro* ER, AR, TR, S assays with added metabolising system.	Note that the EOGRTS provides the most information on endocrine disruption; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive. Consider route of exposures and possible implications for ADME characteristics of the chemical with existing studies. Further mechanistic studies with metabolism may help determine MOA. Equivocal results may indicate chemical has multiple MOA.
M	–	–	+	No evidence of adverse effects in OECD TG 414. Effects seen in existing (lower level) studies do not lead to adverse outcome.	Consider supplemental testing, depending on existing data. To further discern mechanism, could perform *in vitro* ER, AR, TR, S assays with added metabolising system.	If existing data are from adequate *in vivo* studies such as 28-day, 90-day, chronic/carcinogenicity studies, question why there are differences. Note that the EOGRTS provides the most information on endocrine disruption; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive. Consider route of exposures and possible implications for ADME characteristics of the chemical with existing studies. Further mechanistic studies with metabolism may help determine MOA.
N	–	–	–	No evidence of adverse effects in OECD TG 414.	Consider existing data, there may be no need for further testing.	If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption. Note that the EOGRTS provides the most information on endocrine disruption; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive.

Scenarios	Result of OECD TG 414 (Dev tox)	Existing results		Possible conclusions: 1) Indicators of hormonal activity and apical endpoints positive 2) Indicators of hormonal activity positive and apical endpoints negative 3) Indicators of hormonal activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
O	–	–	Eq/0	No evidence of adverse effects in OECD TG 414. No evidence for (anti)-E,A,T,S activity *in vitro*.	Consider existing data, there may be no need for further testing. To further discern mechanism, could perform *in vitro* ER, AR, TR, S assays with added metabolising system.	Note that the EOGRTS provides the most information on endocrine disruption; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive. Further mechanistic studies with metabolism may help determine MOA. Consider route of exposures and possible implications for ADME characteristics of the chemical with existing studies. Check data on chemical analogues.
P	–	Eq/0	+	No evidence of adverse effects in OECD TG 414. Effects seen in existing (lower level) studies do not lead to adverse outcome. Effects seen in existing studies are via unknown mechanism.	Consider supplemental testing, depending on existing data. To further discern mechanism, could perform *in vitro* ER, AR, TR, S assays with added metabolising system.	If existing data are from adequate *in vivo* studies such as 28-day, 90-day, chronic/carcinogenicity studies, question why there are differences. Note that the EOGRTS provides the most information on endocrine disruption; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive. Consider route of exposures and possible implications for ADME characteristics of the chemical with existing studies. Check data on chemical analogues. Equivocal results may indicate chemical has multiple MOA.
Q	–	Eq/0	–	No evidence of adverse effects in OECD TG 414.	To further discern mechanism, could perform in vitro ER, AR, TR, S assays with added metabolising system.	There may be sufficient information to conclude absence of concern for endocrine disruption. Check data on chemical analogues.
R	–	Eq/0	Eq/0	No evidence of adverse effects in OECD TG 414.	To further discern mechanism, could perform *in vitro* ER, AR, TR, S assays with added metabolising system.	Note that the EOGRTS provides the most information on endocrine disruption; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive. Further mechanistic studies may strengthen weight of evidence. Check data on chemical analogues. Equivocal results may indicate chemical has multiple MOA.

C.3.8. Developmental Neurotoxicity Study (OECD TG 426)

Status: Assay validated by the OECD.

820. Modalities detected: (Anti)-eEstrogen, (anti)androgen, thyroid, steroidogenesis.

Endpoints: Gestation length, litter size, sex ratio (F1, F2), litter/pup weight, pup survival index, sexual maturation (age at vaginal opening and preputial separation).

In offspring: motor activity (including habituation), motor and sensory function, learning and memory.

Brain weight and histopathological examination. Morphometric (quantitative) evaluation of the brain.

Other (non-neurological) tissues may be taken at post-mortem on a case-by-case basis.

Background to the assay

821. OECD TG 426 determines the potential for developmental neurotoxicity of chemicals. The assay provides data, including dose-response characterisations, on the potential functional and morphological effects on the developing nervous system of the offspring that may arise from exposure *in utero* and during early life. Developmental neurotoxicity cohorts may also be added to other OECD studies, and is included in the Extended One-Generation Reproductive Toxicity Study (EOGRTS – OECD TG 443). In OECD TG 426, test substance is administered to animals during gestation and lactation (PND 21). It may extend post-weaning in young adulthood (PND 60-70). Dams are tested to assess effects in pregnant and lactating females and may also provide comparative information (dams versus offspring). Offspring are tested during postnatal development and adulthood for gross neurologic and behavioural abnormalities, physical development, behavioural ontogeny, motor activity, motor and sensory function, learning and memory, brain weights, and neuropathology. The preferred species is the rat. The recommended route of administration is oral, by gavage, via the diet or in drinking water. The study is not specifically designed to detect endocrine active substances (EASs), but has endpoints relevant for the assessment of possible endocrine disruption. Some endocrine disruptors (EDs) have been shown to cause developmental neurotoxicity and therefore this assay should detect such effects. Disturbance of the thyroid hormonal system, particularly reduction of thyroid hormones in the fetus, has been shown to cause developmental neurotoxicity (Crofton, 2008; Zoeller, 2010). The exposure of the fetus (which may be a sensitive life stage for endocrine disruption effects) and the duration of dosing makes it an assay that can be used when assessing effects relevant to endocrine disruption. In addition, it provides data on effects related to reproduction and development, in particular the endocrine-sensitive endpoint of sexual maturation.

822. OECD TG 426 was adopted in October 2007 following an expert consultation on neurotoxicity (OECD, 2003a). Additional information on the conduct and interpretation of this test guideline (TG) can be found in Guidance Document (GD) No. 43 on

"Reproductive Toxicity Testing and Assessment" (OECD, 2008b) and GD 20 on "Neurotoxicity Testing" (OECD, 2003b). As all the endpoints are apical, it is difficult to discern mechanism of action from this test alone. Information on mechanism of action needs to be obtained from *in vitro* estrogen/androgen/thyroid/steroidogenesis (E,A,T,S) assays or *in vivo* lower tier tests such as the Uterotrophic Bioassay (UT – OECD TG 440) and the Hershberger Bioassay (H – OECD TG 441).

When/why the assay may be used

823. This assay may form part of the package of studies required for registration of pesticides in some jurisdictions. It is likely to be conducted following a concern for neurotoxicity. It will have at least three dose levels and therefore may be used for hazard identification/characterisation.

824. In order to provide information relevant for assessing whether or not a chemical may fulfil the WHO/IPCS (2002) definition of an ED, the study design has to be sufficiently robust to demonstrate the presence or absence of effects. In the dose selection, the investigator should also consider and ensure that data generated are adequate to fulfil the regulatory requirement across OECD countries as appropriate (e.g. hazard and risk assessment and labelling, ED assessment, etc.). The top dose or concentration should be sufficiently high to give clear systemic (i.e. non endocrine-specific) toxicity in order to ensure that a wide range of exposures (high to low) is tested. However, endocrine effects observed solely in the presence of clear systemic toxicity should be interpreted with caution and may be disregarded when sufficiently justified to be caused by secondary effects which are unlikely to be due to endocrine activity. The reason for this advice is a concern that some EAS-sensitive assays are being run at doses/concentrations of EASs that are too low to trigger direct impacts on the endocrine system. This GD is not the place to address this issue directly, but it should be considered when EAS-sensitive TGs are revised in the future. In addition, the number and spacing of dose/concentration levels should also be adequate to fulfil the objectives of the study (e.g. to demonstrate dose response relationships if this is required).

Introduction to the table of scenarios

825. Table C.3.8 gives guidance on a further step to take in the event of a positive (+) or negative (-) result and in the presence of positive (+), negative (-) or equivocal/absent (Eq/0) existing results. "Existing results" are subdivided into "mechanism" and "effects" data (third and fourth columns). The table is divided horizontally into a series of scenarios that represent all the combinations of these events.

826. The results of OECD TG 426 are given in the second column. As this assay is not a screening test where a yes/no (qualitative) answer is obtained, criteria for positive results for the endocrine endpoints are not given in the test guideline. Results for the endpoints would be considered both individually and as a whole. It is not possible to provide guidance on all endpoints individually and for this test all endpoints are considered to be "apical".

827. For the purpose of this guidance, a positive result is defined as a biologically significant change in any of the endocrine endpoints (e.g. a significant alteration in vaginal opening in the absence of body weight changes).

828. A negative result for OECD TG 426 is taken to be the absence of biologically significant changes in all of the endocrine endpoints.

829. Equivocal results for the guideline are not considered in the table, partly for brevity but also because equivocal results are by nature uncertain. A decision must eventually be reached about whether the endocrine endpoints tend to be positive or negative or whether the result must be put to one side and the test repeated or supplemented by a different test.

Existing data to be considered

830. Existing "mechanism" *in vitro* data are assumed to be available from estrogen receptor (ER-), androgen receptor (AR-) and steroidogenesis-based assays (Level 2). Assays may also be available for interference with thyroid modalities. In practice, it is possible that data from all of these assays may not be available, so judgement will need to be used to decide which assays to perform. Although the current *in vitro* test guidelines do not incorporate metabolic activation, published information on use of metabolic activation systems is available in Jacobs et al. (2008; 2013) and OECD (2008a). These methods, however, have not yet been validated.

831. Existing "effects" data refer to *in vivo* effects that may come from lower level assays, e.g. UT or H Assays (Level 3); Peripubertal (PP) Assays or OECD TG 407 assays (Level 4), or there may be longer term studies (e.g. in the case of pesticide registration packages where 90-day and carcinogenicity studies may be available). Data may also be available on effects in mammalian and non-mammalian wildlife species, although caution should be used when extrapolating between taxa. A chemical causing endocrine effects in non-mammalian environmental species (fish, for example) may also have endocrine effects in mammals, but the physiological consequences of the effects are likely to be different.

832. When considering the results of OECD TG 426, all available data should be used in order to reach a conclusion and a weight of evidence approach taken. This may include high throughput screening data, read-across data from structural analogues and quantitative structure activity relationship (QSAR). Several QSAR models for ER and AR binding/activation are now available (see Sections B.1.1.1 and B.1.1.2).

Scenarios: Positive and negative results combined with existing data

833. The scenarios (A to R) presented in Table C.3.8 represent all the possibilities of positive or negative results in combination with the presence or absence of existing data. The action taken will also depend on the regulatory environment, but the considerations given here are generally science based. Although rats are the preferred species for OECD TG 426, the well-conserved nature of the hormonal pathways across taxa should be an indication that results in this assay may be relevant to other vertebrate species. Effects in laboratory mammal tests are also highly relevant for environmental mammalian species. Wherever possible, the recommended "next step which could be taken" avoids unnecessary animal testing. However, sometimes conducting an animal test will be indicated and then the relevance of species, strain, exposure route and species-specific metabolism should always be considered. The sensitivity and physiological function of the hormone under investigation in the test species should also be considered. In general, lower level tests should be conducted before higher level tests in order to avoid unnecessary animal usage, unless it is apparent that a Level 5 test will be required anyway or will be needed to establish the evidence to conclude on ED properties. Information on some endocrine-related tumours may be detected more comprehensively in carcinogenicity studies (OECD TG 451/453) (Level 4); for example, detection of certain types of thyroid tumors in the absence of reproductive or developmental effects, as well as substances causing tumors in other

endocrine-sensitive tissues. At Level 5, the EOGRTS (OECD TG 443) is the most sensitive reproduction assay for detecting endocrine disruption because it includes evaluation of a number of endocrine endpoints not included in the two-generation study (OECD TG 416) adopted in 2001. Further considerations specific to each scenario are given in the table.

834. Scenarios A to C represent positive results in OECD TG 426 in the presence of positive *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. A positive result in the *in vitro* assays in combination with a positive OECD TG 426 assay is evidence of adverse effects on reproduction/development (including neurodevelopment) via E,A,T,S mechanisms. In the absence of robust upper-level data, the next step may be to conduct an upper-level test. In the presence of robust *in vivo* data, there may be sufficient evidence to conclude concern for endocrine disruption and therefore no need for further testing. Positive results in the OECD TG 426 assay may also indicate the potential for endocrine mediated effects in lower vertebrates. These could be followed up with partial life cycle tests such as the Larval Amphibian Growth and Development Assay (LAGDA) or the Medaka Extended One-Generation Reproduction Test (MEOGRT) if the evidence is strong enough. *In vivo* assays/tests with negative results should be interpreted with caution as they may either indicate that the tests used do not have sufficient power to detect weak effects or, alternatively, that the effects do not present a concern for endocrine disruption. The possibility of other (non-E,A,T,S) mechanisms should also not be overlooked (e.g. involving other receptors or endocrine axes).

835. Scenarios D to F represent positive results in OECD TG 426 in the presence of negative *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. A positive result in OECD TG 426 is evidence of adverse effects on reproduction/development (including neurodevelopment). Differential effects on the different endpoints may assist with interpretation. Negative results in the *in vitro* assays should be viewed with caution in case a metabolite is responsible for the positive OECD TG 416 study. If the metabolic profile of the test substance is not known, performing the *in vitro* assays with addition of a metabolising system may help to understand mechanism. Positive results in the OECD TG 426 assay may also indicate the potential for endocrine mediated effects in lower vertebrates. As in Scenarios A to C, *in vivo* assays/tests with negative results should be interpreted with caution as they may either indicate that the tests used do not have sufficient power to detect weak effects or, alternatively, that the effects do not present a concern for endocrine disruption.

836. Scenarios G to I represent positive results in OECD TG 426 in the presence of various combinations of missing or equivocal data. Positive results in the OECD TG 426 assay may also indicate the potential for endocrine mediated effects in lower vertebrates. The next step to take in these eventualities will depend on the nature of the other available data and the jurisdiction in which it is being used. In some cases, equivocal data may be viewed as positive whilst in others it may or may not contribute to the weight of evidence. The interpretation may also depend on the mode of action (MOA) in question and why the data are considered equivocal, e.g. a study that is equivocal for thyroid effects may still be of value in evaluating (anti)androgenic effects. In all three scenarios, the recommended first step is to obtain reliable mechanistic (*in vitro*) data rather than proceed further with *in vivo* testing. Equivocal and missing data are alternative scenarios and two possibilities for the next step are given in most cases, but the nature of equivocal data means that decisions need to be taken on a case-by-case basis. In all cases, the role of metabolism, route of exposure and data from structural analogues should be considered before deciding on the next step.

837. Scenarios J to L represent negative results in OECD TG 426 in the presence of positive *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. In all scenarios, the small number of (non-neuro) endocrine-sensitive endpoints in OECD TG 426 means that an absence of effect in this assay alone cannot lead to a conclusion that a substance does not have endocrine disrupting effects. Additional data from more comprehensive assays are required. All three scenarios could fit a chemical that is positive in *in vitro* assays but is metabolised to a non-active metabolite, leading to negative results in OECD TG 426. This possibility may be investigated to help understand mechanism. Endocrine active potency may also explain differences between *in vitro* and *in vivo* results (e.g. a chemical with weak endocrine activity may give a positive result *in vitro* but may be negative *in vivo*). Positive *in vivo* effects data may involve E,A,T,S or non-E,A,T,S mechanisms (e.g. involving other receptors or endocrine axes), more sensitive endpoints, greater statistical power, but knowledge of absorption, distribution, metabolism and excretion (ADME) may help to explain differences from the OECD TG 426 data.

838. Scenarios M to O represent negative results in OECD TG 426 in the presence of negative *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. In all scenarios, the small number of endocrine-sensitive endpoints in OECD TG 426 means that an absence of effect in this assay alone cannot lead to a conclusion that a substance does not have endocrine disrupting effects. Positive *in vivo* effects data may involve E,A,T,S or non-E,A,T,S mechanisms (e.g. involving other receptors or endocrine axes) but knowledge of ADME may help to explain differences from the OECD TG 426 data.

839. Scenarios P to R represent negative results in OECD TG 426 in the presence of various combinations of missing or equivocal data. As with the positive result scenarios above, the next step to take in these eventualities will have to be decided on a case-by-case basis. In all cases, the role of metabolism, route of exposure and data from structural analogues should be considered before deciding on the next step.

840. In all scenarios (A to R), the next step to take to strengthen weight of evidence will depend on the existing information. Table C.3.8 is meant to provide a succinct guide and may not cover all circumstances or possibilities. The scenarios may also suggest that chemicals have simple or single MOA, when in practice they may have multiple endocrine and non-endocrine MOA. In some cases, for example, two opposite modes of simultaneous action (e.g. estrogenic and anti-estrogenic) could, depending on dose, lead to a minimisation or abolition of effects, while in others two different MOA (e.g. estrogenic and anti-androgenic) could potentially reinforce effects. Endocrine pathways interact, mixed effects are common and there are many pathways that cannot be distinguished with currently available TGs. If multiple MOA are suspected, either from the existing results or based on QSAR/read-across/integrated approaches, this should be investigated further if needed for regulatory decision making.

References

Crofton, K.M. (2008), "Thyroid disrupting chemicals: Mechanisms and mixtures", *International Journal of Andrology*, Vol. 31/2, pp. 209-223, https://doi.org/10.1111/j.1365-2605.2007.00857.x.

Jacobs, M. et al. (2013), "*In vitro* metabolism and bioavailability tests for endocrine active substances: What is needed next for regulatory purposes?", *ALTEX – Alternatives to Animal Experimentation*, Vol. 30/3, pp. 331-351.

Jacobs, M.N. et al. (2008), "The use of metabolising systems for *in vitro* testing of endocrine disrupters", *Current Drug Metabolism*, Vol. 9/8, pp. 796-826.

OECD (2008a), *Detailed Review Paper on the Use of Metabolising Systems for* In Vitro *Testing of Endocrine Disrupters*, OECD Series on Testing and Assessment, No. 97, OECD Publishing, Paris, https://doi.org/10.1787/9789264085497-en.

OECD (2008b), "Guidance document on mammalian reproductive toxicity testing and assessment", OECD Series on Testing and Assessment, No. 43, OECD, Paris, www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/JM/MONO(2008)16&doclanguage=en.

OECD (2003a), "Report of the OECD Expert Consultation Meeting on Developmental Neurotoxicity Testing", OECD, Paris.

OECD (2003b), "Guidance Document for Neurotoxicity Testing", OECD Series on Testing and Assessment, No. 20, OECD, Paris, www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2004)25&doclanguage=en.

WHO/IPCS (2002), "Global assessment of the state-of-the-science of endocrine disrupters", Damstra, T. et al. (eds.) WHO/PCS/EDC/02.2, World Health Organization, Geneva, www.who.int/ipcs/publications/new_issues/endocrine_disruptors/en.

Zoeller, R.T. (2010), "Environmental chemicals targeting thyroid", *Hormones (Athens, Grece)*, Vol. 9/1, pp. 28-40, http://dx.doi.org/10.14310/horm.2002.1250.

Table C.3.8. **Developmental Neurotoxicity Study (OECD TG 426): Guidance for scenarios of combinations of results with existing data**

This table represents possible conclusions to be drawn from assay data, and a next step which could be taken if further evidence is required about possible endocrine disrupting properties and/or effects. The guidance offered is not meant to be prescriptive, but provides science-based considerations. It encourages the use of all available data and expert judgement in a weight of evidence approach. Regional and national interpretation of results and "next steps" may vary.

The conclusions are grouped into a series of scenarios (A-R), each scenario representing a different combination of assay results, existing *in vitro* data and existing *in vivo* data. The symbol "+" indicates that the data in question represent a positive result, «-" indicates a negative result, and "Eq/0" indicates that the data are either equivocal or are not available.

Existing results: * "Mechanism (*in vitro* mechanistic data)" assumes that mechanistic data are available from estrogen receptor (ER-), androgen receptor (AR-) and steroidogenesis-based assays (Level 2). Thyroid hormone receptor (TR) and other assays concerning mechanisms of thyroid disruption may be available, but they are not in common use. In practice, data from all assays may not be available and therefore this must be taken into account when deciding on the "next step". Quantitative structure activity relationship (QSAR) predictions of estrogen and androgen binding/activation may be made for some substances.

Existing results: ** "Effects (*in vivo* effects of concern)" assumes effects have been observed in other *in vivo* screens/tests which give rise to concern that the test chemical may be an endocrine disrupter. These may be other repeated dose toxicity tests, Uterotrophic Bioassays (UT) and Hershberger Bioassays, Peripubertal (PP) Assays, or read-across from chemical analogues.

*** *Note:* a positive result is defined as a biologically significant change in any of the endocrine endpoints (all "apical endpoints").

Scenarios	Result of OECD TG 426 (DNT study)	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
A	+ ***	+	+	Evidence for adverse effects on (neuro-) development in OECD TG 426, possibly via (anti)-E,A,T,S activity.	Further testing may not be required. May perform assay from Level 5 (e.g. EOGRTS or two-generation assay) if needed.	Note that the EOGRTS provides the most information on endocrine disruption; however, for endocrine disrupting chemicals (EDCs) with a carcinogenic potential, OECD TG 451-3 may be more sensitive. Effects on apical endpoints may indicate E,A,T,S modalities or other mechanisms. Consider potency of effects for existing results and whether E,A,T,S mechanism is credible for reproductive/developmental effects or whether there may be non-endocrine mechanisms. Consider route of exposures for effects data and possible implications of absorption, distribution, metabolism and excretion (ADME) characteristics of the chemical. Endocrine activity possible in lower vertebrates. Consider performing a Larval Amphibian Growth and Development Assay (LAGDA) or Medaka Extended One-Generation Reproduction Test (MEOGRT).
B	+	+	–	Evidence for adverse effects on (neuro-) development in OECD TG 426, possibly via (anti)-E,A,T,S activity.	Further testing may not be required. May perform assay from Level 5 (e.g. EOGRTS or two-generation assay) if needed.	Note that the EOGRTS provides the most information on endocrine disruption; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive. Effects on apical endpoints may indicate E,A,T,S modalities or other mechanisms. Consider potency of effects for existing results and whether E,A,T,S mechanism is credible for reproductive/developmental effects or whether there may be non-endocrine mechanisms. Consider route of exposures for effects data and possible implications of ADME characteristics of the chemical. Hormonal activity possible in lower vertebrates. Consider performing a LAGDA or MEOGRT.
C	+	+	Eq/0	Evidence for adverse effects on (neuro-) development in OECD TG 426, possibly via (anti)-E,A,T,S activity.	Further testing may not be required. May perform assay from Level 5 (e.g. EOGRTS or two-generation assay) if needed.	Note that the EOGRTS provides the most information on endocrine disruption; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive. Effects on apical endpoints may indicate E,A,T,S modalities or other mechanisms. Consider potency of effects for existing results and whether E,A,T,S mechanism is credible for reproductive/developmental effects or whether there may be non-endocrine mechanisms. Consider route of exposures for effects data and possible implications of ADME characteristics of the chemical. Hormonal activity possible in lower vertebrates. Consider performing a LAGDA or MEOGRT. Equivocal results may indicate chemical has multiple modes of action (MOA).

Scenarios	Result of OECD TG 426 (DNT study)	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
D	+	–	+	Evidence for adverse effects on (neuro-) development in OECD TG 426, but not via E,A,T,S mechanism or requires metabolic activation for activity.	Further testing may not be required. To further discern mechanism, could perform *in vitro* estrogen receptor (ER-), androgen receptor (AR-), thyroid hormone receptor (TR), steroidogenesis (S) assays with added metabolising system.	Note that the EOGRTS provides the most information on endocrine disruption; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive. Effects on apical endpoints may indicate E,A,T,S modalities or other mechanisms. Consider potency of effects for existing results and whether endocrine disruption mechanism is credible for reproductive/developmental effects or whether there may be non-endocrine mechanisms. Consider route of exposures for effects data and possible implications of ADME characteristics of the chemical. Hormonal activity possible in lower vertebrates. Consider performing a LAGDA or MEOGRT.
E	+	–	–	Evidence for adverse effects on (neuro-) development in OECD TG 426 via non-E,A,T,S/non-endocrine disruption mechanism or requires metabolic activation for activity.	Further testing may not be required. To further discern mechanism, could perform *in vitro* ER, AR, TR, S assays with added metabolising system.	Sufficient information to conclude evidence of concern for reproductive toxicity via unknown mechanism. Note that the EOGRTS provides the most information on endocrine disruption; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive. Effects on apical endpoints may indicate E,A,T,S modalities or other mechanisms. Consider potency of effects for existing results and whether endocrine disruption mechanism is credible for reproductive/developmental effects or whether there may be non-endocrine mechanisms. Consider route of exposures for effects data and possible implications of ADME characteristics of the chemical. Hormonal activity possible in lower vertebrates. Consider performing a LAGDA or MEOGRT.
F	+	–	Eq/0	Evidence for adverse effects on (neuro-) development in OECD TG 426 via non-E,A,T,S/non-endocrine disruption mechanism or requires metabolic activation for activity.	Further testing may not be required. To further discern mechanism, could perform *in vitro* ER, AR, TR, S assays with added metabolising system.	Note that the EOGRTS provides the most information on endocrine disruption; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive. Effects on apical endpoints may indicate E,A,T,S modalities or other mechanisms. Consider potency of effects for existing results and whether endocrine disruption mechanism is credible for reproductive/developmental effects or whether there may be non-endocrine mechanisms. Equivocal results may indicate chemical has multiple MOA. Consider route of exposures for effects data and possible implications of ADME characteristics of the chemical. Hormonal activity possible in lower vertebrates. Consider performing a LAGDA or MEOGRT.

Scenarios	Result of OECD TG 426 (DNT study)	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
G	+	Eq/0	+	Evidence for adverse effects on (neuro-) development in OECD TG 426, may act via E,A,T,S mechanism and may require metabolic activation for activity.	Further testing may not be required. To further discern mechanism, could perform *in vitro* ER, AR, TR, S assays with added metabolising system.	Note that the EOGRTS provides the most information on endocrine disruption; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive. Effects on apical endpoints may indicate E,A,T,S modalities or other mechanisms. Equivocal results may indicate chemical has multiple MOA. Consider potency of effects for existing results and whether endocrine disruption mechanism is credible for reproductive/developmental effects or whether there may be non-endocrine mechanisms. Consider route of exposures for effects data and possible implications of ADME characteristics of the chemical. Hormonal activity possible in lower vertebrates. Consider performing a LAGDA or MEOGRT.
H	+	Eq/0	–	Evidence for adverse effects on (neuro-) development in OECD TG 426 via non-E,A,T,S/non-endocrine disruption mechanism or requires metabolic activation for activity.	Further testing may not be required. To further discern mechanism, could perform *in vitro* ER, AR, TR, S assays with added metabolising system.	Effects on apical endpoints may indicate E,A,T,S modalities or other mechanisms. Equivocal results may indicate chemical has multiple MOA. Consider potency of effects for existing results and whether endocrine disruption mechanism is credible for reproductive/developmental effects or whether there may be non-endocrine mechanisms. Consider route of exposures for effects data and possible implications of ADME characteristics of the chemical. Hormonal activity possible in lower vertebrates. Consider performing a LAGDA or MEOGRT.
I	+	Eq/0	Eq/0	Evidence for adverse effects on (neuro-) development in OECD TG 426 via unknown mechanism.	Further testing may not be required. To further discern mechanism, could perform *in vitro* ER, AR, TR, S assays with added metabolising system.	Consider existing results and whether endocrine disruption mechanism is credible for reproductive/developmental effects or whether there may be non-endocrine mechanisms. Equivocal results may indicate chemical has multiple MOA. Consider route of exposures for effects data and possible implications of ADME characteristics of the chemical. Hormonal activity possible in lower vertebrates. Consider performing a LAGDA or MEOGRT.
J	–	+	+	No evidence of adverse effects on (neuro-) development in OECD TG 426. Effects seen in existing (lower level) studies do not lead to adverse outcome. Metabolism or potency explains the difference from existing *in vitro* and *in vivo* data.	Consider supplemental testing, depending on existing data. To further discern mechanism, could perform *in vitro* ER, AR, TR, S assays with added metabolising system.	Note that the EOGRTS provides the most information on endocrine disruption; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive. Consider route of exposures and possible implications for ADME characteristics of the chemical with existing studies. Further mechanistic studies with metabolism may help determine MOA.

Scenarios	Result of OECD TG 426 (DNT study)	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
K	–	+	–	No evidence of adverse effects in OECD TG 426. Metabolism or potency explains *in vitro*/*in vivo* differences.	Consider supplemental testing, depending on existing data. To further discern mechanism, could perform *in vitro* ER, AR, TR, S assays with added metabolising system.	If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption. Further mechanistic studies with metabolism may help determine MOA.
L	–	+	Eq/0	No evidence of adverse effects in OECD TG 426. Metabolism or potency explains *in vitro*/*in vivo* differences. Effects seen in existing (lower level) studies do not lead to adverse outcome.	Consider supplemental testing, depending on existing data. To further discern mechanism, could perform *in vitro* ER, AR, TR, S assays with added metabolising system.	Note that the EOGRTS provides the most information on endocrine disruption; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive. Consider route of exposures and possible implications for ADME characteristics of the chemical with existing studies. Further mechanistic studies with metabolism may help determine MOA. Equivocal results may indicate chemical has multiple MOA.
M	–	–	+	No evidence of adverse effects in OECD TG 426. Effects seen in existing (lower level) studies do not lead to adverse outcome.	Consider supplemental testing, depending on existing data.	If existing data are from adequate *in vivo* studies such as 28-day, 90-day, chronic/carcinogenicity studies, than question why there are differences. Note that the EOGRTS provides the most information on endocrine disruption; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive. Consider route of exposures and possible implications for ADME characteristics of the chemical with existing studies. Further mechanistic studies with metabolism may help determine MOA.
N	–	–	–	No evidence of adverse effects in OECD TG 426.	Consider supplemental testing, depending on existing data.	If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption. Note that the EOGRTS provides the most information on endocrine disruption; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive.
O	–	–	Eq/0	No evidence of adverse effects in OECD TG 426. No evidence for (anti)-E,A,T,S activity *in vitro*.	Consider supplemental testing, depending on existing data. To further discern mechanism, could perform *in vitro* ER, AR, TR, S assays with added metabolising system.	Note that the EOGRTS provides the most information on endocrine disruption; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive. Further mechanistic studies with metabolism may help determine MOA. Consider route of exposures and possible implications for ADME characteristics of the chemical with existing studies. Check data on chemical analogues.

Scenarios	Result of OECD TG 426 (DNT study)	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
P	–	Eq/0	+	No evidence of adverse effects in OECD TG 426. Effects seen in existing (lower level) studies do not lead to adverse. Effects seen in existing studies are via unknown mechanism.	Consider supplemental testing, depending on existing data. To further discern mechanism, could perform *in vitro* ER, AR, TR, S assays with added metabolising system.	If existing data are from adequate *in vivo* studies such as 28-day, 90-day, chronic/carcinogenicity studies, than question why there are differences. Note that the EOGRTS provides the most information on endocrine disruption; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive. Consider route of exposures and possible implications for ADME characteristics of the chemical with existing studies. Check data on chemical analogues. Equivocal results may indicate chemical has multiple MOA.
Q	–	Eq/0	–	No evidence of adverse effects in OECD TG 426.	Consider supplemental testing, depending on existing data.	There may be sufficient information to conclude absence of concern for endocrine disruption. Check data on chemical analogues.
R	–	Eq/0	Eq/0	No evidence of adverse effects in OECD TG 426	Consider supplemental testing, depending on existing data.	Note that the EOGRTS provides the most information on endocrine disruption; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive. Further mechanistic studies may strengthen weight of evidence. Equivocal results may indicate chemical has multiple MOA. Check data on chemical analogues.

C.3.9. Repeated Dose Dermal Toxicity: 21/28-Day Study (OECD TG 410)

Status: Assay validated by the OECD.

841. Modalities detected: (anti)estrogen, (anti)androgen, steroidogenesis.

Endpoints: Weight of adrenals, testes. Other (target) organs may also be examined. Histopathological examination of tissues may take place.

Background to the assay

842. This assay determines the subchronic dermal toxicity of chemicals after initial information on toxicity has been obtained by acute testing. It provides information on possible health hazards likely to arise from repeated exposure by the dermal route over a limited period of time. Dosing duration is 28 days and the preferred species are the adult rat, rabbit or guinea pig. Test substance is applied to the dorsal area of the trunk, held in place with a dressing and protected from ingestion. OECD TG 410 was adopted in May 1981. Although it has not been validated for the detection of endocrine active substances (EASs), this assay may contain endpoints that are suitable for the determination of endocrine effects. It should be noted that the only endocrine-relevant tissues that are required are testes and adrenals, therefore the information provided may be very limited. Other tissues may be taken as required and therefore the utility of this assay for detecting EASs will vary on a case-to-case basis. As all the endpoints are apical, it is difficult to discern mechanism of action from this test alone. Information on mechanism of action needs to be obtained from *in vitro* estrogen/androgen/thyroid/steroidogenesis (E,A,T,S) assays or *in vivo* lower tier tests such as the Uterotrophic Bioassay (UT – OECD TG 440) and the Hershberger Bioassay (H – OECD TG 441). Hormone measurements are **not** included in this assay.

843. A comparison can be made with OECD TG 407 (28-Day Oral Toxicity Study) for endocrine endpoints, although dermal absorption of test substances is likely to result in lower internal doses compared to oral administration. However, the duration of dosing means that it may be a relevant assay to assess when determining potential endocrine activity. Dermal exposure may be a relevant route of human exposure to certain substances.

When/why the assay may be used

844. This assay may be used as part of a chemical submission package and forms part of the standard information requirements in certain chemical legislations. At least three dose levels are included so that an estimate of no-adverse-effect-level can be determined and the assay used for hazard identification/characterisation. It should be noted that as this assay is not primarily designed to detect endocrine disruption, a higher degree of systemic toxicity is typically induced than is the case with the other Level 3 and 4 assays. The possibly confounding effect of systemic toxicity on endocrine endpoints therefore needs to be considered.

845. In order to provide information relevant for assessing whether or not a chemical may fulfil the WHO/IPCS (2002) definition of an endocrine disruptor (ED), the study design has to be sufficiently robust to demonstrate the presence or absence of effects. In the dose selection, the investigator should also consider and ensure that data generated are adequate to fulfil the regulatory requirement across OECD countries as appropriate (e.g. hazard and risk assessment and labelling, ED assessment, etc.). The top dose or concentration should be sufficiently high to give clear systemic (i.e. non endocrine-specific) toxicity in order to ensure that a wide range of exposures (high to low) is tested. However, endocrine effects observed solely in the presence of clear systemic toxicity should be interpreted with caution and may be disregarded when sufficiently justified to be caused by secondary effects which are unlikely to be due to endocrine activity. The reason for this advice is a concern that some EAS-sensitive assays are being run at doses/concentrations of EASs that are too low to trigger direct impacts on the endocrine system. This guidance document is not the place to address this issue directly, but it should be considered when EAS-sensitive test guidelines (TGs) are revised in the future. In addition, the number and spacing of dose/concentration levels should also be adequate to fulfil the objectives of the study (e.g. to demonstrate dose response relationships if this is required).

Introduction to the table of scenarios

846. Table C.3.9 gives guidance on a further step to take in the event of a positive (+) or negative (-) result and in the presence of positive (+), negative (-) or equivocal/absent (Eq/0) existing results. “Existing results” are subdivided into “mechanism” and “effects” data (thirrd and fourth columns). The table is divided horizontally into a series of scenarios that represent all the combinations of these events.

847. The results of OECD TG 410 are given in the second column. As OECD TG 410 is not a screening test where a yes/no (qualitative) answer is obtained for the test as a whole, positive results would generally be assessed for individual endpoints. For the purposes of this guidance, however, a positive result is defined as a biologically significant change in any of the endocrine endpoints listed above (e.g. statistically significant reductions in reproductive organ weights). Changes in related endpoints will increase their biological significance (e.g. changes in the weights of testes and accompanied by histopathological changes). The guidance on histopathologic changes in endocrine tests (OECD, 2009) may be helpful in interpretation. A negative result for OECD TG 410 is taken to be the absence of biologically significant changes in all endocrine endpoints.

848. In the absence of other pertinent lines of evidence, negative results in this test alone cannot be taken as evidence that the substance is not an ED. Further studies will be required as confirmation.

849. Equivocal results for the guideline are not considered in the table, partly for brevity but also because equivocal results are by nature uncertain. A decision must eventually be reached about whether the endocrine endpoints tend to be positive or negative or whether the result must be put to one side and the test repeated (using the same or a different test guideline). Factors which may have interfered with the result (e.g. composition of the diet used, environmental influences) should be considered.

Existing data to be considered

850. Existing “mechanism” *in vitro* data are assumed to be available from estrogen receptor (ER-), androgen receptor (AR-) and steroidogenesis-based assays (Level 2). Assays

may also be available for interference with thyroid modalities. In practice, it is possible that data from all of these assays may not be available, so judgement will need to be used to decide which assays to perform. Although the current *in vitro* test guidelines do not incorporate metabolic activation, published information on use of metabolic activation systems is available in Jacobs et al. (2008; 2013) and OECD (2008). These methods, however, have not yet been validated.

851. Existing "effects" data refer to *in vivo* effects that may come from Level 3 or 4 tests in the Conceptual Framework (e.g. UT or H assays). In these cases, it should be remembered that these assays are specifically designed to be sensitive to EASs. As mentioned above, the results of the study may be interpreted as part of a battery or group of tests carried out for regulatory purposes. Data may also be available on effects in mammalian and non-mammalian wildlife species, although caution should be used when extrapolating between taxa. A chemical causing endocrine effects in non-mammalian environmental species (fish, for example) may also have endocrine effects in mammals, but the physiological consequences of the effects are likely to be different.

852. When considering the results of the OECD TG 410 assay, all available data should be used in order to reach a conclusion and a weight of evidence approach taken. This may include high throughput screening data, read-across data from structural analogues and quantitative structure activity relationship (QSAR). Several QSAR models for ER and AR binding/activation are now available (see Sections B.1.1.1 and B.1.1.2).

Scenarios: Positive and negative results combined with existing data

853. A series of scenarios (A to R) are presented in Table C.3.9 and represent all the possibilities of positive or negative results in combination with the presence or absence of existing data. The action taken will also depend on the regulatory environment, but the considerations given here are generally science based. Although the OECD TG 410 assay uses mammals, the well-conserved nature of the hormonal pathways across taxa indicate that results on endocrine endpoints in this assay may be relevant to other vertebrate species. Effects in laboratory mammal tests are also highly relevant for environmental mammalian species. Wherever possible, the recommended "next step which could be taken" avoids unnecessary animal testing. However, sometimes conducting an animal test will be indicated and then the relevance of species, strain and exposure route should always be considered. The sensitivity and physiological function of the hormone under investigation in the test species should also be considered. In general, Level 2 or 3 tests should be conducted before Level 5 tests in order to avoid unnecessary animal usage, unless it is apparent that a Level 5 test will be required anyway or will be needed to establish the evidence to conclude on ED properties. Information on some endocrine-related tumours may be detected more comprehensively in carcinogenicity studies (OECD TG 451/453) (Level 4); for example, detection of certain types of thyroid tumors in the absence of reproductive or developmental effects, as well as substances causing tumors in other endocrine-sensitive tissues. It is recognised, however, that some jurisdictions may require a two-generation study. Further considerations specific to each scenario are given in the table.

854. Scenarios A to C represent positive results in the OECD TG 410 assay in the presence of positive *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. A positive result in the *in vitro* assays in combination with a positive OECD TG 410 assay is moderate or strong evidence for E,A,T,S-mediated activity that may or may not be supported by the *in vivo* effects data. In the absence of robust upper-level data,

the next step may be to conduct an upper-level test. In the presence of robust *in vivo* data, there may be sufficient evidence to conclude concern for endocrine disruption and therefore no need for further testing. Positive results in the OECD TG 410 assay may also indicate the potential for endocrine mediated effects in lower vertebrates. These could be followed up with partial life cycle tests such as the Fish Sexual Development Test (FSDT), the Larval Amphibian Growth and Development Assay (LAGDA) or the Medaka Extended One-Generation Reproduction Test (MEOGRT) if the evidence were strong enough. *In vivo* assays/tests with negative results should be interpreted with caution as they may either indicate that the tests used do not have sufficient power to detect weak effects or, alternatively, that the effects do not present a concern for endocrine disruption. The possibility of other (non-E,A,T,S) mechanisms should also not be overlooked (e.g. involving other receptors or endocrine axes).

855. Scenarios D to F represent positive results in the OECD TG 410 assay in the presence of negative *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. Negative results in the *in vitro* assays should be viewed with caution in case a metabolite is responsible for the positive OECD TG 410 assay. Unless the metabolic profile of the test substance is known, one option may be to conduct these assays with an added metabolising system. If the metabolic profile is known, then a Level 5 *in vivo* test may be advisable. The choice of tests will depend on the available *in vivo* effects data. Positive results in the OECD TG 410 assay may also indicate the potential for endocrine mediated effects in lower vertebrates. As in Scenarios A to C, *in vivo* assays/tests with negative results should be interpreted with caution as they may either indicate that the tests used do not have sufficient power to detect weak effects or, alternatively, that the effects do not present a concern for endocrine disruption.

856. Scenarios G to I represent positive results in the OECD TG 410 assay in the presence of various combinations of missing or equivocal data. Positive results in the OECD TG 410 assay may also indicate the potential for endocrine mediated effects in lower vertebrates. The next step to take in these eventualities will depend on the nature of the other available data and the jurisdiction in which it is being used. In some cases, equivocal data may be viewed as positive whilst in others it may or may not contribute to the weight of evidence. The interpretation may also depend on the mode of action (MOA) in question and why the data are considered equivocal. In all three scenarios, the recommended first step is to obtain reliable mechanistic (*in vitro*) data rather than proceed further with *in vivo* testing. Equivocal and missing data are alternative scenarios and two possibilities for the next step are given in most cases, but the nature of equivocal data means that decisions need to be taken on a case-by-case basis. In all cases, the role of metabolism, route of exposure and data from structural analogues should be considered before deciding on the next step.

857. Scenarios J to L represent negative results in the OECD TG 410 assay in the presence of positive *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. Negative outcomes in the OECD TG 410 should be viewed with caution because of the power of the assay to detect (anti)estrogens and androgens may be limited. All three scenarios could also arise from a chemical that is positive in *in vitro* assays, but is metabolised to a non-active metabolite leading to negative results in the OECD TG 410 assay. This should be considered first when investigating the next step. Endocrine active potency may also explain differences between *in vitro* and *in vivo* results (e.g. a chemical with weak endocrine activity may give a positive result *in vitro* but may be negative *in vivo*). Positive *in vivo* effects data may involve other E,A,T,S, non-E,A,T,S mechanisms (e.g. involving other receptors or endocrine axes), more sensitive endpoints, greater

statistical power or life stages that are more sensitive to the substance than the adult dermally exposed animals in OECD TG 410.

858. Scenarios M to O represent negative results in the OECD TG 410 assay in the presence of negative *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. Negative results for all tests (Scenario N) may be sufficient to enable a conclusion of no concern for endocrine disruption. This will depend on the weight of evidence and may not be possible. Where there are positive *in vivo* effects data there could still be an E,A,T,S-related mechanism, the effects may be related to length of exposure, route of exposure or exposure at different life stages. Other E,A,T,S or non-E,A,T,S mechanisms may also be involved.

859. Scenarios P to R represent negative results in the OECD TG 410 assay in the presence of various combinations of missing or equivocal data. As with the positive result scenarios above (see Paragraph 857) the next step to take in these eventualities will have to be decided on a case-by-case basis. However, the recommended first step is generally to obtain reliable mechanistic (*in vitro*) data rather than proceed further with *in vivo* testing. In all cases, the role of metabolism, route of exposure and data from structural analogues should be considered before deciding on the next step.

860. In all scenarios (A to R), the next step to take to strengthen weight of evidence will depend on the existing information. The table is meant to provide a succinct guide and may not cover all circumstances or possibilities. The scenarios may also suggest that chemicals have simple or single MOA, when in practice they may have multiple endocrine and non-endocrine MOA. In some cases, for example, two opposite modes of simultaneous action (e.g. estrogenic and anti-estrogenic) could, depending on dose, lead to a minimisation or abolition of effects, while in others two different MOA (e.g. estrogenic and anti-androgenic) could potentially reinforce effects. Endocrine pathways interact, mixed effects are common and there are many pathways that cannot be distinguished with currently available TGs. If multiple MOA are suspected, either from the existing results or based on QSAR/read-across/integrated approaches, this should be investigated further if needed for regulatory decision making.

References

Jacobs, M. et al. (2013), "*In vitro* metabolism and bioavailability tests for endocrine active substances: What is needed next for regulatory purposes?", *ALTEX – Alternatives to Animal Experimentation*, Vol. 30/3, pp. 331-351.

Jacobs, M.N. et al. (2008), "The use of metabolising systems for *in vitro* testing of endocrine disrupters", *Current Drug Metabolism*, Vol. 9/8, pp. 796-826.

OECD (2009), "Guidance document for histologic evaluation of endocrine and reproductive tests in rodents", OECD Series on Testing and Assessment, No. 106, OECD, Paris, www.oecd.org/env/ehs/testing/43411534.pdf.

OECD (2008), *Detailed Review Paper on the Use of Metabolising Systems for* In Vitro *Testing of Endocrine Disrupters*, OECD Series on Testing and Assessment, No. 97, OECD Publishing, Paris, https://doi.org/10.1787/9789264085497-en.

WHO/IPCS (2002), "Global assessment of the state-of-the-science of endocrine disrupters", Damstra, T. et al. (eds.) WHO/PCS/EDC/02.2, World Health Organization, Geneva, www.who.int/ipcs/publications/new_issues/endocrine_disruptors/en.

Table C.3.9. **Repeated Dose Dermal Toxicity: 21/28-Day Study (OECD TG 410): Guidance for scenarios of combinations of results with existing data**

This table represents possible conclusions to be drawn from assay data, and a next step which could be taken if further evidence is required about possible endocrine disrupting properties and/or effects. The guidance offered is not meant to be prescriptive, but provides science-based considerations. It encourages the use of all available data and expert judgement in a weight of evidence approach. Regional and national interpretation of results and "next steps" may vary.

The conclusions are grouped into a series of scenarios (A-R), each scenario representing a different combination of assay results, existing *in vitro* data and existing *in vivo* data. The symbol "+" indicates that the data in question represent a positive result, «-" indicates a negative result, and "Eq/0" indicates that the data are either equivocal or are not available.

Existing results: * "Mechanism (*in vitro* mechanistic data)" assumes that mechanistic data are available from estrogen receptor (ER-), androgen receptor (AR-) and steroidogenesis-based assays (Level 2). Thyroid hormone receptor (TR) and other assays concerning mechanisms of thyroid disruption may be available, but they are not in common use. In practice, data from all assays may not be available and therefore this must be taken into account when deciding on the "next step". Quantitative structure activity relationship (QSAR) predictions of estrogen and androgen binding/activation may be made for some substances.

Existing results: ** "Effects (*in vivo* effects of concern)" assumes various information, such as data from repeat dose oral toxicity studies, reproduction/developmental toxicity screen tests, read-across from analogues, will be available.

*** *Note:* a positive result is defined as a biologically significant change in any of the endocrine endpoints.

Scenarios	Result of OECD TG 410 (28-day dermal assay)	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
A	+ ***	+	+	(Anti)-E,A,T,S activity. Increased evidence of (anti)-E,A,T,S activity.	Perform assay from Level 5 (e.g. Extended One-Generation Reproduction Toxicity Study [EOGRTS] or two-generation assay).	If existing data are from a Level 5 assay, there may be sufficient information to conclude evidence of concern for endocrine disruption (the EOGRTS provides the most information; however, for endocrine disrupting chemicals [EDCs] with a carcinogenic potential, OECD TG 451-3 may be more sensitive). Endocrine activity possible in lower vertebrates. Consider performing Fish Sexual Development Test (FSDT), Larval Amphibian Growth and Development Assay (LAGDA) or Medaka Extended One-Generation Reproduction Test (MEOGRT).
B	+	+	–	(Anti)-E,A,T,S activity. Increased evidence of (anti)-E,A,T,S activity.	Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	If existing data are from an adequate Level 5 assay, question why there are differences. If existing data are from a less sensitive assay, a higher level test may be required. Consider route of exposures and possible implications of absorption, distribution, metabolism and excretion (ADME) characteristics of the chemical. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT.
C	+	+	Eq/0	(Anti)-E,A,T,S activity. Increased evidence of (anti)-E,A,T,S activity.	Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	Check data on chemical analogues. Consider route of exposures and possible implications of ADME characteristics of the chemical. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT. Equivocal results may indicate chemical has multiple modes of action (MOA).
D	+	–	+	(Anti)-E,A,T,S activity. Acts via non-ER, AR, TR, S mechanism or may require metabolic activation for activity.	Perform *in vitro* estrogen receptor (ER), androgen receptor (AR), thyroid hormone receptor (TR), steroidogenesis (S) assays with added metabolising system.	If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude evidence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). Consider route of exposures for effects data and possible implications of ADME characteristics of the chemical. Further mechanistic studies may help determine MOA. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT.
E	+	–	–	(Anti)-E,A,T,S activity. Acts via non-ER, AR, TR, S mechanism or may require metabolic activation for activity. Route of exposure may account for the differences between OECD TG 411 and existing data.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system OR Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	If existing data are from an adequate Level 5 assay, question why there are differences. If existing data are from a less sensitive assay, a higher level test may be required. Consider route of exposures and possible implications of ADME characteristics of the chemical. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT.

Scenarios	Result of OECD TG 410 (28-day dermal assay)	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
F	+	–	Eq/0	(Anti)-E,A,T,S activity. Acts via non-ER, AR, TR, S mechanism or may require metabolic activation for activity.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system OR Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	Check data on chemical analogues. Further mechanistic studies may help determine MOA. If existing data are from an adequate Level 5 assay, question why there are differences. If existing data are from a less sensitive assay, a higher level test may be required. Equivocal results may indicate chemical has multiple MOA. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT.
G	+	Eq/0	+	(Anti)-E,A,T,S activity. May act via ER, AR, TR, S mechanism (metabolic activation may be needed).	Perform *in vitro* ER, AR, TR, S assays (for the "0" scenario, otherwise Eq result available) OR Perform *in vitro* ER, AR, TR, S assays with added metabolising system.	If existing data are from Level 5 then may be sufficient information to conclude evidence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). Check data on chemical analogues. Further mechanistic studies may help determine MOA. Equivocal results may indicate chemical has multiple MOA. Consider route of exposures for effects data and possible implications of ADME characteristics of the chemical. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT.
H	+	Eq/0	–	(Anti)-E,A,T,S activity. Acts via unknown mechanism or may require metabolic activation for activity. Route of exposure may account for the differences between OECD TG 410 and existing data. Unknown potential for adverse effects.	For the "0" scenario, perform *in vitro* ER, AR, TR, S assays, maybe with added metabolising system (otherwise Eq result available).	If existing data are from an adequate Level 5 assay, question why there are differences. Consider route of exposures and possible implications of ADME characteristics of the chemical. If existing data are from a less sensitive assay, a higher level test may be required. Check data on chemical analogues. Further mechanistic studies may help determine MOA. Equivocal results may indicate chemical has multiple MOA. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT.
I	+	Eq/0	Eq/0	(Anti)-E,A,T,S activity. Acts via unknown mechanism. Unknown potential for adverse effects.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system OR Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	Check data on chemical analogues. Further mechanistic studies may help determine MOA. Equivocal results may indicate chemical has multiple MOA. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT.
J	–	+	+	No evidence for (anti)-E,A,T,S activity in OECD TG 410. Weak (anti)-E,A,S activity may not be detected by this assay. Metabolism or potency may explain the difference from existing *in vitro* and *in vivo* data.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system OR Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	If existing data are from an adequate Level 5 assay, question why there are differences. Effects seen in existing studies may be in a more sensitive life stage. Consider route of exposures and possible implications of ADME characteristics of the chemical. Further mechanistic studies may help determine MOA.

Scenarios	Result of OECD TG 410 (28-day dermal assay)	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
K	–	+	–	No evidence for (anti)-E,A,T,S activity in OECD TG 410. Weak (anti)-E,A,S activity may not be detected by this assay. Metabolism or potency may explain *in vitro*/*in vivo* differences.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system OR Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from a less sensitive assay, a higher level test may be required. Further mechanistic studies with metabolism may help determine MOA.
L	–	+	Eq/0	No evidence for (anti)-E,A,T,S activity in OECD TG 410. Weak (anti)-E,A,S activity may not be detected by this assay. Metabolism or potency may explain *in vitro*/*in vivo* differences. Unknown potential for adverse effects.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system OR Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	Metabolic deactivation of chemical may occur *in vivo* so that possible *in vitro* activity is not realised. Consider possible routes of exposure implications of metabolism. Equivocal results may indicate chemical has multiple MOA.
M	–	–	+	No evidence for (anti)-E,A,T,S activity in OECD TG 410. Weak (anti)-E,A,S activity may not be detected by this assay. Effects seen in existing studies are via non-E,A,T,S mechanism.	Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	If existing data are from an adequate Level 5 assay, question why there are differences. Effects seen in existing studies may be in a more sensitive life stage. Consider route of exposures and possible implications of ADME characteristics of the chemical.
N	–	–	–	No evidence for (anti)-E,A,T,S activity in OECD TG 410. Weak (anti)-E,A,S activity may not be detected by this assay. No evidence for (anti)-E,A,T,S activity *in vitro*. No evidence of adverse effects.	Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive).
O	–	–	Eq/0	No evidence for (anti)-E,A,T,S activity in OECD TG 411. Weak (anti)-E,A,S activity may not be detected by this assay. No evidence for (anti)-E,A,T,S activity *in vitro*. Unknown potential for adverse effects.	Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	Consider route of exposures and possible implications for ADME characteristics of the chemical in follow-up assay.

Scenarios	Result of OECD TG 410 (28-day dermal assay)	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
P	–	Eq/0	+	No evidence for (anti)-E,A,T,S activity in OECD TG 410. Weak (anti)-E,A,S activity may not be detected by this assay. Potential for adverse effects via unknown mechanism.	Perform *in vitro* ER, AR,TR, S assays with added metabolising system.	Consider route of exposure for OECD TG 411 assay and possible implications for differences from existing assay. Effects seen in existing studies may be in a more sensitive life stage. Further mechanistic studies may strengthen weight of evidence. Equivocal results may indicate chemical has multiple MOA.
Q	–	Eq/0	–	No evidence for (anti)-E,A,T,S activity in OECD TG 410. Weak (anti)-E,A,S activity may not be detected by this assay. No evidence of adverse effects.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system.	If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). Further mechanistic studies may strengthen weight of evidence.
R	–	Eq/0	Eq/0	No evidence for (anti)-E,A,T,S activity in OECD TG 411. Weak (anti)-E,A,S activity may not be detected by this assay. Unknown potential for adverse effects.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system, otherwise Eq result available.	Further mechanistic studies may strengthen weight of evidence. Equivocal results may indicate chemical has multiple MOA. Check data on chemical analogues.

C.3.10. Subchronic Dermal Toxicity: 90-Day Study (OECD TG 411)

Status: Assay validated by the OECD.

861. Modalities detected: (anti)estrogen, (anti)androgen, thyroid, steroidogenesis.

Endpoints: Weight of adrenals, testes.

Histopathologic changes in pituitary, thyroid gland, gonads, accessory sex organs, female mammary gland and adrenals.

Background to the assay

862. This assay determines the subchronic dermal toxicity of chemicals after initial information on toxicity has been obtained by acute testing. It provides information on possible health hazards likely to arise from repeated exposure by the dermal route over a limited period of time. Dosing duration is 90 days and the preferred species are the adult rat, rabbit or guinea pig. Test substance is applied to the dorsal area of the trunk, held in place with a dressing and protected from ingestion. OECD TG 411 was adopted in May 1981. Although it has not been validated for the detection of endocrine active substances (EASs), this assay contains several endpoints that are suitable for the determination of endocrine effects. As all the endpoints are apical, it is difficult to discern mechanism of action from this test alone. Information on mechanism of action needs to be obtained from *in vitro* estrogen/androgen/thyroid/steroidogenesis (E,A,T,S) assays or *in vivo* lower tier tests such as the Uterotrophic Bioassay (UT – OECD TG 440) and the Hershberger Bioassay (H – OECD TG 441). Hormone measurements are **not** included in this assay.

863. A comparison can be made with OECD TG 408 (90-day oral toxicity study) for endocrine endpoints, although dermal absorption of test substances is likely to result in lower internal doses compared to oral administration. Nevertheless, the number of animals per group (ten male and ten female) and the duration of dosing means that it is a relevant assay to assess when determining potential endocrine activity. Dermal exposure may be a relevant route of human exposure to certain substances.

When/why the assay may be used

864. This assay may be used as part of a chemical submission package and forms part of the standard information requirements in certain chemical legislations. At least three dose levels are included so that an estimate of no-adverse-effect-level can be determined and the assay used for hazard identification/characterisation. It should be noted that as this assay is not primarily designed to detect endocrine disruption, a higher degree of systemic toxicity is typically induced than is the case with the other Level 3 and 4 assays. The possibly confounding effect of systemic toxicity on endocrine endpoints therefore needs to be considered.

865. In order to provide information relevant for assessing whether or not a chemical may fulfil the WHO/IPCS (2002) definition of an endocrine disruptor (ED), the study design has to be sufficiently robust to demonstrate the presence or absence of effects. In the dose selection, the investigator should also consider and ensure that data generated are adequate to fulfil the regulatory requirement across OECD countries as appropriate (e.g. hazard and risk assessment and labelling, ED assessment, etc.). The top dose or concentration should be sufficiently high to give clear systemic (i.e. non endocrine-specific) toxicity in order to ensure that a wide range of exposures (high to low) is tested. However, endocrine effects observed solely in the presence of clear systemic toxicity should be interpreted with caution and may be disregarded when sufficiently justified to be caused by secondary effects which are unlikely to be due to endocrine activity. The reason for this advice is a concern that some endocrine active substance (EAS) sensitive assays are being run at doses/concentrations of EASs that are too low to trigger direct impacts on the endocrine system. This guidance document is not the place to address this issue directly, but it should be considered when EAS-sensitive test guidelines (TGs) are revised in the future. In addition, the number and spacing of dose/concentration levels should also be adequate to fulfil the objectives of the study (e.g. to demonstrate dose response relationships if this is required).

Introduction to the table of scenarios

866. Table C.3.10 gives guidance on a further step to take in the event of a positive (+) or negative (-) result and in the presence of positive (+), negative (-) or equivocal/absent (Eq/0) existing results. “Existing results” are subdivided into “mechanism” and “effects” data (third and fourth columns). The table is divided horizontally into a series of scenarios that represent all the combinations of these events.

867. The results of OECD TG 411 are given in the second column. As OECD TG 411 is not a screening test where a yes/no (qualitative) answer is obtained for the test as a whole, positive results would generally be assessed for individual endpoints. For the purposes of this guidance, however, a positive result is defined as a biologically significant change in any of the endocrine endpoints listed above (e.g. statistically significant reductions in reproductive organ weights). Changes in related endpoints will increase their biological significance (e.g. changes in the weights of testes and epididymides accompanied by histopathological changes). The guidance on histopathologic changes in endocrine tests (OECD, 2009) may be helpful in interpretation. A negative result for OECD TG 411 is taken to be the absence of biologically significant changes in all endocrine endpoints.

868. In the absence of other pertinent lines of evidence, negative results in this test alone cannot be taken as evidence that the substance is not an ED. Further studies will be required as confirmation.

869. Equivocal results for the guideline are not considered in the table, partly for brevity but also because equivocal results are by nature uncertain. A decision must eventually be reached about whether the endocrine endpoints tend to be positive or negative or whether the result must be put to one side and the test repeated (using the same or a different test guideline). Factors which may have interfered with the result (e.g. composition of the diet used, environmental influences) should be considered.

Existing data to be considered

870. Existing "mechanism" *in vitro* data are assumed to be available from estrogen receptor (ER-), androgen receptor (AR-) and steroidogenesis-based assays (Level 2). Assays may also be available for interference with thyroid modalities. In practice, it is possible that data from all of these assays may not be available, so judgement will need to be used to decide which assays to perform. Although the current *in vitro* test guidelines do not incorporate metabolic activation, published information on use of metabolic activation systems is available in Jacobs et al. (2008; 2013) and OECD (2008). These methods, however, have not yet been validated.

871. Existing "effects" data refer to *in vivo* effects that may come from Level 3 or 4 tests in the Conceptual Framework (e.g. the UT or H assays). In these cases, it should be remembered that these assays are specifically designed to be sensitive to EASs. As mentioned above, the results of the study may be interpreted as part of a battery or group of tests carried out for regulatory purposes. Data may also be available on effects in mammalian and non-mammalian wildlife species, although caution should be used when extrapolating between taxa. A chemical causing endocrine effects in non-mammalian environmental species (fish, for example) may also have endocrine effects in mammals, but the physiological consequences of the effects are likely to be different.

872. When considering the results of the OECD TG 411 assay, all available data should be used in order to reach a conclusion and a weight of evidence approach taken. This may include high throughput screening data, read-across data from structural analogues and quantitative structure activity relationship (QSAR). Several QSAR models for ER and AR binding/activation are now available (see Sections B.1.1.1 and B.1.1.2).

Scenarios: Positive and negative results combined with existing data

873. A series of scenarios (A to R) are presented in Table C.3.10 and represent all the possibilities of positive or negative results in combination with the presence or absence of existing data. The action taken will also depend on the regulatory environment, but the considerations given here are generally science based. Although OECD TG 411 assay uses mammals, the well-conserved nature of the hormonal pathways across taxa indicate that results on endocrine endpoints in this assay may be relevant to other vertebrate species. Effects in laboratory mammal tests are also highly relevant for environmental mammalian species. Wherever possible, the recommended "next step which could be taken" avoids unnecessary animal testing. However, sometimes conducting an animal test will be indicated and then the relevance of species, strain and exposure route should always be considered. The sensitivity and physiological function of the hormone under investigation in the test species should also be considered. In general, lower level tests should be conducted before higher level tests in order to avoid unnecessary animal usage, unless it is apparent that a Level 5 test will be required anyway or will be needed to establish the evidence to conclude on ED properties. Information on some endocrine-related tumours may be detected more comprehensively in carcinogenicity studies (OECD TG 451/453) (Level 4); for example, detection of certain types of thyroid tumors in the absence of reproductive or developmental effects, as well as substances causing tumors in other endocrine-sensitive tissues. It is recognised, however, that some jurisdictions may require a two-generation study. Further considerations specific to each scenario are given in the table.

874. Scenarios A to C represent positive results in the OECD TG 411 assay in the presence of positive *in vitro* mechanistic data and positive, negative or equivocal *in vivo*

effects data. A positive result in the *in vitro* assays in combination with a positive OECD TG 411 assay is moderate or strong evidence for E,A,T,S-mediated activity that may or may not be supported by the *in vivo* effects data. In the absence of robust upper-level data, the next step may be to conduct an upper-level test. In the presence of robust *in vivo* data, there may be sufficient evidence to conclude concern for endocrine disruption and therefore no need for further testing. Positive results in the OECD TG 411 assay may also indicate the potential for endocrine mediated effects in lower vertebrates. These could be followed up with partial life cycle tests such as the Fish Sexual Development Test (FSDT), the Larval Amphibian Growth and Development Assay (LAGDA) or the Medaka Extended One-Generation Reproduction Test (MEOGRT) if the evidence were strong enough. *In vivo* assays/tests with negative results should be interpreted with caution as they may either indicate that the tests used do not have sufficient power to detect weak effects or, alternatively, that the effects do not present a concern for endocrine disruption. The possibility of other (non-E,A,T,S) mechanisms should also not be overlooked (e.g. involving other receptors or endocrine axes).

875. Scenarios D to F represent positive results in the OECD TG 411 assay in the presence of negative *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. Negative results in the *in vitro* assays should be viewed with caution in case a metabolite is responsible for the positive OECD TG 411 assay. Unless the metabolic profile of the test substance is known, one option may be to conduct these assays with an added metabolising system. If the metabolic profile is known, then a higher level *in vivo* test may be advisable. The choice of tests will depend on the available *in vivo* effects data. Positive results in the OECD TG 411 assay may also indicate the potential for endocrine mediated effects in lower vertebrates. As in Scenarios A to C, *in vivo* assays/tests with negative results should be interpreted with caution as they may either indicate that the tests used do not have sufficient power to detect weak effects or, alternatively, that the effects do not present a concern for endocrine disruption.

876. Scenarios G to I represent positive results in the OECD TG 411 assay in the presence of various combinations of missing or equivocal data. Positive results in the OECD TG 411 assay may also indicate the potential for endocrine mediated effects in lower vertebrates. The next step to take in these eventualities will depend on the nature of the other available data and the jurisdiction in which it is being used. In some cases, equivocal data may be viewed as positive whilst in others it may or may not contribute to the weight of evidence. The interpretation may also depend on the mode of action (MOA) in question and why the data are considered equivocal, e.g. a study that is equivocal for thyroid effects may still be of value in evaluating (anti)androgenic effects. In all three scenarios, the recommended first step is to obtain reliable mechanistic (*in vitro*) data rather than proceed further with *in vivo* testing. Equivocal and missing data are alternative scenarios and two possibilities for the next step are given in most cases, but the nature of equivocal data means that decisions need to be taken on a case-by-case basis. In all cases, the role of metabolism, route of exposure and data from structural analogues should be considered before deciding on the next step.

877. Scenarios J to L represent negative results in the OECD TG 411 assay in the presence of positive *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. Negative outcomes in the OECD TG 411 should be viewed with caution because of the power of the assay to detect (anti)estrogens and androgens may be limited. All three scenarios could also arise from a chemical that is positive in *in vitro* assays, but is metabolised to a non-active metabolite leading to negative results in the OECD TG 411 assay. This should be considered first when investigating the next step. Endocrine active

potency may also explain differences between *in vitro* and *in vivo* results (e.g. a chemical with weak endocrine activity may give a positive result *in vitro* but may be negative *in vivo*). Positive *in vivo* effects data may involve other E,A,T,S, non-E,A,T,S mechanisms (e.g. involving other receptors or endocrine axes), more sensitive endpoints, greater statistical power or life stages that are more sensitive to the substance than the adult dermally exposed animals in OECD TG 411.

878. Scenarios M to O represent negative results in the OECD TG 411 assay in the presence of negative *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. Negative results for all tests (Scenario N) may be sufficient to enable a conclusion of no concern for endocrine disruption. This will depend on the weight of evidence and may not be possible. Where there are positive *in vivo* effects data, there could still be an E,A,T,S-related mechanism, the effects may be related to length of exposure, route of exposure or exposure at different life stages. Other E,A,T,S or non-E,A,T,S mechanisms may also be involved.

879. Scenarios P to R represent negative results in the OECD TG 411 assay in the presence of various combinations of missing or equivocal data. As with the positive result scenarios above (see Paragraph 877) the next step to take in these eventualities will have to be decided on a case-by-case basis. However, the recommended first step is generally to obtain reliable mechanistic (*in vitro*) data rather than proceed further with *in vivo* testing. In all cases, the role of metabolism, route of exposure and data from structural analogues should be considered before deciding on the next step.

880. In all scenarios (A to R), the next step to take to strengthen weight of evidence will depend on the existing information. The table is meant to provide a succinct guide and may not cover all circumstances or possibilities. The scenarios may also suggest that chemicals have simple or single MOA, when in practice they may have multiple endocrine and non-endocrine MOA. In some cases, for example, two opposite modes of simultaneous action (e.g. estrogenic and anti-estrogenic) could, depending on dose, lead to a minimisation or abolition of effects, while in others two different MOA (e.g. estrogenic and anti-androgenic) could potentially reinforce effects. Endocrine pathways interact, mixed effects are common and there are many pathways that cannot be distinguished with currently available TGs. If multiple MOA are suspected, either from the existing results or based on QSAR/read-across/integrated approaches, this should be investigated further if needed for regulatory decision making.

References

Jacobs, M. et al. (2013), "*In vitro* metabolism and bioavailability tests for endocrine active substances: What is needed next for regulatory purposes?", *ALTEX – Alternatives to Animal Experimentation*, Vol. 30/3, pp. 331-351.

Jacobs, M.N. et al. (2008), "The use of metabolising systems for *in vitro* testing of endocrine disrupters", *Current Drug Metabolism*, Vol. 9/8, pp. 796-826.

OECD (2009), "Guidance document for histologic evaluation of endocrine and reproductive tests in rodents", OECD Series on Testing and Assessment, No. 106, OECD, Paris, www.oecd.org/env/ehs/testing/43411534.pdf.

OECD (2008), *Detailed Review Paper on the Use of Metabolising Systems for* In Vitro *Testing of Endocrine Disrupters*, OECD Series on Testing and Assessment, No. 97, OECD Publishing, Paris, https://doi.org/10.1787/9789264085497-en.

WHO/IPCS (2002), "Global assessment of the state-of-the-science of endocrine disrupters", Damstra, T. et al. (eds.) WHO/PCS/EDC/02.2, World Health Organization, Geneva, www.who.int/ipcs/publications/new_issues/endocrine_disruptors/en.

Table C.3.10. **Subchronic Dermal Toxicity: 90-Day Study (OECD TG 411): Guidance for scenarios of combinations of results with existing data**

This table represents possible conclusions to be drawn from assay data, and a next step which could be taken if further evidence is required about possible endocrine disrupting properties and/or effects. The guidance offered is not meant to be prescriptive, but provides science-based considerations. It encourages the use of all available data and expert judgement in a weight of evidence approach. Regional and national interpretation of results and "next steps" may vary.

The conclusions are grouped into a series of scenarios (A-R), each scenario representing a different combination of assay results, existing *in vitro* data and existing *in vivo* data. The symbol "+" indicates that the data in question represent a positive result, «-" indicates a negative result, and "Eq/0" indicates that the data are either equivocal or are not available.

Existing results: * "Mechanism (*in vitro* mechanistic data)" assumes that mechanistic data are available from estrogen receptor (ER-), androgen receptor (AR-) and steroidogenesis-based assays (Level 2). Thyroid hormone receptor (TR) and other assays concerning mechanisms of thyroid disruption may be available, but they are not in common use. In practice, data from all assays may not be available and therefore this must be taken into account when deciding on the "next step". Quantitative structure activity relationship (QSAR) predictions of estrogen and androgen binding/activation may be made for some substances.

Existing results: ** "Effects (*in vivo* effects of concern)" assumes various information, such as data from repeat dose oral toxicity studies, reproduction/developmental toxicity screen tests, read-across from analogues, will be available.

*** *Note:* a positive result is defined as a biologically significant change in any of the endocrine endpoints.

Scenarios	Result of OECD TG 411 (90-day dermal assay)	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
A	+ ***	+	+	(Anti)-E,A,T,S activity. Increased evidence of (anti)-E,A,T,S activity.	Perform assay from Level 5 (e.g. Extended One-Generation Reproduction Toxicity Study [EOGRTS] or two-generation assay).	If existing data are from a Level 5 assay, there may be sufficient information to conclude evidence of concern for endocrine disruption (the EOGRTS provides the most information; however, for endocrine disrupting chemicals [EDCs] with a carcinogenic potential, OECD TG 451-3 may be more sensitive). Endocrine activity possible in lower vertebrates. Consider performing a Fish Sexual Development Test (FSDT), Larval Amphibian Growth and Development Assay (LAGDA) or Medaka Extended One-Generation Reproduction Test (MEOGRT).
B	+	+	–	(Anti)-E,A,T,S activity. Increased evidence of (anti)-E,A,T,S activity.	Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	If existing data are from an adequate Level 5 assay, question why there are differences. If existing data are from a less sensitive assay, a higher level test may be required. Consider route of exposures and possible implications of absorption, distribution, metabolism and excretion (ADME) characteristics of the chemical. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT.
C	+	+	Eq/0	(Anti)-E,A,T,S activity. Increased evidence of (anti)-E,A,T,S activity.	Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	Check data on chemical analogues. Consider route of exposures and possible implications of ADME characteristics of the chemical. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT. Equivocal results may indicate chemical has multiple modes of action (MOA).
D	+	–	+	(Anti)-E,A,T,S activity. Acts via non-estrogen receptor (ER-), androgen receptor (AR-), thyroid hormone receptor (TR), steroidogenesis (S) mechanism or may require metabolic activation for activity.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system.	If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude evidence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). Consider route of exposures for effects data and possible implications of ADME characteristics of the chemical. Further mechanistic studies may help determine MOA. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT.
E	+	–	–	(Anti)-E,A,T,S activity. Acts via non-ER, AR, TR, S mechanism or may require metabolic activation for activity. Route of exposure may account for the differences between OECD TG 411 and existing data.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system OR Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	If existing data are from an adequate Level 5 assay, question why there are differences. If existing data are from a less sensitive assay, a higher level test may be required. Consider route of exposures and possible implications of ADME characteristics of the chemical. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT.

Scenarios	Result of OECD TG 411 (90-day dermal assay)	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
F	+	–	Eq/0	(Anti)-E,A,T,S activity. Acts via non-ER, AR, TR, S mechanism or may require metabolic activation for activity.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system OR Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	Check data on chemical analogues. Further mechanistic studies may help determine MOA. If existing data are from an adequate Level 5 assay, question why there are differences. If existing data are from a less sensitive assay, a higher level test may be required Equivocal results may indicate chemical has multiple MOA. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT.
G	+	Eq/0	+	(Anti)-E,A,T,S activity. May act via ER, AR, TR, S mechanism (metabolic activation may be needed).	Perform *in vitro* ER, AR, TR, S assays (for the "0" scenario, otherwise Eq result available) OR Perform *in vitro* ER, AR, TR, S assays with added metabolising system.	If existing data are from a Level 5 assay, there may be sufficient information to conclude evidence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). Check data on chemical analogues. Further mechanistic studies may help determine MOA. Equivocal results may indicate chemical has multiple MOA. Consider route of exposures for effects data and possible implications of ADME characteristics of the chemical. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT.
H	+	Eq/0	–	(Anti)-E,A,T,S activity. Acts via unknown mechanism or may require metabolic activation for activity. Route of exposure may account for the differences between OECD TG 411 and existing data. Unknown potential for adverse effects.	For the "0" scenario, perform *in vitro* ER, AR, TR, S assays, maybe with added metabolising system (otherwise Eq result available).	If existing data are from an adequate Level 5 assay, question why there are differences. Consider route of exposures and possible implications of ADME characteristics of the chemical. If existing data are from a less sensitive assay, a higher level test may be required. Check data on chemical analogues. Further mechanistic studies may help determine MOA. Equivocal results may indicate chemical has multiple MOA. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT.
I	+	Eq/0	Eq/0	(Anti)-E,A,T,S activity. Acts via unknown mechanism. Unknown potential for adverse effects.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system OR Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	Check data on chemical analogues. Further mechanistic studies may help determine MOA. Equivocal results may indicate chemical has multiple MOA. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT.

Scenarios	Result of OECD TG 411 (90-day dermal assay)	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
J	–	+	+	No evidence for (anti)-E,A,T,S activity in OECD TG 411. Weak (anti)-E,A,S activity may not be detected by this assay. Metabolism or potency may explain the difference from existing *in vitro* and *in vivo* data.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system OR Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	If existing data are from an adequate Level 5 assay, question why there are differences. Effects seen in existing studies may be in a more sensitive life stage. Consider route of exposures and possible implications of ADME characteristics of the chemical. Further mechanistic studies may help determine MOA.
K	–	+	–	No evidence for (anti)-E,A,T,S activity in OECD TG 411. Weak (anti)-E,A,S activity may not be detected by this assay. Metabolism or potency may explain *in vitro*/*in vivo* differences.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system OR Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from a less sensitive assay, a higher level test may be required. Further mechanistic studies with metabolism may help determine MOA.
L	–	+	Eq/0	No evidence for (anti)-E,A,T,S activity in OECD TG 411. Weak (anti)-E,A,S activity may not be detected by this assay. Metabolism or potency may explain *in vitro*/*in vivo* differences. Unknown potential for adverse effects.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system OR Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	Metabolic deactivation of chemical may occur *in vivo* so that possible *in vitro* activity is not realised. Consider possible routes of exposure implications of metabolism. Equivocal results may indicate chemical has multiple MOA.
M	–	–	+	No evidence for (anti)-E,A,T,S activity in OECD TG 411. Weak (anti)-E,A,S activity may not be detected by this assay. Effects seen in existing studies are via non-E,A,T,S mechanism.	Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	If existing data are from an adequate Level 5 assay, question why there are differences. Effects seen in existing studies may be in a more sensitive life stage. Consider route of exposures and possible implications of ADME characteristics of the chemical.
N	–	–	–	No evidence for (anti)-E,A,T,S activity in OECD TG 411. Weak (anti)-E,A,S activity may not be detected by this assay. No evidence for (anti)-E,A,T,S activity *in vitro*. No evidence of adverse effects.	Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive).

Scenarios	Result of OECD TG 411 (90-day dermal assay)	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
O	–	–	Eq/0	No evidence for (anti)-E,A,T,S activity in OECD TG 411. Weak (anti)-E,A,S activity may not be detected by this assay. No evidence for (anti)-E,A,T,S activity *in vitro*. Unknown potential for adverse effects.	Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	Consider route of exposures and possible implications for ADME characteristics of the chemical in follow-up assay.
P	–	Eq/0	+	No evidence for (anti)-E,A,T,S activity in OECD TG 411. Weak (anti)-E,A,S activity may not be detected by this assay. Potential for adverse effects via unknown mechanism.	Perform *in vitro* ER, AR,TR, S assays with added metabolising system.	Consider route of exposure for OECD TG 411 assay and possible implications for differences from existing assay. Effects seen in existing studies may be in a more sensitive life stage. Further mechanistic studies may strengthen weight of evidence. Equivocal results may indicate chemical has multiple MOA.
Q	–	Eq/0	–	No evidence for (anti)-E,A,T,S activity in OECD TG 411. Weak (anti)-E,A,S activity may not be detected by this assay. No evidence of adverse effects.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system.	If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). Further mechanistic studies may strengthen weight of evidence.
R	–	Eq/0	Eq/0	No evidence for (anti)-E,A,T,S activity in OECD TG 411. Weak (anti)-E,A,S activity may not be detected by this assay. Unknown potential for adverse effects.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system, otherwise Eq result available.	Further mechanistic studies may strengthen weight of evidence. Equivocal results may indicate chemical has multiple MOA. Check data on chemical analogues.

C.3.11. 28-Day (Subacute) Inhalation Toxicity Study (OECD TG 412)

Status: Assay validated by the OECD.

881. Modalities detected: (anti)estrogen, (anti)androgen, thyroid, steroidogenesis.

Endpoints: Weight of adrenals, epididymides, ovaries, testes, thyroid, uterus.

Histopathologic changes in adrenals, epididymides, female mammary gland, ovaries, pituitary, prostate, seminal vesicles, testes, thyroid, uterus.

Background to the assay

882. This assay determines the subchronic inhalation toxicity of chemicals to provide data for quantitative inhalational hazard identification/characterisation. It is primarily used to derive regulatory concentrations for assessing worker risk in occupational settings. It may also be used to assess human residential, transportation and environmental risk. Dosing duration is 28 days and the preferred species is the rat. Animals are exposed to test substance in inhalation chambers for six hours per day on a five day per week basis. The preferred mode of exposure is nose only, but whole body exposure may also be used. OECD TG 412 was first adopted in 1981, revised in 2009 and 2017. Although it has not been validated for the detection of endocrine active substances (EASs), this assay contains several endpoints that are suitable for the determination of endocrine effects. As all the endpoints are apical, it is difficult to discern mechanism of action from this test alone. Information on mechanism of action may need to be obtained from *in vitro* estrogen/androgen/thyroid/steroidogenesis (E,A,T,S) assays or *in vivo* lower tier tests such as the Uterotrophic Bioassay (UT – OECD TG 440) and the Hershberger Bioassay (H – OECD TG 441). Hormone measurements are **not** included in this assay.

883. A comparison can be made with OECD TG 407 (28-Day Oral Toxicity Study) for endocrine endpoints, although inhalational absorption of test substances is likely to result in lower internal doses compared to oral administration. Nevertheless, the duration of dosing means that it is a relevant assay to assess when determining potential endocrine activity. Inhalation exposure may be a relevant route of human exposure to certain substances.

When/why the assay may be used

884. This assay may be used as part of a pesticide submission package and may form part of the standard information requirements in certain chemical legislations (e.g. REACH). At least three dose levels are included so that an estimate of no-adverse-effect-level can be determined and the assay used for hazard identification/characterisation. It should be noted that as this assay is not primarily designed to detect endocrine disruption, a higher degree of systemic toxicity is typically induced than is the case with the other Level 3 and 4 assays. The possibly confounding effect of systemic toxicity on endocrine endpoints therefore needs to be considered.

885. In order to provide information relevant for assessing whether or not a chemical may fulfil the WHO/IPCS (2002) definition of an endocrine disruptor (ED), the study design has to be sufficiently robust to demonstrate the presence or absence of effects. In the dose selection, the investigator should also consider and ensure that data generated are adequate to fulfil the regulatory requirement across OECD countries as appropriate (e.g. hazard and risk assessment and labelling, ED assessment, etc.). The top dose or concentration should be sufficiently high to give clear systemic (i.e. non endocrine-specific) toxicity in order to ensure that a wide range of exposures (high to low) is tested. However, endocrine effects observed solely in the presence of clear systemic toxicity should be interpreted with caution and may be disregarded when sufficiently justified to be caused by secondary effects which are unlikely to be due to endocrine activity. The reason for this advice is a concern that some endocrine active substance (EAS) sensitive assays are being run at doses/concentrations of EASs that are too low to trigger direct impacts on the endocrine system. This guidance document is not the place to address this issue directly, but it should be considered when EAS-sensitive test guidelines (TGs) are revised in the future. In addition, the number and spacing of dose/concentration levels should also be adequate to fulfil the objectives of the study (e.g. to demonstrate dose response relationships if this is required).

Introduction to the table of scenarios

886. Table C.3.11 gives guidance on a further step to take in the event of a positive (+) or negative (-) result and in the presence of positive (+), negative (-) or equivocal/absent (Eq/0) existing results. "Existing results" are subdivided into "mechanism" and "effects" data (third and fourth columns). The table is divided horizontally into a series of scenarios that represent all the combinations of these events.

887. The results of OECD TG 412 are given in the second column. As OECD TG 412 is not a screening test where a yes/no (qualitative) answer is obtained for the test as a whole, positive results would generally be assessed for individual endpoints. For the purposes of this guidance, however, a positive result is defined as a biologically significant change in any of the endocrine endpoints listed above (e.g. biologically significant reductions in reproductive organ weights). Changes in related endpoints will increase their biological significance (e.g. changes in the weights of testes and epididymides accompanied by histopathological changes). The guidance on histopathologic changes in endocrine tests (OECD, 2009) may be helpful in interpretation. A negative result for the OECD TG 412 is taken to be the absence of biologically significant changes in all endocrine endpoints.

888. In the absence of other pertinent lines of evidence, negative results in this test alone cannot be taken as evidence that the substance is not an ED. Further studies will be required as confirmation.

889. Equivocal results for the guideline are not considered in the table, partly for brevity but also because equivocal results are by nature uncertain. A decision must eventually be reached about whether the endocrine endpoints tend to be positive or negative or whether the result must be put to one side and the test repeated (using the same or a different test guideline). Factors which may have interfered with the result (e.g. composition of the diet used, environmental influences) should be considered.

Existing data to be considered

890. Existing "mechanism" *in vitro* data are assumed to be available from estrogen receptor (ER-), androgen receptor (AR-) and steroidogenesis-based assays (Level 2).

Assays may also be available for interference with thyroid modalities. In practice, it is possible that data from all of these assays may not be available, so judgement will need to be used to decide which assays to perform. Although the current *in vitro* test guidelines do not incorporate metabolic activation, published information on use of metabolic activation systems is available in Jacobs et al. (2008; 2013) and OECD (2008). These methods, however, have not yet been validated.

891. Existing "effects" data refer to *in vivo* effects that may come from Level 3 or 4 tests in the Conceptual Framework (e.g. UT or H assays). In these cases, it should be remembered that these assays are specifically designed to be sensitive to EASs. As mentioned above, the results of the study may be interpreted as part of a battery or group of tests carried out for regulatory purposes. Data may also be available on effects in mammalian and non-mammalian wildlife species, although caution should be used when extrapolating between taxa. A chemical causing endocrine effects in non-mammalian environmental species (fish, for example) may also have endocrine effects in mammals, but the physiological consequences of the effects are likely to be different.

892. When considering the results of the OECD TG 412 assay, all available data should be used in order to reach a conclusion and a weight of evidence approach taken. This may include high throughput screening data, read-across data from structural analogues and quantitative structure activity relationship (QSAR). Several QSAR models for ER and AR binding/activation are now available (see Sections B.1.1.1 and B.1.1.2).

Scenarios: Positive and negative results combined with existing data

893. A series of scenarios (A to R) are presented in Table C.3.11 and represent all the possibilities of positive or negative results in combination with the presence or absence of existing data. The action taken will also depend on the regulatory environment, but the considerations given here are generally science based. Although OECD TG 412 assay uses mammals, the well-conserved nature of the hormonal pathways across taxa indicate that results on endocrine endpoints in this assay may be relevant to other vertebrate species. Effects in laboratory mammal tests are also highly relevant for environmental mammalian species. Wherever possible, the recommended "next step which could be taken" avoids unnecessary animal testing. However, sometimes conducting an animal test will be indicated and then the relevance of species, strain and exposure route should always be considered. The sensitivity and physiological function of the hormone under investigation in the test species should also be considered. In general, Level 3 tests should be conducted before Level 5 tests in order to avoid unnecessary animal usage, unless it is apparent that a Level 5 test will be required anyway or will be needed to establish the evidence to conclude on ED properties. Information on some endocrine-related tumours may be detected more comprehensively in carcinogenicity studies (OECD TG 451/453) (Level 4); for example, detection of certain types of thyroid tumors in the absence of reproductive or developmental effects, as well as substances causing tumors in other endocrine-sensitive tissues. At Level 5, the Extended One-Generation Reproduction Toxiticy Study (EOGRTS – OECD TG 443) is the most sensitive reproduction assay for detecting endocrine disruption because it includes evaluation of a number of endocrine endpoints not included in the two-generation study (OECD TG 416) adopted in 2001. It is recognised, however, that some jurisdictions may require a two-generation study. Further considerations specific to each scenario are given in the table.

894. Scenarios A to C represent positive results in the OECD TG 412 assay in the presence of positive *in vitro* mechanistic data and positive, negative or equivocal *in vivo*

effects data. A positive result in the *in vitro* assays in combination with a positive OECD TG 412 assay is moderate or strong evidence for E,A,T,S-mediated activity that may or may not be supported by the *in vivo* effects data. In the absence of robust upper-level data, the next step may be to conduct an upper-level test. In the presence of robust *in vivo* data, there may be sufficient evidence to conclude concern for endocrine disruption and therefore no need for further testing. Positive results in the OECD TG 412 assay may also indicate the potential for endocrine mediated effects in lower vertebrates. These could be followed up with partial life cycle tests such as the Fish Sexual Development Test (FSDT), the Larval Amphibian Growth and Development Assay (LAGDA) or the Medaka Extended One-Generation Reproduction Test (MEOGRT) if the evidence were strong enough. *In vivo* assays/tests with negative results should be interpreted with caution as they may either indicate that the tests used do not have sufficient power to detect weak effects or, alternatively, that the effects do not present a concern for endocrine disruption. The possibility of other (non-E,A,T,S) mechanisms should also not be overlooked (e.g. involving other receptors or endocrine axes).

895. Scenarios D to F represent positive results in the OECD TG 412 assay in the presence of negative *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. Negative results in the *in vitro* assays should be viewed with caution in case a metabolite is responsible for the positive OECD TG 412 assay. Unless the metabolic profile of the test substance is known, one option may be to conduct these assays with an added metabolising system. If the metabolic profile is known, then a higher level *in vivo* test may be advisable. The choice of tests will depend on the available *in vivo* effects data. Positive results in the OECD TG 412 assay may also indicate the potential for endocrine mediated effects in lower vertebrates. As in Scenarios A to C, *in vivo* assays/tests with negative results should be interpreted with caution as they may either indicate that the tests used do not have sufficient power to detect weak effects or, alternatively, that the effects do not present a concern for endocrine disruption.

896. Scenarios G to I represent positive results in the OECD TG 412 assay in the presence of various combinations of missing or equivocal data. Positive results in the OECD TG 412 assay may also indicate the potential for endocrine mediated effects in lower vertebrates. The next step to take in these eventualities will depend on the nature of the other available data and the jurisdiction in which it is being used. In some cases, equivocal data may be viewed as positive whilst in others it may or may not contribute to the weight of evidence. The interpretation may also depend on the mode of action (MOA) in question and why the data are considered equivocal, e.g. a study that is equivocal for thyroid effects may still be of value in evaluating (anti)androgenic effects. In all three scenarios, the recommended first step is to obtain reliable mechanistic (*in vitro*) data rather than proceed further with *in vivo* testing. Equivocal and missing data are alternative scenarios and two possibilities for the next step are given in most cases, but the nature of equivocal data means that decisions need to be taken on a case-by-case basis. In all cases, the role of metabolism, route of exposure and data from structural analogues should be considered before deciding on the next step.

897. Scenarios J to L represent negative results in the OECD TG 412 assay in the presence of positive *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. Negative outcomes in the OECD TG 412 should be viewed with caution because of the power of the assay to detect (anti)estrogens and androgens may be limited All three scenarios could also arise from a chemical that is positive in *in vitro* assays, but is metabolised to a non-active metabolite leading to negative results in the OECD TG 412 assay. This should be considered first when investigating the next step. Endocrine active potency may also explain differences between *in vitro* and *in vivo* results (e.g. a weak chemical may give a positive result *in vitro* but may be negative *in vivo*). Positive *in vivo* effects data may involve other E,A,T,S, non-E,A,T,S mechanisms (e.g. involving other receptors or endocrine axes), more sensitive endpoints, greater statistical power or life stages that are more sensitive to the substance than the adult inhalationally exposed animals in OECD TG 412.

898. Scenarios M to O represent negative results in the OECD TG 412 assay in the presence of negative *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. Negative results for all tests (Scenario N) may be sufficient to enable a conclusion of no concern for endocrine disruption. This will depend on the weight of evidence and may not be possible. Where there are positive *in vivo* effects data there could still be an E,A,T,S-related mechanism, the effects may be related to length of exposure, route of exposure or exposure at different life stages. Other E,A,T,S or non-E,A,T,S mechanisms may also be involved.

899. Scenarios P to R represent negative results in the OECD TG 412 assay in the presence of various combinations of missing or equivocal data. As with the positive result scenarios above, the next step to take in these eventualities will have to be decided on a case-by-case basis. However, the recommended first step is generally to obtain reliable mechanistic (*in vitro*) data rather than proceed further with *in vivo* testing. In all cases, the role of metabolism, route of exposure and data from structural analogues should be considered before deciding on the next step.

900. In all scenarios (A to R), the next step to take to strengthen weight of evidence will depend on the existing information. The table is meant to provide a succinct guide and may not cover all circumstances or possibilities. The scenarios may also suggest that chemicals have simple or single MOA, when in practice they may have multiple endocrine and non-endocrine MOA. In some cases, for example, two opposite modes of simultaneous action (e.g. estrogenic and anti-estrogenic) could, depending on dose, lead to a minimisation or abolition of effects, while in others two different MOA (e.g. estrogenic and anti-androgenic) could potentially reinforce effects. Endocrine pathways interact, mixed effects are common and there are many pathways that cannot be distinguished with currently available TGs. If multiple MOA are suspected, either from the existing results or based on QSAR/read-across/integrated approaches, this should be investigated further if needed for regulatory decision making.

References

Jacobs, M. et al. (2013), "*In vitro* metabolism and bioavailability tests for endocrine active substances: What is needed next for regulatory purposes?", *ALTEX – Alternatives to Animal Experimentation*, Vol. 30/3, pp. 331-351.

Jacobs, M.N. et al. (2008), "The use of metabolising systems for *in vitro* testing of endocrine disrupters", *Current Drug Metabolism*, Vol. 9/8, pp. 796-826.

OECD (2009), "Guidance document for histologic evaluation of endocrine and reproductive tests in rodents", OECD Series on Testing and Assessment, No. 106, OECD, Paris, www.oecd.org/env/ehs/testing/43411534.pdf.

OECD (2008), *Detailed Review Paper on the Use of Metabolising Systems for* In Vitro *Testing of Endocrine Disrupters*, OECD Series on Testing and Assessment, No. 97, OECD Publishing, Paris, https://doi.org/10.1787/9789264085497-en.

WHO/IPCS (2002), "Global assessment of the state-of-the-science of endocrine disrupters", Damstra, T. et al. (eds.) WHO/PCS/EDC/02.2, World Health Organization, Geneva, www.who.int/ipcs/publications/new_issues/endocrine_disruptors/en.

Table C.3.11. **28-Day (Subacute) Inhalation Toxicity Study (OECD TG 412): Guidance for scenarios of combinations of results with existing data**

This table represents possible conclusions to be drawn from assay data, and a next step which could be taken if further evidence is required about possible endocrine disrupting properties and/or effects. The guidance offered is not meant to be prescriptive, but provides science-based considerations. It encourages the use of all available data and expert judgement in a weight of evidence approach. Regional and national interpretation of results and "next steps" may vary.

The conclusions are grouped into a series of scenarios (A-R), each scenario representing a different combination of assay results, existing *in vitro* data and existing *in vivo* data. The symbol "+" indicates that the data in question represent a positive result, «-" indicates a negative result, and "Eq/0" indicates that the data are either equivocal or are not available.

Existing results: * "Mechanism (*in vitro* mechanistic data)" assumes that mechanistic data are available from estrogen receptor (ER-), androgen receptor (AR-) and steroidogenesis-based assays (Level 2). Thyroid hormone receptor (TR) and other assays concerning mechanisms of thyroid disruption may be available, but they are not in common use. In practice, data from all assays may not be available and therefore this must be taken into account when deciding on the "next step". Quantitative structure activity relationship (QSAR) predictions of estrogen and androgen binding/activation may be made for some substances.

Existing results: ** "Effects (*in vivo* effects of concern)" assumes various information, such as data from repeat dose oral toxicity studies, reproduction/developmental toxicity screen tests, read-across from analogues, will be available.

*** *Note:* a positive result is defined as a biologically significant change in any of the endocrine endpoints.

Scenarios	Result of OECD TG 412 (28-day inhalation assay)	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
A	+ ***	+	+	(Anti)-E,A,T,S activity. Increased evidence of (anti)-E,A,T,S activity.	Perform assay from Level 5 (e.g. Extended One-Generation Reproduction Toxicity Study [EOGRTS] or two-generation assay).	If existing data are from a Level 5 assay, there may be sufficient information to conclude evidence of concern for endocrine disruption (the EOGRTS provides the most information; however, for endocrine disrupting chemicals [EDCs] with a carcinogenic potential, OECD TG 451-3 may be more sensitive). Endocrine activity possible in lower vertebrates. Consider performing Fish Sexual Development Test (FSDT), Larval Amphibian Growth and Development Assay (LAGDA) or Medaka Extended One-Generation Reproduction Test (MEOGRT).
B	+	+	–	(Anti)-E,A,T,S activity. Increased evidence of (anti)-E,A,T,S activity.	Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	If existing data are from an adequate level 5 assay, question why there are differences. If existing data are from a less sensitive assay, a higher level test may be required. Consider route of exposures and possible implications of absorption, distribution, metabolism and excretion (ADME) characteristics of the chemical. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT.
C	+	+	Eq/0	(Anti)-E,A,T,S activity. Increased evidence of (anti)-E,A,T,S activity.	Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	Check data on chemical analogues. Consider route of exposure for OECD TG 408 and follow-up assay. Possible implications of ADME characteristics of the chemical. Equivocal results may indicate chemical has multiple modes of action (MOA). Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT.
D	+	–	+	(Anti)-E,A,T,S activity. Acts via non-estrogen receptor (ER-),androgen receptor (AR-), thyroid hormone receptor (TR), steroidogenesis (S) mechanism or may require metabolic activation for activity.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system.	If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude evidence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). Consider route of exposures for effects data and possible implications of ADME characteristics of the chemical. Further mechanistic studies may help determine MOA. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT.
E	+	–	–	(Anti)-E,A,T,S activity. Acts via non-ER, AR, TR, S mechanism or may require metabolic activation for activity. Route of exposure may account for the differences between OECD TG 413 and existing data.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system OR Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	If existing data are from an adequate Level 5 assay, question why there are differences. If existing data are from a less sensitive assay, a higher level test may be required. Consider route of exposures and possible implications of ADME characteristics of the chemical. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT.

Scenarios	Result of OECD TG 412 (28-day inhalation assay)	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
F	+	–	Eq/0	(Anti)-E,A,T,S activity. Acts via non-ER, AR, TR, S mechanism or may require metabolic activation for activity.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system OR Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	Check data on chemical analogues. Further mechanistic studies may help determine MOA. Equivocal results may indicate chemical has multiple MOA. If existing data are from an adequate Level 5 assay, question why there are differences. If existing data are from a less sensitive assay, a higher level test may be required. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT.
G	+	Eq/0	+	(Anti)-E,A,T,S activity. May act via ER, AR, TR, S mechanism (metabolic activation needed).	Perform *in vitro* ER, AR, TR, S assays (for the "0" scenario, otherwise Eq result available) OR Perform *in vitro* ER, AR, TR, S assays with added metabolising system.	If existing data are from a Level 5 assay, there may be sufficient information to conclude evidence of concern for endocrine disruption (the EOGRTS provides the most information; however for, EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). Check data on chemical analogues. Further mechanistic studies may help determine MOA. Equivocal results may indicate chemical has multiple MOA. Consider route of exposures for effects data and possible implications of ADME characteristics of the chemical. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT.
H	+	Eq/0	–	(Anti)-E,A,T,S activity. Acts via unknown mechanism or may require metabolic activation for activity. Route of exposure may account for the differences between OECD TG 413 and existing data. Unknown potential for adverse effects.	For the "0" scenario, perform *in vitro* ER, AR, TR, S assays, maybe with added metabolising system (otherwise Eq result available).	If existing data are from an adequate Level 5 assay, question why there are differences. Consider route of exposures and possible implications of ADME characteristics of the chemical. If existing data are from a less sensitive assay, a higher level test may be required. Check data on chemical analogues. Further mechanistic studies may help determine MOA. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT. Equivocal results may indicate chemical has multiple MOA.
I	+	Eq/0	Eq/0	(Anti)-E,A,T,S activity. Acts via unknown mechanism. Unknown potential for adverse effects.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system OR Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	Check data on chemical analogues. Further mechanistic studies may help determine MOA. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT. Equivocal results may indicate chemical has multiple MOA.

Scenarios	Result of OECD TG 412 (28-day inhalation assay)	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
J	–	+	+	No evidence for (anti)-E,A,T,S activity in OECD TG 413. Weak (anti)-E,A,S activity may not be detected by this assay. Metabolism or potency may explain the difference from existing *in vitro* and *in vivo* data.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system OR Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	If existing data are from an adequate Level 5 assay, question why there are differences. Effects seen in existing studies may be in a more sensitive life stage. Consider route of exposures and possible implications of ADME characteristics of the chemical. Further mechanistic studies may help determine MOA.
K	–	+	–	No evidence for (anti)-E,A,T,S activity in OECD TG 413. Weak (anti)-E,A,S activity may not be detected by this assay. Metabolism or potency may explain *in vitro*/*in vivo* differences.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system OR Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from a less sensitive assay, a higher level test may be required. Further mechanistic studies with metabolism may help determine MOA.
L	–	+	Eq/0	No evidence for (anti)-E,A,T,S activity in OECD TG 413. Weak (anti)-E,A,S activity may not be detected by this assay. Metabolism or potency may explain *in vitro*/*in vivo* differences. Unknown potential for adverse effects.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system OR Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	Metabolic deactivation of chemical may occur *in vivo* so that possible *in vitro* activity is not realised. Consider possible routes of exposure implications of metabolism. Equivocal results may indicate chemical has multiple MOA.
M	–	–	+	No evidence for (anti)-E,A,T,S activity in OECD TG 413. Weak (anti)-E,A,S activity may not be detected by this assay. Effects seen in existing studies are via non-E,A,T,S mechanism.	Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	If existing data are from an adequate Level 5 assay, question why there are differences. Effects seen in existing studies may be in a more sensitive life stage. Consider route of exposures and possible implications of ADME characteristics of the chemical.
N	–	–	–	No evidence for (anti)-E,A,T,S activity in OECD TG 413. Weak (anti)-E,A,S activity may not be detected by this assay. No evidence for (anti)-E,A,T,S activity *in vitro*. No evidence of adverse effects.	Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive).
O	–	–	Eq/0	No evidence for (anti)-E,A,T,S activity in OECD TG 413. Weak (anti)-E,A,S activity may not be detected by this assay. No evidence for (anti)-E,A,T,S activity *in vitro*. Unknown potential for adverse effects.	Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	Consider route of exposures and possible implications for ADME characteristics of the chemical in follow-up assay.

Scenarios	Result of OECD TG 412 (28-day inhalation assay)	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
P	–	Eq/0	+	No evidence for (anti)-E,A,T,S activity in OECD TG 413. Weak (anti)-E,A,S activity may not be detected by this assay. Potential for adverse effects via unknown mechanism.	Perform *in vitro* ER, AR,TR, S assays with added metabolising system.	Consider route of exposure for OECD TG 413 assay and possible implications for differences from existing assay. Effects seen in existing studies may be in a more sensitive life stage. Further mechanistic studies may strengthen weight of evidence. Equivocal results may indicate chemical has multiple MOA.
Q	–	Eq/0	–	No evidence for (anti)-E,A,T,S activity in OECD TG 413. Weak (anti)-E,A,S activity may not be detected by this assay. No evidence of adverse effects.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system.	If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). Further mechanistic studies may strengthen weight of evidence.
R	–	Eq/0	Eq/0	No evidence for (anti)-E,A,T,S activity in OECD TG 413. Weak (anti)-E,A,S activity may not be detected by this assay. Unknown potential for adverse effects.	Perform *in vitro* ER, AR, TR, S assays, otherwise Eq result available.	Further mechanistic studies may strengthen weight of evidence. Equivocal results may indicate chemical has multiple MOA. Check data on chemical analogues.

C.3.12. Subchronic Inhalation Toxicity: 90-Day Study (OECD TG 413)

Status: Assay validated by the OECD.

901. Modalities detected: (anti)estrogen, (anti)androgen, thyroid, steroidogenesis.

Endpoints: Weight of adrenals, epididymides, ovaries, testes, thyroid, uterus.

Histopathologic changes in adrenals, epididymides, female mammary gland, ovaries, pituitary, prostate, seminal vesicles, testes, thyroid, uterus.

Background to the assay

902. This assay determines the subchronic inhalation toxicity of chemicals to provide robust data for quantitative inhalational hazard identification/characterisation. It is primarily used to derive regulatory concentrations for assessing worker risk in occupation settings. It is also used to assess human residential, transportation and environmental risk. Dosing duration is 90 days and the preferred species is the rat. Animals are exposed to test substance in inhalation chambers for six hours per day on a five day per week basis. The preferred mode of exposure is nose only, but whole body exposure may also be used. OECD TG 413 was first adopted in May 1981 and revised in September 2009. Although it has not been validated for the detection of endocrine active substances (EASs), this assay contains several endpoints that are suitable for the determination of endocrine effects. As all the endpoints are apical, it is difficult to discern mechanism of action from this test alone. Information on mechanism of action may need to be obtained from *in vitro* estrogen/androgen/thyroid/steroidogenesis (E,A,T,S) assays or *in vivo* lower tier tests such as the Uterotrophic Bioassay (UT – OECD TG 440) and the Hershberger Bioassay (H – OECD TG 441). Hormone measurements are **not** included in this assay.

903. A comparison can be made with OECD TG 408 (90-Day Oral Toxicity Study) for endocrine endpoints, although inhalational absorption of test substances is likely to result in lower internal doses compared to oral administration. Nevertheless, the number of animals per group (ten male and ten female) and the duration of dosing means that it is a relevant assay to assess when determining potential endocrine activity. Inhalation exposure may be a relevant route of human exposure to certain substances.

When/why the assay may be used

904. This assay may be used as part of a pesticide submission package and may form part of the standard information requirements in certain chemical legislations (e.g. REACH). At least three dose levels are included so that an estimate of no-adverse-effect-level can be determined and the assay used for hazard identification/characterisation. It should be noted that as this assay is not primarily designed to detect endocrine disruption, a higher degree of systemic toxicity is typically induced than is the case with the other Level 3 and 4 assays. The possibly confounding effect of systemic toxicity on endocrine endpoints therefore needs to be considered.

905. In order to provide information relevant for assessing whether or not a chemical may fulfil the WHO/IPCS (2002) definition of an endocrine disruptor (ED), the study design has to be sufficiently robust to demonstrate the presence or absence of effects. In the dose selection, the investigator should also consider and ensure that data generated are adequate to fulfil the regulatory requirement across OECD countries as appropriate (e.g. hazard and risk assessment and labelling, ED assessment, etc.). The top dose or concentration should be sufficiently high to give clear systemic (i.e. non endocrine-specific) toxicity in order to ensure that a wide range of exposures (high to low) is tested. However, endocrine effects observed solely in the presence of clear systemic toxicity should be interpreted with caution and may be disregarded when sufficiently justified to be caused by secondary effects which are unlikely to be due to endocrine activity. The reason for this advice is a concern that some endocrine active substance (EAS) sensitive assays are being run at doses/concentrations of EASs that are too low to trigger direct impacts on the endocrine system. This guidance document is not the place to address this issue directly, but it should be considered when EAS-sensitive test guidelines (TGs) are revised in the future. In addition, the number and spacing of dose/concentration levels should also be adequate to fulfil the objectives of the study (e.g. to demonstrate dose response relationships if this is required).

Introduction to the table of scenarios

906. Table C.3.10 gives guidance on a further step to take in the event of a positive (+) or negative (-) result and in the presence of positive (+), negative (-) or equivocal/absent (Eq/0) existing results. “Existing results” are subdivided into “mechanism” and “effects” data (third and fourth columns). The table is divided horizontally into a series of scenarios that represent all the combinations of these events.

907. The results of OECD TG 413 are given in the second column. As OECD TG 413 is not a screening test where a yes/no (qualitative) answer is obtained for the test as a whole, positive results would generally be assessed for individual endpoints. For the purposes of this guidance, however, a positive result is defined as a biologically significant change in any of the endocrine endpoints listed above (e.g. biologically significant reductions in reproductive organ weights). Changes in related endpoints will increase their biological significance (e.g. changes in the weights of testes and epididymides accompanied by histopathological changes). The guidance on histopathologic changes in endocrine tests (OECD, 2009) may be helpful in interpretation. A negative result for the OECD TG 413 is taken to be the absence of biologically significant changes in all endocrine endpoints.

908. In the absence of other pertinent lines of evidence, negative results in this test alone cannot be taken as evidence that the substance is not an ED. Further studies will be required as confirmation.

909. Equivocal results for the guideline are not considered in the table, partly for brevity but also because equivocal results are by nature uncertain. A decision must eventually be reached about whether the endocrine endpoints tend to be positive or negative or whether the result must be put to one side and the test repeated (using the same or a different test guideline). Factors which may have interfered with the result (e.g. composition of the diet used, environmental influences) should be considered.

Existing data to be considered

910. Existing "mechanism" *in vitro* data are assumed to be available from estrogen receptor (ER-), androgen receptor (AR-) and steroidogenesis-based assays (Level 2). Assays may also be available for interference with thyroid modalities. In practice, it is possible that data from all of these assays may not be available, so judgement will need to be used to decide which assays to perform. Although the current *in vitro* test guidelines do not incorporate metabolic activation, published information on use of metabolic activation systems is available in Jacobs et al. (2008; 2013) and OECD (2008). These methods, however, have not yet been validated.

911. Existing "effects" data refer to *in vivo* effects that may come from Level 3 or 4 tests in the Conceptual Framework (e.g. UT or H assays). In these cases, it should be remembered that these assays are specifically designed to be sensitive to EASs. As mentioned above, the results of the study may be interpreted as part of a battery or group of tests carried out for regulatory purposes. Data may also be available on effects in mammalian and non-mammalian wildlife species, although caution should be used when extrapolating between taxa. A chemical causing endocrine effects in non-mammalian environmental species (fish, for example) may also have endocrine effects in mammals, but the physiological consequences of the effects are likely to be different.

912. When considering the results of the OECD TG 413 assay, all available data should be used in order to reach a conclusion and a weight of evidence approach taken. This may include high throughput screening data, read-across data from structural analogues and quantitative structure activity relationship (QSAR). Several QSAR models for ER and AR binding/activation are now available (see Sections B.1.1.1 and B.1.1.2).

Scenarios: Positive and negative results combined with existing data

913. A series of scenarios (A to R) are presented in Table C.3.12 and represent all the possibilities of positive or negative results in combination with the presence or absence of existing data. The action taken will also depend on the regulatory environment, but the considerations given here are generally science based. Although TG 413 assay uses mammals, the well-conserved nature of the hormonal pathways across taxa indicate that results on endocrine endpoints in this assay may be relevant to other vertebrate species. Effects in laboratory mammal tests are also highly relevant for environmental mammalian species. Wherever possible, the recommended "next step which could be taken" avoids unnecessary animal testing. However, sometimes conducting an animal test will be indicated and then the relevance of species, strain and exposure route should always be considered. The sensitivity and physiological function of the hormone under investigation in the test species should also be considered. In general, lower level tests should be conducted before higher level tests in order to avoid unnecessary animal usage, unless it is apparent that a Level 5 test will be required anyway or will be needed to establish the evidence to conclude on ED properties. Information on some endocrine-related tumours may be detected more comprehensively in carcinogenicity studies (OECD TG 451/453) (Level 4); for example, detection of certain types of thyroid tumors in the absence of reproductive or developmental effects, as well as substances causing tumors in other endocrine-sensitive tissues. At Level 5, the Extended One-Generation Reproduction Toxicity Study (EOGRTS – OECD TG 443) is the most sensitive reproduction assay for detecting endocrine disruption because it includes evaluation of a number of endocrine endpoints not included in the two-generation study (OECD TG 416) adopted in 2001. It is recognised, however, that some jurisdictions

may require a two-generation study. Further considerations specific to each scenario are given in the table.

914. Scenarios A to C represent positive results in the OECD TG 413 assay in the presence of positive *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. A positive result in the *in vitro* assays in combination with a positive OECD TG 413 assay is moderate or strong evidence for E,A,T,S-mediated activity that may or may not be supported by the *in vivo* effects data. In the absence of robust upper-level data, the next step may be to conduct an upper-level test. In the presence of robust *in vivo* data, there may be sufficient evidence to conclude concern for endocrine disruption and therefore no need for further testing. Positive results in the OECD TG 413 assay may also indicate the potential for endocrine mediated effects in lower vertebrates. These could be followed up with partial life cycle tests such as the Fish Sexual Development Test (FSDT), the Larval Amphibian Growth and Development Assay (LAGDA) or the Medaka Extended One-Generation Reproduction Test (MEOGRT) if the evidence were strong enough. *In vivo* assays/tests with negative results should be interpreted with caution as they may either indicate that the tests used do not have sufficient power to detect weak effects or, alternatively, that the effects do not present a concern for endocrine disruption. The possibility of other (non-E,A,T,S) mechanisms should also not be overlooked (e.g. involving other receptors or endocrine axes).

915. Scenarios D to F represent positive results in the OECD TG 413 assay in the presence of negative *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. Negative results in the *in vitro* assays should be viewed with caution in case a metabolite is responsible for the positive OECD TG 413 assay. Unless the metabolic profile of the test substance is known, one option may be to conduct these assays with an added metabolising system. If the metabolic profile is known, then a higher level *in vivo* test may be advisable. The choice of tests will depend on the available *in vivo* effects data. Positive results in the OECD TG 413 assay may also indicate the potential for endocrine mediated effects in lower vertebrates. As in Scenarios A to C, *in vivo* assays/tests with negative results should be interpreted with caution as they may either indicate that the tests used do not have sufficient power to detect weak effects or, alternatively, that the effects do not present a concern for endocrine disruption.

916. Scenarios G to I represent positive results in the OECD TG 413 assay in the presence of various combinations of missing or equivocal data. Positive results in the OECD TG 413 assay may also indicate the potential for endocrine mediated effects in lower vertebrates. The next step to take in these eventualities will depend on the nature of the other available data and the jurisdiction in which it is being used. In some cases, equivocal data may be viewed as positive whilst in others it may or may not contribute to the weight of evidence. The interpretation may also depend on the mode of action (MOA) in question and why the data are considered equivocal, e.g. a study that is equivocal for thyroid effects may still be of value in evaluating (anti)androgenic effects. In all three scenarios, the recommended first step is to obtain reliable mechanistic (*in vitro*) data rather than proceed further with *in vivo* testing. Equivocal and missing data are alternative scenarios and two possibilities for the next step are given in most cases, but the nature of equivocal data means that decisions need to be taken on a case-by-case basis. In all cases, the role of metabolism, route of exposure and data from structural analogues should be considered before deciding on the next step.

917. Scenarios J to L represent negative results in the OECD TG 413 assay in the presence of positive *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. Negative outcomes in the OECD TG 413 should be viewed with caution

because of the power of the assay to detect (anti)estrogens and androgens may be limited All three scenarios could also arise from a chemical that is positive in *in vitro* assays, but is metabolised to a non-active metabolite leading to negative results in the OECD TG 413 assay. This should be considered first when investigating the next step. Endocrine active potency may also explain differences between *in vitro* and *in vivo* results (e.g. a weak chemical may give a positive result *in vitro* but may be negative *in vivo*). Positive *in vivo* effects data may involve other E,A,T,S, non-E,A,T,S mechanisms (e.g. involving other receptors or endocrine axes), more sensitive endpoints, greater statistical power or life stages that are more sensitive to the substance than the adult dermally exposed animals in OECD TG 413.

918. Scenarios M to O represent negative results in the OECD TG 413 assay in the presence of negative *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. Negative results for all tests (Scenario N) may be sufficient to enable a conclusion of no concern for endocrine disruption. This will depend on the weight of evidence and may not be possible. Where there are positive *in vivo* effects data there could still be an E,A,T,S-related mechanism, the effects may be related to length of exposure, route of exposure or exposure at different life stages. Other E,A,T,S or non-E,A,T,S mechanisms may also be involved.

919. Scenarios P to R represent negative results in the OECD TG 413 assay in the presence of various combinations of missing or equivocal data. As with the positive result scenarios above (see Paragraph 918), the next step to take in these eventualities will have to be decided on a case-by-case basis. However, the recommended first step is generally to obtain reliable mechanistic (*in vitro*) data rather than proceed further with *in vivo* testing. In all cases, the role of metabolism, route of exposure and data from structural analogues should be considered before deciding on the next step.

920. In all scenarios (A to R), the next step to take to strengthen weight of evidence will depend on the existing information. The table is meant to provide a succinct guide and may not cover all circumstances or possibilities. The scenarios may also suggest that chemicals have simple or single MOA, when in practice they may have multiple endocrine and non-endocrine MOA. In some cases for example, two opposite modes of simultaneous action (e.g. estrogenic and anti-estrogenic) could, depending on dose, lead to a minimisation or abolition of effects, while in others two different MOA (e.g. estrogenic and anti-androgenic) could potentially reinforce effects. Endocrine pathways interact, mixed effects are common and there are many pathways that cannot be distinguished with currently available TGs. If multiple MOA are suspected, either from the existing results or based on QSAR/read-across/integrated approaches, this should be investigated further if needed for regulatory decision making.

References

Jacobs, M. et al. (2013), "*In vitro* metabolism and bioavailability tests for endocrine active substances: What is needed next for regulatory purposes?", *ALTEX – Alternatives to Animal Experimentation*, Vol. 30/3, pp. 331-351.

Jacobs, M.N. et al. (2008), "The use of metabolising systems for *in vitro* testing of endocrine disrupters", *Current Drug Metabolism*, Vol. 9/8, pp. 796-826.

OECD (2009), "Guidance document for histologic evaluation of endocrine and reproductive tests in rodents", OECD Series on Testing and Assessment, No. 106, OECD, Paris, www.oecd.org/env/ehs/testing/43411534.pdf.

OECD (2008), *Detailed Review Paper on the Use of Metabolising Systems for* In Vitro *Testing of Endocrine Disrupters*, OECD Series on Testing and Assessment, No. 97, OECD Publishing, Paris, https://doi.org/10.1787/9789264085497-en.

WHO/IPCS (2002), "Global assessment of the state-of-the-science of endocrine disrupters", Damstra, T. et al. (eds.) WHO/PCS/EDC/02.2, World Health Organization, Geneva, www.who.int/ipcs/publications/new_issues/endocrine_disruptors/en.

Table C.3.12. **Subchronic Inhalation Toxicity: 90-Day Study (OECD TG 413): Guidance for scenarios of combinations of results with existing data**

This table represents possible conclusions to be drawn from assay data, and a next step which could be taken if further evidence is required about possible endocrine disrupting properties and/or effects. The guidance offered is not meant to be prescriptive, but provides science-based considerations. It encourages the use of all available data and expert judgement in a weight of evidence approach. Regional and national interpretation of results and "next steps" may vary.

The conclusions are grouped into a series of scenarios (A-R), each scenario representing a different combination of assay results, existing *in vitro* data and existing *in vivo* data. The symbol "+" indicates that the data in question represent a positive result, «-" indicates a negative result, and "Eq/0" indicates that the data are either equivocal or are not available.

Existing results: * "Mechanism (*in vitro* mechanistic data)" assumes that mechanistic data are available from estrogen receptor (ER-), androgen receptor (AR-) and steroidogenesis-based assays (Level 2). Thyroid hormone receptor (TR) and other assays concerning mechanisms of thyroid disruption may be available, but they are not in common use. In practice, data from all assays may not be available and therefore this must be taken into account when deciding on the "next step". Quantitative structure activity relationship (QSAR) predictions of estrogen and androgen binding/activation may be made for some substances.

Existing results: ** "Effects (*in vivo* effects of concern)" assumes various information, such as data from repeat dose oral toxicity studies, reproduction/developmental toxicity screen tests, read-across from analogues, will be available.

*** *Note:* a positive result is defined as a biologically significant change in any of the endocrine endpoints.

Scenarios	Result of OECD TG 413 (90-day inhalation assay)	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
A	+ ***	+	+	(Anti)-E,A,T,S activity. Increased evidence of (anti)-E,A,T,S activity.	Perform assay from Level 5 (e.g. Extended One-Generation Reproduction Toxicity Study [EOGRTS] or two-generation assay).	If existing data are from a Level 5 assay, there may be sufficient information to conclude evidence of concern for endocrine disruption (the EOGRTS provides the most information; however, for endocrine disrupting chemicals [EDCs] with a carcinogenic potential, OECD TG 451-3 may be more sensitive). Endocrine activity possible in lower vertebrates. Consider performing a Fish Sexual Development Test (FSDT), Larval Amphibian Growth and Development Assay (LAGDA) or Medaka Extended One-Generation Reproduction Test (MEOGRT).
B	+	+	–	(Anti)-E,A,T,S activity. Increased evidence of (anti)-E,A,T,S activity.	Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	If existing data are from an adequate Level 5 assay, question why there are differences. If existing data are from a less sensitive assay, a higher level test may be required. Consider route of exposures and possible implications of absorption, distribution, metabolism and excretion (ADME) characteristics of the chemical. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT.
C	+	+	Eq/0	(Anti)-E,A,T,S activity. Increased evidence of (anti)-E,A,T,S activity.	Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	Check data on chemical analogues. Consider route of exposure for OECD TG 408 and follow-up assay. Possible implications of ADME characteristics of the chemical. Equivocal results may indicate chemical has multiple modes of action (MOA). Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT.
D	+	–	+	(Anti)-E,A,T,S activity. Acts via non-estrogen receptor (ER-),androgen receptor (AR-), thyroid hormone receptor (TR), steroidogenesis (S) mechanism or may require metabolic activation for activity.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system.	If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude evidence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). Consider route of exposures for effects data and possible implications of ADME characteristics of the chemical. Further mechanistic studies may help determine MOA. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT.
E	+	–	–	(Anti)-E,A,T,S activity. Acts via non-ER, AR, TR, S mechanism or may require metabolic activation for activity. Route of exposure may account for the differences between OECD TG 413 and existing data.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system OR Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	If existing data are from an adequate Level 5 assay, question why there are differences. If existing data are from a less sensitive assay, a higher level test may be required. Consider route of exposures and possible implications of ADME characteristics of the chemical. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT.

Scenarios	Result of OECD TG 413 (90-day inhalation assay)	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
F	+	–	Eq/0	(Anti)-E,A,T,S activity. Acts via non-ER, AR, TR, S mechanism or may require metabolic activation for activity.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system OR Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	Check data on chemical analogues. Further mechanistic studies may help determine MOA. Equivocal results may indicate chemical has multiple MOA. If existing data are from an adequate Level 5 assay, question why there are differences. If existing data are from a less sensitive assay, a higher level test may be required. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT.
G	+	Eq/0	+	(Anti)-E,A,T,S activity. May act via ER, AR, TR, S mechanism (metabolic activation needed).	Perform *in vitro* ER, AR, TR, S assays. (for the "0" scenario, otherwise Eq result available) OR Perform *in vitro* ER, AR, TR, S assays with added metabolising system.	If existing data are from a Level 5 assay, there may be sufficient information to conclude evidence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). Check data on chemical analogues. Further mechanistic studies may help determine MOA. Equivocal results may indicate chemical has multiple MOA. Consider route of exposures for effects data and possible implications of ADME characteristics of the chemical. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT.
H	+	Eq/0	–	(Anti)-E,A,T,S activity. Acts via unknown mechanism or may require metabolic activation for activity. Route of exposure may account for the differences between OECD TG 413 and existing data. Unknown potential for adverse effects.	For the "0" scenario, perform *in vitro* ER, AR, TR, S assays, maybe with added metabolising system (otherwise Eq result available).	If existing data are from an adequate Level 5 assay, question why there are differences. Consider route of exposures and possible implications of ADME characteristics of the chemical. If existing data are from a less sensitive assay, a higher level test may be required. Check data on chemical analogues. Further mechanistic studies may help determine MOA. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT. Equivocal results may indicate chemical has multiple MOA.
I	+	Eq/0	Eq/0	(Anti)-E,A,T,S activity. Acts via unknown mechanism. Unknown potential for adverse effects.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system OR Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	Check data on chemical analogues. Further mechanistic studies may help determine MOA. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT. Equivocal results may indicate chemical has multiple MOA.

Scenarios	Result of OECD TG 413 (90-day inhalation assay)	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
J	–	+	+	No evidence for (anti)-E,A,T,S activity in OECD TG 413. Weak (anti)-E,A,S activity may not be detected by this assay. Metabolism or potency may explain the difference from existing *in vitro* and *in vivo* data.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system OR Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	If existing data are from an adequate Level 5 assay, question why there are differences. Effects seen in existing studies may be in a more sensitive life stage. Consider route of exposures and possible implications of ADME characteristics of the chemical. Further mechanistic studies may help determine MOA.
K	–	+	–	No evidence for (anti)-E,A,T,S activity in OECD TG 413. Weak (anti)-E,A,S activity may not be detected by this assay. Metabolism or potency may explain *in vitro*/*in vivo* differences.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system OR Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from a less sensitive assay, a higher level test may be required. Further mechanistic studies with metabolism may help determine MOA.
L	–	+	Eq/0	No evidence for (anti)-E,A,T,S activity in OECD TG 413. Weak (anti)-E,A,S activity may not be detected by this assay. Metabolism or potency may explain *in vitro*/*in vivo* differences. Unknown potential for adverse effects.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system OR Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	Metabolic deactivation of chemical may occur *in vivo* so that possible *in vitro* activity is not realised. Consider possible routes of exposure implications of metabolism. Equivocal results may indicate chemical has multiple MOA.
M	–	–	+	No evidence for (anti)-E,A,T,S activity in OECD TG 413. Weak (anti)-E,A,S activity may not be detected by this assay. Effects seen in existing studies are via non-E,A,T,S mechanism.	Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	If existing data are from an adequate Level 5 assay, question why there are differences. Effects seen in existing studies may be in a more sensitive life stage. Consider route of exposures and possible implications of ADME characteristics of the chemical.
N	–	–	–	No evidence for (anti)-E,A,T,S activity in OECD TG 413. Weak (anti)-E,A,S activity may not be detected by this assay. No evidence for (anti)-E,A,T,S activity *in vitro*. No evidence of adverse effects.	Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive).

Scenarios	Result of OECD TG 413 (90-day inhalation assay)	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
O	–	–	Eq/0	No evidence for (anti)-E,A,T,S activity in OECD TG 413. Weak (anti)-E,A,S activity may not be detected by this assay. No evidence for (anti)-E,A,T,S activity *in vitro*. Unknown potential for adverse effects.	Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	Consider route of exposures and possible implications for ADME characteristics of the chemical in follow-up assay.
P	–	Eq/0	+	No evidence for (anti)-E,A,T,S activity in OECD TG 413. Weak (anti)-E,A,S activity may not be detected by this assay. Potential for adverse effects via unknown mechanism.	Perform *in vitro* ER, AR,TR, S assays with added metabolising system.	Consider route of exposure for OECD TG 413 assay and possible implications for differences from existing assay. Effects seen in existing studies may be in a more sensitive life stage. Further mechanistic studies may strengthen weight of evidence. Equivocal results may indicate chemical has multiple MOA.
Q	–	Eq/0	–	No evidence for (anti)-E,A,T,S activity in OECD TG 413. Weak (anti)-E,A,S activity may not be detected by this assay. No evidence of adverse effects.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system.	If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). Further mechanistic studies may strengthen weight of evidence.
R	–	Eq/0	Eq/0	No evidence for (anti)-E,A,T,S activity in OECD TG 413. Weak (anti)-E,A,S activity may not be detected by this assay. Unknown potential for adverse effects.	Perform *in vitro* ER, AR, TR, S assays, otherwise Eq result available.	Further mechanistic studies may strengthen weight of evidence. Equivocal results may indicate chemical has multiple MOA. Check data on chemical analogues.

C.3.13. Repeated Dose 90-Day Oral Toxicity Study in Non-Rodents (OECD TG 409)

Status: Assay validated by the OECD.

921. Modalities detected: (anti)estrogen, (anti)androgen, thyroid, steroidogenesis.

Endpoints: Weight of adrenals, testes, epididymides, uterus, ovaries, thyroid.

Histopathologic changes in pituitary, thyroid gland, gonads, uterus, accessory sex organs, female mammary gland, testes and adrenals.

Background to the assay

922. This assay determines the general toxicity of chemicals in non-rodents after 90 days of oral dosing (by gavage, via the diet, in drinking water or in capsules). The most commonly used non-rodent species is the dog, which should be of a defined breed (beagle usually). It provides information on major toxic effects and target organ toxicity likely to arise from the post-weaning period until well into adulthood. OECD TG 409 was adopted in September 1981 and revised in September 1998.

923. Although it has not been validated for the detection of endocrine active substances (EASs), OECD TG 409 contains many endpoints that are suitable for the determination of endocrine effects. The well-conserved nature of many endocrine receptors and systems means that many endpoints validated in rat studies will have similar responses in non-rodent species used in OECD TG 409. A comparison can be made with validation of the OECD TG 407 (28-Day Oral Toxicity Study) for endocrine endpoints where substances that were moderate and strong endocrine disruptors (EDs) for (anti)estrogenicity and (anti)androgenicity (e.g. ethinylestradiol and flutamide) and weak and strong modulators of thyroid hormone-related effects (e.g. propylthiouracil, T4 and methyl testosterone) were detected (OECD, 2006). Steroidogenesis inhibition was also detected, although only one (potent) chemical was used in the validation study (CGS 18320B). OECD TG 409 may be more sensitive than OECD TG 407 because of the extended dosing period, although the number of animals per group is similar (at least four male and four female, compared with five per sex per group in OECD TG 407). OECD TG 409, however, does not contain some endocrine-sensitive endpoints (e.g. thyroid hormones, estrous cyclicity) that may be included in OECD TG 407. As all the endpoints are apical, it is difficult to discern mechanism of action from this test alone. Information on mechanism of action may need to be obtained from *in vitro* estrogen/androgen/thyroid/steroidogenesis (E,A,T,S) assays or *in vivo* lower tier tests such as the Uterotrophic Bioassay (UT – OECD TG 440) and Hershberger Bioassay (H – OECD TG 441). Possible species differences in response, physiological differences and species differences in test substance metabolism should also be considered.

When/why the assay may be used

924. This assay is likely to be used as part of a pesticide submission package and may form part of the standard information requirements in certain chemical legislations (e.g. Plant Protection Product Regulations in the European Union). At least three dose levels are included so that an estimate of no-adverse-effect-level can be determined and the assay used for hazard identification/characterisation. It should be noted that as this assay is not primarily designed to detect endocrine disruption, a higher degree of systemic toxicity is typically induced than is the case with the other Level 3 and 4 assays. The possibly confounding effect of systemic toxicity on endocrine endpoints therefore needs to be considered.

925. In order to provide information relevant for assessing whether or not a chemical may fulfil the WHO/IPCS (2002) definition of an ED, the study design has to be sufficiently robust to demonstrate the presence or absence of effects. In the dose selection, the investigator should also consider and ensure that data generated are adequate to fulfil the regulatory requirement across OECD countries as appropriate (e.g. hazard and risk assessment and labelling, ED assessment, etc.). The top dose or concentration should be sufficiently high to give clear systemic (i.e. non endocrine-specific) toxicity in order to ensure that a wide range of exposures (high to low) is tested. However, endocrine effects observed solely in the presence of clear systemic toxicity should be interpreted with caution and may be disregarded when sufficiently justified to be caused by secondary effects which are unlikely to be due to endocrine activity. The reason for this advice is a concern that some EAS-sensitive assays are being run at doses/concentrations of EASs that are too low to trigger direct impacts on the endocrine system. This guidance document is not the place to address this issue directly, but it should be considered when EAS-sensitive test guidelines (TGs) are revised in the future. In addition, the number and spacing of dose/concentration levels should also be adequate to fulfil the objectives of the study (e.g. to demonstrate dose response relationships if this is required).

Introduction to the table of scenarios

926. Table C.3.13 gives guidance on a further step to take in the event of a positive (+) or negative (-) result and in the presence of positive (+), negative (-) or equivocal/absent (Eq/0) existing results. “Existing results” are subdivided into “mechanism” and “effects” data (third and fourth columns). The table is divided horizontally into a series of scenarios that represent all the combinations of these events.

927. The results of OECD TG 409 are given in the second column. As OECD TG 409 is not a screening test where a yes/no (qualitative) answer is obtained for the test as a whole, positive results would generally be assessed for individual endpoints. For the purposes of this guidance, however, a positive result is defined as a biologically significant change in any of the endocrine endpoints listed above (e.g. statistically significant reductions in reproductive organ weights). Changes in related endpoints will increase their biological significance (e.g. changes in the weights of testes and epididymides accompanied by histopathological changes). The guidance on histopathologic changes in endocrine tests (OECD, 2009) may be helpful in interpretation. A negative result for OECD TG 409 is taken to be the absence of biologically significant changes in all endocrine endpoints.

928. In the absence of other pertinent lines of evidence, negative results in this test alone cannot be taken as evidence that the substance is not an ED. Further studies will be required as confirmation.

929. Equivocal results for the guideline are not considered in the table, partly for brevity but also because equivocal results are by nature uncertain. A decision must eventually be reached about whether the endocrine endpoints tend to be positive or negative or whether the result must be put to one side and the test repeated (using the same or a different test guideline). Factors which may have interfered with the result (e.g. composition of the diet used, environmental influences) should be considered.

Existing data to be considered

930. Existing "mechanism" *in vitro* data are assumed to be available from estrogen receptor (ER-), androgen receptor (AR-) and steroidogenesis-based assays (Level 2). Assays may also be available for interference with thyroid modalities. In practice, it is possible that data from all of these assays may not be available, so judgement will need to be used to decide which assays to perform. Although the current *in vitro* test guidelines do not incorporate metabolic activation, published information on use of metabolic activation systems is available in Jacobs et al. (2008; 2013) and OECD (2008). These methods, however, have not yet been validated.

931. Existing "effects" data refer to *in vivo* effects that may come from Level 3 or 4 tests in the Conceptual Framework (e.g. UT or H assays). In these cases, it should be remembered that these assays are specifically designed to be sensitive to EDs. As mentioned above, the results of the study may be interpreted as part of a battery or group of tests carried out for regulatory purposes. Data may also be available on effects in mammalian and non-mammalian wildlife species, although caution should be used when extrapolating between taxa. A chemical causing endocrine effects in non-mammalian environmental species (fish, for example) may also have endocrine effects in mammals, but the physiological consequences of the effects are likely to be different.

932. When considering the results of the OECD TG 409 assay, all available data should be used in order to reach a conclusion and a weight of evidence approach taken. This may include high throughput screening data, read-across data from structural analogues and quantitative structure activity relationship (QSAR). Several QSAR models for estrogen and androgen binding/activation are now available (see Sections B.1.1.1 and B.1.1.2).

Scenarios: Positive and negative results combined with existing data

933. A series of scenarios (A to R) are presented in Table C.3.13 and represent all the possibilities of positive or negative results in combination with the presence or absence of existing data. The action taken will also depend on the regulatory environment, but the considerations given here are generally science based. Although the OECD TG 409 assay uses mammals, the well-conserved nature of the hormonal pathways across taxa indicate that results on endocrine endpoints in this assay may be relevant to other vertebrate species. Effects in laboratory mammal tests are also highly relevant for environmental mammalian species. Wherever possible, the recommended "next step which could be taken" avoids unnecessary animal testing. However, sometimes conducting an animal test will be indicated and then the relevance of species, strain and exposure route should always be considered. The sensitivity and physiological function of the hormone under investigation in the test species should also be considered. As OECD TG 409 is a non-rodent assay, it provides insight into endocrine effects across species. In general, lower level tests should be conducted before higher level tests in order to avoid unnecessary animal usage, unless it is apparent that a Level 5 test will be required anyway or will be needed to establish the

evidence to conclude on ED properties. Information on some endocrine-related tumours may be detected more comprehensively in carcinogenicity studies (OECD TG 451/453) (Level 4); for example, detection of certain types of thyroid tumors in the absence of reproductive or developmental effects, as well as substances causing tumors in other endocrine-sensitive tissues. At Level 5, the Extended One-Generation Reproduction Toxiticy Study (EOGRTS – OECD TG 443) is the most sensitive reproduction assay for detecting endocrine disruption because it includes evaluation of a number of endocrine endpoints not included in the two-generation study (OECD TG 416) adopted in 2001. It is recognised, however, that some jurisdictions may require a two-generation study. Further considerations specific to each scenario are given in the table.

934. Scenarios A to C represent positive results in the OECD TG 409 assay in the presence of positive *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. A positive result in the *in vitro* assays in combination with a positive OECD TG 409 assay is moderate or strong evidence for E,A,T,S-mediated activity that may or may not be supported by the *in vivo* effects data. OECD TG 409 uses non-rodents and therefore a positive result in non-rodent species in combination with positive rodent data indicates a higher level of concern for human health. In the absence of robust upper-level data, the next step may be to conduct an upper-level test. In the presence of robust *in vivo* data, there may be sufficient evidence to conclude concern for endocrine disruption and therefore no need for further testing. Positive results in the OECD TG 409 assay may also indicate the potential for endocrine mediated effects in lower vertebrates. These could be followed up with partial life cycle tests such as the Fish Sexual Development Test (FSDT), the Larval Amphibian Growth and Development Assay (LAGDA) or the Medaka Extended One-Generation Reproduction Test (MEOGRT) if the evidence were strong enough. *In vivo* assays/tests with negative results should be interpreted with caution as they may either indicate that the tests used do not have sufficient power to detect weak effects or, alternatively, that the effects do not present a concern for endocrine disruption. There may also be species differences in the effects on the endocrine system. The possibility of other (non-E,A,T,S) mechanisms should also not be overlooked (e.g. involving other receptors or endocrine axes).

935. Scenarios D to F represent positive results in the OECD TG 409 assay in the presence of negative *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. Negative results in the *in vitro* assays should be viewed with caution in case a metabolite is responsible for the positive OECD TG 409 assay. Unless the metabolic profile of the test substance is known, one option may be to conduct these assays with an added metabolising system. If the metabolic profile is known, then a higher level *in vivo* test may be advisable. The choice of tests will depend on the available *in vivo* effects data. Positive results in the OECD TG 408 assay may also indicate the potential for endocrine mediated effects in lower vertebrates. As in Scenarios A to C, *in vivo* assays/tests with negative results should be interpreted with caution as they may either indicate that the tests used do not have sufficient power to detect weak effects or, alternatively, that the effects do not present a concern for endocrine disruption. There may also be species differences in the effects on the endocrine system.

936. Scenarios G to I represent positive results in the OECD TG 409 assay in the presence of various combinations of missing or equivocal data. Positive results in the OECD TG 409 assay may also indicate the potential for endocrine mediated effects in lower vertebrates. The next step to take in these eventualities will depend on the nature of the other available data and the jurisdiction in which it is being used. In some cases, equivocal data may be viewed as positive whilst in others it may or may not contribute to the weight

of evidence. The interpretation may also depend on the mode of action (MOA) in question and why the data are considered equivocal, e.g. a study that is equivocal for thyroid effects may still be of value in evaluating (anti)androgenic effects. In all three scenarios, the recommended first step is to obtain reliable mechanistic (*in vitro*) data rather than proceed further with *in vivo* testing. Equivocal and missing data are alternative scenarios and two possibilities for the next step are given in most cases, but the nature of equivocal data means that decisions need to be taken on a case-by-case basis. As OECD TG 409 is a non-rodent assay, the possibility of species differences in response should be considered. In all cases, the role of metabolism, route of exposure and data from structural analogues should be considered before deciding on the next step.

937. Scenarios J to L represent negative results in the OECD TG 409 assay in the presence of positive *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. Negative outcomes in OECD TG 409 should be viewed with caution because of the power of the assay to detect (anti)estrogens and androgens may be limited. All three scenarios could also arise from a chemical that is positive in *in vitro* assays, but is metabolised to a non-active metabolite leading to negative results in the OECD TG 409 assay. This should be considered first when investigating the next step. Endocrine active potency may also explain differences between *in vitro* and *in vivo* results (e.g. a weak chemical may give a positive result *in vitro* but may be negative *in vivo*). Species differences may also account for a negative result in the non-rodent assay and positive *in vivo* effects data in rodents. Positive *in vivo* effects data may involve other E,A,T,S, non-E,A,T,S mechanisms (e.g. involving other receptors or endocrine axes), more sensitive endpoints, greater statistical power or life stages that are more sensitive to the substance than the young adult exposed animals in OECD TG 408.

938. Scenarios M to O represent negative results in the OECD TG 409 assay in the presence of negative *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. Negative results for all tests (Scenario N) may be sufficient to enable a conclusion of no concern for endocrine disruption. OECD TG 409 uses non-rodents and therefore a negative result in non-rodent species in combination with negative rodent data may allow more confidence in a conclusion of no concern for human health. This will depend on the weight of evidence and may not be possible. Where there are positive *in vivo* effects data, there could still be an E,A,T,S-related mechanism, the effects may be related to length of exposure, route of exposure, exposure at different life stages or species differences. Other E,A,T,S or non-E,A,T,S mechanisms may also be involved.

939. Scenarios P to R represent negative results in the OECD TG 409 assay in the presence of various combinations of missing or equivocal data. As with the positive result scenarios above (see Paragraph 937), the next step to take in these eventualities will have to be decided on a case-by-case basis. However, the recommended first step is generally to obtain reliable mechanistic (*in vitro*) data rather than proceed further with *in vivo* testing. In all cases, the role of metabolism, route of exposure, species differences and data from structural analogues should be considered before deciding on the next step.

940. In all scenarios (A to R), the next step to take to strengthen weight of evidence will depend on the existing information. The table is meant to provide a succinct guide and may not cover all circumstances or possibilities. The scenarios may also suggest that chemicals have simple or single MOA, when in practice they may have multiple endocrine and non-endocrine MOA. In some cases, for example, two opposite modes of simultaneous action (e.g. estrogenic and anti-estrogenic) could, depending on dose, lead to a minimisation or abolition of effects, while in others two different MOA (e.g. estrogenic and anti-androgenic)

could potentially reinforce effects. Endocrine pathways interact, mixed effects are common and there are many pathways that cannot be distinguished with currently available TGs. If multiple MOA are suspected, either from the existing results or based on QSAR/read-across/integrated approaches, this should be investigated further if needed for regulatory decision making.

References

Jacobs, M. et al. (2013), "*In vitro* metabolism and bioavailability tests for endocrine active substances: What is needed next for regulatory purposes?", *ALTEX – Alternatives to Animal Experimentation*, Vol. 30/3, pp. 331-351.

Jacobs, M.N. et al. (2008), "The use of metabolising systems for *in vitro* testing of endocrine disrupters", *Current Drug Metabolism*, Vol. 9/8, pp. 796-826.

OECD (2009), "Guidance document for histologic evaluation of endocrine and reproductive tests in rodents", OECD Series on Testing and Assessment, No. 106, OECD, Paris, www.oecd.org/env/ehs/testing/43411534.pdf.

OECD (2008), *Detailed Review Paper on the Use of Metabolising Systems for* In Vitro *Testing of Endocrine Disrupters*, OECD Series on Testing and Assessment, No. 97, OECD Publishing, Paris, https://doi.org/10.1787/9789264085497-en.

OECD (2006), "Report of the validation of the updated test guideline 407: Repeat dose 28-day oral toxicity study in laboratory rats", OECD Series on Testing and Assessment, No. 59, OECD, Paris, www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2006)26&doclanguage=en.

WHO/IPCS (2002), "Global assessment of the state-of-the-science of endocrine disrupters", Damstra, T. et al. (eds.) WHO/PCS/EDC/02.2, World Health Organization, Geneva, www.who.int/ipcs/publications/new_issues/endocrine_disruptors/en.

Table C.3.13. **Repeated Dose 90-Day Oral Toxicity Study in Non-Rodents (OECD TG 409): Guidance for scenarios of combinations of results with existing data**

This table represents possible conclusions to be drawn from assay data, and a next step which could be taken if further evidence is required about possible endocrine disrupting properties and/or effects. The guidance offered is not meant to be prescriptive, but provides science-based considerations. It encourages the use of all available data and expert judgement in a weight of evidence approach. Regional and national interpretation of results and "next steps" may vary.

The conclusions are grouped into a series of scenarios (A-R), each scenario representing a different combination of assay results, existing *in vitro* data and existing *in vivo* data. The symbol "+" indicates that the data in question represent a positive result, «-" indicates a negative result, and "Eq/0" indicates that the data are either equivocal or are not available.

Existing results: * "Mechanism (*in vitro* mechanistic data)" assumes that mechanistic data are available from estrogen receptor (ER-), androgen receptor (AR-) and steroidogenesis-based assays (Level 2). Thyroid hormone receptor (TR) and other assays concerning mechanisms of thyroid disruption may be available, but they are not in common use. In practice, data from all assays may not be available and therefore this must be taken into account when deciding on the "next step". Quantitative structure activity relationship (QSAR) predictions of estrogen and androgen binding/activation may be made for some substances.

Existing results: ** "Effects (*in vivo* effects of concern)" assumes various information, such as data from repeat dose oral toxicity studies, reproduction/developmental toxicity screen tests, read-across from analogues, will be available.

*** *Note:* a positive result is defined as a biologically significant change in any of the endocrine endpoints.

Scenarios	Result of OECD TG 409 (non-rodent 90-day assay)	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
A	+ ***	+	+	Increased evidence of (anti)-E,A,T,S activity in multiple species.	Perform assay from Level 5 (e.g. Extended One-Generation Reproduction Toxiticy Study [EOGRTS] or two-generation assay).	If existing data are from a Level 5 assay, there may be sufficient information to conclude evidence of concern for endocrine disruption (the EOGRTS provides the most information; however, for endocrine disrupting chemicals [EDCs] with a carcinogenic potential, OECD TG 451-3 may be more sensitive). Effects on indicators of hormonal activity alone may be indicative of changes not detected by apical endpoints. Effects on apical endpoints alone may indicate estrogen/androgen/ thyroid/steroidogenesis (E,A,T,S) modalities or other mechanisms. Consider route of exposures for effects data and possible implications of absorption, distribution, metabolism and excretion (ADME) characteristics of the chemical. Endocrine activity possible in lower vertebrates. Consider performing a Fish Sexual Development Test (FSDT), Larval Amphibian Growth and Development Assay (LAGDA) or Medaka Extended One-Generation Reproduction Test (MEOGRT).
B	+	+	–	Evidence of (anti)-E,A,T,S activity in non-rodent species. (Anti)-E,A,T,S activity. Route of exposure or species differences may account for the differences between OECD TG 409 and existing data.	Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	If existing data are from an adequate Level 5 assay, question why there are differences. If existing data are from a less sensitive assay, a higher level test may be required. Consider route of exposures and possible implications of ADME characteristics of the chemical in different species. Consider species differences in physiology and response. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT.
C	+	+	Eq/0	Evidence of (anti)-E,A,T,S activity in non-rodent species. (Anti)-E,A,T,S activity. Route of exposure or species differences may account for the differences between OECD TG 409 and existing data.	Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	Check data on chemical analogues. Consider species and route of exposure for OECD TG 409 and follow-up assay. Possible implications of ADME characteristics of the chemical in different species. Consider species differences in physiology and response. Equivocal results may indicate chemical has multiple modes of action (MOA). Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT.

Scenarios	Result of OECD TG 409 (non-rodent 90-day assay)	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
D	+	–	+	Increased evidence of (anti)-E,A,T,S activity in multiple species. Acts via non-estrogen receptor (ER-),androgen receptor (AR-), thyroid hormone receptor (TR), steroidogenesis (S) mechanism or may require metabolic activation for activity.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system.	If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude evidence of concern for endocrine disruption (the EOGRTS provides the most information; however, for endocrine disrupting chemicals [EDCs] with a carcinogenic potential, OECD TG 451-3 may be more sensitive). Consider route of exposures for effects data and possible implications of ADME characteristics of the chemical in different species. Further mechanistic studies may help determine MOA. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT.
E	+	–	–	Evidence of (anti)-E,A,T,S activity in non-rodent species. Acts via non-ER, AR, TR, S mechanism or may require metabolic activation for activity. Route of exposure or species differences may account for the differences between OECD TG 409 and existing data.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system OR Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	If existing data are from an adequate Level 5 assay, question why there are differences. If existing data are from a less sensitive assay, a higher level test may be required. Consider route of exposures and possible implications of ADME characteristics of the chemical in different species. Consider species differences in physiology and response. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT.
F	+	–	Eq/0	Evidence of (anti)-E,A,T,S activity in non-rodent species. Acts via non-ER, AR, TR, S mechanism or may require metabolic activation for activity. Route of exposure or species differences may account for the differences between OECD TG 409 and existing data.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system OR Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	Check data on chemical analogues. Further mechanistic studies may help determine MOA. If existing data are from an adequate Level 5 assay, question why there are differences. If existing data are from a less sensitive assay, a higher level test may be required. Consider route of exposures and possible implications of ADME characteristics of the chemical in different species. Consider species differences in physiology and response. Equivocal results may indicate chemical has multiple MOA. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT.

Scenarios	Result of OECD TG 409 (non-rodent 90-day assay)	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
G	+	Eq/0	+	Increased evidence of (anti)-E,A,T,S activity in multiple species. May act via ER, AR, TR, S mechanism (metabolic activation may be needed).	Perform *in vitro* ER, AR, TR, S assays with added metabolising system (for the "0" scenario, otherwise Eq result available).	If existing data are from a Level 5 assay, there may be sufficient information to conclude evidence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). Check data on chemical analogues. Further mechanistic studies may help determine MOA. Consider route of exposures for effects data and possible implications of ADME characteristics of the chemical in different species. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT. Equivocal results may indicate chemical has multiple MOA.
H	+	Eq/0	–	Evidence of (anti)-E,A,T,S activity in non-rodent species. Acts via unknown mechanism or may require metabolic activation for activity. Route of exposure may account for the differences between OECD TG 409 and existing data. Unknown potential for adverse effects.	For the "0" scenario, perform *in vitro* ER, AR, TR, S assays with added metabolising system (otherwise Eq result available).	If existing data are from an adequate Level 5 assay, question why there are differences. Consider route of exposures and possible implications of ADME characteristics of the chemical. Consider species differences in physiology and response. If existing data are from a less sensitive assay, a higher level test may be required. Check data on chemical analogues. Further mechanistic studies may help determine MOA. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT. Equivocal results may indicate chemical has multiple MOA.
I	+	Eq/0	Eq/0	Evidence of (anti)-E,A,T,S activity in non-rodent species. Acts via unknown mechanism or requires metabolic activation for activity. Route of exposure may account for the differences between OECD TG 409 and existing data.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system OR Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	Check data on chemical analogues. Further mechanistic studies may help determine MOA. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT. Equivocal results may indicate chemical has multiple MOA.
J	–	+	+	No evidence for (anti)-E,A,T,S activity in non-rodent species. Weak (anti)-E,A,S activity may not be detected by this assay. Route of exposure or species differences may account for the differences between OECD TG 409 and existing data. Metabolism or potency may explain *in vitro*/*in vivo* differences.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system OR Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	If existing data are from an adequate Level 5 assay, question why there are differences. Effects seen in existing studies may be in a more sensitive species or life stage. Consider route of exposures and possible implications of ADME characteristics of the chemical in different species. Consider species differences in physiology and response. Further mechanistic studies may help determine MOA.

Scenarios	Result of OECD TG 409 (non-rodent 90-day assay)	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
K	–	+	–	No evidence for (anti)-E,A,T,S activity in multiple species. Weak (anti)-E,A,S activity may not be detected by this assay. Metabolism or potency may explain *in vitro*/*in vivo* differences. Unknown potential for adverse effects.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system.	If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from a less sensitive assay, a higher level test may be required. Further mechanistic studies with metabolism may help determine MOA. Consider species differences in physiology and response.
L	–	+	Eq/0	No evidence for (anti)-E,A,T,S activity in non-rodent species. Weak (anti)-E,A,S activity may not be detected by this assay.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system.	Metabolic deactivation of chemical may occur *in vivo* so that possible *in vitro* activity is not realised. Consider route of exposures and possible implications of ADME characteristics of the chemical in different species. Consider species differences in physiology and response. Equivocal results may indicate chemical has multiple MOA.
M	–	–	+	No evidence for (anti)-E,A,T,S activity in non-rodent species. Weak (anti)-E,A,S activity may not be detected by this assay. Effects seen in existing studies are via non-E,A,T,S mechanism.	Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	If existing data are from an adequate Level 5 assay, question why there are differences. Effects seen in existing studies may be in a more sensitive species or life stage. Consider route of exposures and possible implications of ADME characteristics of the chemical in different species. Consider species differences in physiology and response.
N	–	–	–	No evidence for (anti)-E,A,T,S activity in multiple species. Weak (anti)-E,A,S activity may not be detected by this assay. No evidence for (anti)-E,A,T,S activity *in vitro*. No evidence of adverse effects.	Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive).
O	–	–	Eq/0	No evidence for (anti)-E,A,T,S activity in non-rodent species. Weak (anti)-E,A,S activity may not be detected by this assay. No evidence for (anti)-E,A,T,S activity *in vitro*. Unknown potential for adverse effects.	Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	Consider route of exposures and possible implications of ADME characteristics of the chemical in different species.

Scenarios	Result of OECD TG 409 (non-rodent 90-day assay)	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
P	–	Eq/0	+	No evidence for (anti)-E,A,T,S activity in non-rodent species. Weak (anti)-E,A,S activity may not be detected by this assay. Potential for adverse effects via unknown mechanism.	Perform *in vitro* ER, AR,TR, S assays with added metabolising system.	Consider route of exposures and possible implications of ADME characteristics of the chemical in different species. Effects seen in existing studies may be in a more sensitive species or life stage. Further mechanistic may would strengthen weight of evidence. Equivocal results may indicate chemical has multiple MOA.
Q	–	Eq/0	–	No evidence for (anti)-E,A,T,S activity in multiple species. Weak (anti)-E,A,S activity may not be detected by this assay. No evidence of adverse effects.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system.	If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). Further mechanistic studies may strengthen weight of evidence.
R	–	Eq/0	Eq/0	No evidence for (anti)-E,A,T,S activity in non-rodent species. Weak (anti)-E,A,S activity may not be detected by this assay. Unknown potential for adverse effects.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system, otherwise Eq result available.	Further mechanistic studies may strengthen weight of evidence. Check data on chemical analogues. Equivocal results may indicate chemical has multiple MOA.

C.3.14. Two-Generation Reproduction Toxicity Study (OECD TG 416)

Status: Assay validated by the OECD.

941. Modalities detected: (anti)estrogen, (anti)androgen, thyroid, steroidogenesis.

Endpoints:

Time to mating, male fertility, female fertility, gestation length, number of implantations and corpora lutea, number of live births and post-implantation loss, litter size, sex ratio (F1, F2), litter/pup weight, pup survival index.

Estrus cyclicity (P, F1), sexual maturation (age at vaginal opening [VO] and preputial separation [PPS] (F1)), anogenital distance (F2, if triggered by changes in sex ratio or sexual maturation in F1), pup development (F1, F2).

Weights of: (P, F1) uterus, ovaries, testes, epididymides, prostate, seminal vesicles (+ coagulating glands) thyroid, adrenals.

Histopathologic changes in vagina, uterus (+ cervix), ovaries, testis, epididymis, prostate, seminal vesicles and coagulating glands.

Sperm numbers (testicular homogenisation resistant spermatids and cauda epididymides sperm reserve), sperm motility, sperm morphology (P, F1).

Background to the assay

942. The OECD Two-Generation Reproduction Toxicity Study is an apical assay designed to provide general information concerning the effects of a chemical on the male and female reproductive systems including gonadal function, the estrus cycle, mating, conception, gestation, parturition, lactation, weaning, and growth and development of the offspring. The rat is the preferred species. The recommended route of administration is oral, via the diet, by gavage or in drinking water. The study is not specifically designed to detect endocrine active substances (EASs), but has many endpoints relevant for the assessment of possible endocrine disruption and provides data on adverse effects related to reproduction and development. OECD TG 416 was revised in January 2001 to include a more comprehensive range of endpoints. These endpoints include sexual maturation (VO and PPS) which are particularly sensitive to EASs. One-generation studies and two-generation studies conducted prior to the adoption of the revised OECD TG 416 are therefore unlikely to provide as much data as studies conducted to the revised OECD TG 416, particularly with respect to endocrine disruption. They do, however, provide a great deal of useful data, particularly on adverse effects on reproduction, that may be sufficient for hazard assessment purposes even if the etiology of the effect(s) is not fully characterised.

943. The Extended One-Generation Reproduction Toxicity Study (EOGRTS – OECD TG 443) is the preferred Conceptual Framework Level 5 study for investigating potential endocrine disruption and is likely to replace OECD TG 416 in time. The EOGRTS contains

more endpoints than OECD TG 416 that are sensitive to endocrine disruption and may also include cohorts for investigating developmental neurotoxicity and developmental immunotoxicity. Decisions on which assay to use may also depend on regulatory considerations.

944. All versions of the test guideline (TG) require that parental males be dosed for a period of time encompassing at least one spermatogenic cycle and that parental females be dosed for at least several estrus cycles. Dosing is continuous during mating and throughout production of subsequent generations. The exposure of the fetus (which may be a sensitive life stage for endocrine disruption effects), the duration of dosing and the diversity of endpoints means that the revised OECD TG 416 may be considered to be more predictive for endocrine disruptor (ED-) mediated adverse effects via estrogen/androgen/thyroid/steroidogenesis (E,A,T,S) modalities than previous versions. As all the endpoints are apical, it is difficult to discern mechanism of action from this test alone. Information on mechanism of action needs to be obtained from *in vitro* E,A,T,S assays or *in vivo* lower tier tests such as the Uterotrophic Bioassay (UT) and the Hershberger Bioassay (H).

945. Although formal validation of OECD TG 416 with EDs has not taken place, studies have been published showing that estrogen receptor (ER) agonists (such as ethinylestradiol [NTP, 2010]), androgen receptor (AR) antagonists (such as vinclozolin [Matsuura et al., 2005]), steroidogenesis inhibitors (such as myclobutanil [Rockett et al., 2006]) and thyroid hormone modulators (such as propylthiouracil [Axelstad et al., 2008; NTP, 2003]) can all be detected by reproductive toxicity tests. However, OECD TG 416 does not include measurement of thyroid hormones. Thyroid endpoints are limited to thyroid weight and histopathology. In addition, OECD TG 416 lacks apical endpoints of developmental neurotoxicity, such as motor activity, sensory function, learning and memory, which are included in the EOGRTS (OECD TG 443). Endocrine modalities other than E,A,T,S are also detected (e.g. chemicals acting on the hypothalamic/pituitary/gonadal [HPG] axis or other hormone systems). Some chemicals interacting weakly with endocrine disrupting modalities in lower tier tests, designed to have greater sensitivity than specificity, may not have effects in this test as functional HPG axes in parents and offspring may allow compensation for weak effects. In these cases it could be interpreted that the weak effects do not lead to adverse outcomes in more comprehensive studies. Nonylphenol, for example, is a weak ER agonist in *in vitro* ER assays and in the *in vivo* UT assay, but has no effect on reproduction or development in reproductive tests (Tyl et al., 2006) although there were some effects on the offspring (slight changes in the estrous cycle length, the timing of VO and possibly also in ovarian weight and sperm/spermatid count, although in the absence of functional changes in reproduction at the dose levels tested). The observed perturbations in offspring were concluded (ECBI/48/99 HSE, UK) to be compatible with the predicable or hypothesised effects of exogenous estrogenic activity. Octylphenol is a further example of a weak ER agonist in *in vitro* ER binding assays and in the UT assay, but did not reveal effects on reproduction or development in a good quality test conducted according to OECD TG 416 (Tyl et al., 1999).

946. The adequacy of the protocol in studies where no endocrine-related effects are reported needs to be confirmed so that the absence of effects is not due to inadequacy of methods or reporting.

947. If the adequacy of the protocol is suspect, or the test was conducted before OECD TG 416 was revised, it may be possible to conduct or to use additional studies to support the reproductive toxicity test. For example, a reproduction toxicity study not including data

on sexual maturation could be supplemented by male and female Peripubertal Assays if needed. However, existing knowledge should be considered before embarking on further testing.

When/why the assay may be used

948. This assay forms part of the package of studies required for registration of pesticides in many jurisdictions. It may be carried out for high production volume chemicals of high concern, as well as being a more comprehensive test at Level 5 of the Conceptual Framework. It is likely to have at least three dose levels and therefore may be used for hazard identification/characterisation. OECD TG 416 has been replaced in many regions by the extended one-generation study (OECD TG 443), which is preferable for detecting endocrine disruption because it provides an evaluation of a number of endocrine endpoints in the juvenile and adult F1 which are not included in the two-generation study.

949. In order to provide information relevant for assessing whether or not a chemical may fulfil the WHO/IPCS (2002) definition of an ED, the study design has to be sufficiently robust to demonstrate the presence or absence of effects. In the dose selection, the investigator should also consider and ensure that data generated are adequate to fulfil the regulatory requirement across OECD countries as appropriate (e.g. hazard and risk assessment and labelling, ED assessment, etc.). The top dose or concentration should be sufficiently high to give clear systemic (i.e. non endocrine-specific) toxicity in order to ensure that a wide range of exposures (high to low) is tested. However, endocrine effects observed solely in the presence of clear systemic toxicity should be interpreted with caution and may be disregarded when sufficiently justified to be caused by secondary effects which are unlikely to be due to endocrine activity. The reason for this advice is a concern that some endocrine active substance (EAS) sensitive assays are being run at doses/concentrations of EASs that are too low to trigger direct impacts on the endocrine system. This guidance document is not the place to address this issue directly, but it should be considered when EAS-sensitive test guidelines (TGs) are revised in the future. In addition, the number and spacing of dose/concentration levels should also be adequate to fulfil the objectives of the study (e.g. to demonstrate dose response relationships if this is required).

Introduction to the table of scenarios

950. Table C.3.12 gives guidance on a further step to take in the event of a positive (+) or negative (-) result and in the presence of positive (+), negative (-) or equivocal/absent (Eq/0) existing results. “Existing results” are subdivided into “mechanism” and “effects” data (third and fourth columns). The table is divided horizontally into a series of scenarios that represent all the combinations of these events.

951. The results of OECD TG 416 are given in the second column. As this assay is not a screening test where a yes/no (qualitative) answer is obtained, criteria for positive results for the endocrine endpoints are not given in the test guideline. Results for the endpoints would be considered both individually and as a whole. It is not possible to provide guidance on all endpoints individually and for this test all endpoints are considered to be “apical”. Serum hormone determinations are not included in OECD TG 416, therefore (unlike with some of the other Level 4 and 5 assays) the division of the endpoints into “apical” and “indicators of hormonal activity” has not been possible.

952. For the purpose of this guidance, a positive result is defined as a biologically significant change in any of the endocrine endpoints (e.g. biologically significant reductions in reproductive organ weight). Changes in related endpoints will increase their biological significance (e.g. abnormal estrous cyclicity combined with reduced fertility).

953. A negative result for OECD TG 416 is taken to be the absence of biologically significant changes in all of the endocrine endpoints measured in this TG. Studies conducted to current standards are considered to be more predictive for absence of reproductive and developmental effects.

954. Equivocal results for the guideline are not considered in the table, partly for brevity but also because equivocal results are by nature uncertain. A decision must eventually be reached about whether the endocrine endpoints tend to be positive or negative or whether the result must be put to one side and the test repeated or supplemented by a different test. Factors which may have interfered with the result (e.g. composition of the diet used, environmental influences) should be considered.

Existing data to be considered

955. Existing "mechanism" *in vitro* data are assumed to be available from ER, AR and steroidogenesis-based assays (Level 2). Assays may also be available for interference with thyroid modalities. In practice, it is possible that data from all of these assays may not be available, so judgement will need to be used to decide which assays to perform, if any. Although the current *in vitro* test guidelines do not incorporate metabolic activation, published information on use of metabolic activation systems is available in Jacobs et al. (2008; 2013) and OECD (2008). These methods, however, have not yet been validated.

956. Existing "effects" data refer to *in vivo* effects that may come from lower level assays (e.g. UT or H Assays) (Level 3); Peripubertal (PP) Assays or OECD TG 407 assays (Level 4), or there may be longer term studies (e.g. in the case of pesticide registration packages where 90-day and carcinogenicity studies may be available). Data may also be available on effects in mammalian and non-mammalian wildlife species, although caution should be used when extrapolating between taxa. A chemical causing endocrine effects in non-mammalian environmental species (fish, for example) may also have endocrine effects in mammals, but the physiological consequences of the effects are likely to be different.

957. When considering the results of the OECD TG 416, all available data should be used in order to reach a conclusion and a weight of evidence approach taken. This may include high throughput screening data, read-across data from structural analogues and quantitative structure activity relationship (QSAR). Several QSAR models for ER and AR binding/activation are now available (see Sections B.1.1.1 and B.1.1.2).

Scenarios: Positive and negative results combined with existing data

958. The scenarios (A to R) presented in Table C.3.14 represent all the possibilities of positive or negative results in combination with the presence or absence of existing data. The action taken will also depend on the regulatory environment, but the considerations given here are generally science based. Although rats are the preferred species for OECD TG 416, the well-conserved nature of the hormonal pathways across taxa should be a strong indication that results in this assay may be relevant to other vertebrate species. Effects in laboratory mammal tests are also highly relevant for environmental mammalian species. Wherever possible, the recommended "next step which could be taken" avoids unnecessary

animal testing. However, sometimes conducting an animal test will be indicated and then the relevance of species, strain, exposure route and species-specific metabolism should always be considered. The sensitivity and physiological function of the hormone under investigation in the test species should also be considered. In general, lower level tests should be conducted before higher level tests in order to avoid unnecessary animal usage, unless it is apparent that a Level 5 test will be required anyway or will be needed to establish the evidence to conclude on ED properties. Information on some endocrine-related tumours may be detected more comprehensively in carcinogenicity studies (OECD TG 451/453) (Level 4); for example, detection of certain types of thyroid tumors in the absence of reproductive or developmental effects, as well as substances causing tumors in other endocrine-sensitive tissues. At Level 5, the Extended One-Generation Reproduction Toxiticy Study (EOGRTS – OECD TG 443) is the most sensitive reproduction assay for detecting endocrine disruption because it includes evaluation of a number of endocrine endpoints not included in the two-generation study (OECD TG 416) adopted in 2001. Further considerations specific to each scenario are given in the table.

959. Scenarios A to C represent positive results in OECD TG 416 in the presence of positive *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. A positive result in the *in vitro* assays in combination with a positive OECD TG 416 assay is strong evidence of adverse effects on reproduction/development and/or endocrine organs via E,A,T,S mechanisms. Effects on the different endpoints may assist with interpretation. In all scenarios a robust OECD TG 416 study should provide sufficient information to conclude evidence of concern for reproductive toxicity via an endocrine disruption mechanism. If the study is not considered to be robust (for reasons given in the paragraph above), then supplemental testing could be considered. Positive results in the OECD TG 416 assay may also indicate the potential for endocrine mediated effects in lower vertebrates. These could be followed up with partial life cycle tests such as the Fish Sexual Development Test (FSDT), the Larval Amphibian Growth and Development Assay (LAGDA) or the Medaka Extended One-Generation Reproduction Test (MEOGRT) if the evidence were strong enough.

960. Scenarios D to F represent positive results in OECD TG 416 in the presence of negative *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. A positive result in OECD TG 416 is strong evidence of adverse effects on reproduction/development and/or endocrine organs. Differential effects on the different endpoints may assist with interpretation. In all scenarios a robust OECD TG 416 study should provide sufficient information to conclude evidence of concern for reproductive toxicity via an endocrine disruption mechanism. If the study is not considered to be robust (for reasons given above), then supplemental testing could be considered. Positive results in the OECD TG 416 assay may also indicate the potential for endocrine mediated effects in lower vertebrates. These could be followed up with partial life cycle tests such as the FSDT, LAGDA or MEOGRT if the evidence were strong enough. Negative results in the *in vitro* assays should be viewed with caution in case a metabolite is responsible for the positive OECD TG 416 study. If the metabolic profile of the test substance is not known, then performing the *in vitro* assays with addition of a metabolising system may help to understand mechanism.

961. Scenarios G to I represent positive results in OECD TG 416 in the presence of various combinations of missing or equivocal data. The next step to take in these eventualities will depend on the nature of the other available data and the jurisdiction in which it is being used. In some cases, equivocal data may be viewed as positive whilst in others it may or may not contribute to the weight of evidence. The interpretation may also

depend on the mode of action (MOA) in question and why the data are considered equivocal, e.g. a study that is equivocal for thyroid effects may still be of value in evaluating (anti)androgenic effects. In all three scenarios, the recommended first step is to obtain reliable mechanistic (*in vitro*) data rather than proceed further with *in vivo* testing. Equivocal and missing data are alternative scenarios and two possibilities for the next step are given in most cases, but the nature of equivocal data means that decisions need to be taken on a case-by-case basis. In all cases, the role of metabolism, route of exposure and data from structural analogues should be considered before deciding on the next step. Positive results in the OECD TG 416 assay may also indicate the potential for endocrine mediated effects in lower vertebrates. These could be followed up with partial life cycle tests such as the FSDT, LAGDA or MEOGRT if the evidence were strong enough.

962. Scenarios J to L represent negative results in OECD TG 416 in the presence of positive *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. In all scenarios, a robust OECD TG 416 study may provide sufficient information to conclude absence of concern for reproductive toxicity via an endocrine disruption mechanism. If the study is not considered to be robust (for reasons given above), then supplemental testing could be considered. All three scenarios could fit a chemical that is positive in *in vitro* assays but is metabolised to a non-active metabolite, leading to negative results in OECD TG 416. This possibility may be investigated to help understand mechanism. Endocrine active potency may also explain differences between *in vitro* and *in vivo* results (e.g. a weak chemical may give a positive result *in vitro* but may be negative *in vivo*). Positive *in vivo* effects data may involve E,A,T,S or non-E,A,T,S mechanisms (e.g. involving other receptors or endocrine axes), more sensitive endpoints, greater statistical power but knowledge of absorption, distribution, metabolism and excretion (ADME) may help to explain differences from the OECD TG 416 data.

963. Scenarios M to O represent negative results in OECD TG 416 in the presence of negative *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. Negative results for all tests (Scenario N) may provide sufficient information to conclude absence of concern for reproductive toxicity via an endocrine disruption mechanism. If the study is not considered to be robust (for reasons given above), then supplemental testing could be considered. Positive *in vivo* effects data may involve E,A,T,S or non-E,A,T,S mechanisms (e.g. involving other receptors or endocrine axes), but knowledge of ADME may help to explain differences from the OECD TG 416 data.

964. Scenarios P to R represent negative results in OECD TG 416 in the presence of various combinations of missing or equivocal data. As with the positive result scenarios above (see Paragraph 920), the next step to take in these eventualities will have to be decided on a case-by-case basis. In all cases, the role of metabolism, route of exposure and data from structural analogues should be considered before deciding on the next step.

965. In all scenarios (A to R), the next step to take to strengthen weight of evidence will depend on the existing information. Table C.3.14 is meant to provide a succinct guide and may not cover all circumstances or possibilities. The scenarios may also suggest that chemicals have simple or single MOA, when in practice they may have multiple endocrine and non-endocrine MOA. In some cases, for example, two opposite modes of simultaneous action (e.g. estrogenic and anti-estrogenic) could, depending on dose, lead to a minimisation or abolition of effects, while in others two different MOA (e.g. estrogenic and anti-androgenic) could potentially reinforce effects. Endocrine pathways interact, mixed effects are common and there are many pathways that cannot be distinguished with currently available TGs. If multiple MOA are suspected, either from the existing results or based on QSAR/read-

across/integrated approaches, this should be investigated further if needed for regulatory decision making.

References

Axelstad, M. et al. (2008), "Developmental neurotoxicity of propylthiouracil (PTU) in rats: Relationship between transient hypothyroxinemia during development and long-lasting behavioural and functional changes", *Toxicology and Applied Pharmacology*, Vol. 232/1, pp. 1-13, https://doi.org/10.1016/j.taap.2008.05.020.

Jacobs, M. et al. (2013), "*In vitro* metabolism and bioavailability tests for endocrine active substances: What is needed next for regulatory purposes?", *ALTEX – Alternatives to Animal Experimentation*, Vol. 30/3, pp. 331-351.

Jacobs, M.N. et al. (2008), "The use of metabolising systems for *in vitro* testing of endocrine disrupters", *Current Drug Metabolism*, Vol. 9/8, pp. 796-826.

Matsuura, I. et al. (2005), "Evaluation of a two-generation reproduction toxicity study adding endpoints to detect endocrine disrupting activity using vinclozolin", *The Journal of Toxicological Sciences*, 30 Spec No., pp. 163-188.

NTP (2010), "Multigenerational reproductive toxicology study of ethinyl estradiol (CAS No. 57-63-6) in Sprague-Dawley rats (feed studies)", NTP TR-547, US Department of Health and Human Services, https://ntp.niehs.nih.gov/ntp/htdocs/lt_rpts/tr547.pdf.

NTP (2003), "Two-generation reproduction toxicity study of propylthiouracil (CAS#51-52-5) when administered to Sprague-Dawley rats in the drinking water", Report No. RACB20102, US Department of Health and Human Services.

OECD (2008), *Detailed Review Paper on the Use of Metabolising Systems for* In Vitro *Testing of Endocrine Disrupters*, OECD Series on Testing and Assessment, No. 97, OECD Publishing, Paris, https://doi.org/10.1787/9789264085497-en.

Rockett, J.C. et al. (2006), "Effect of conazole fungicides on reproductive development in the female rat", *Reproductive Toxicology*, Vol. 22/4, pp. 647-658, https://doi.org/10.1016/j.reprotox.2006.05.008.

Tyl, R.W. et al. (2006), "Three-generation evaluation of dietary para-nonylphenol in CD (Sprague-Dawley) rats", *Toxicological Sciences*, Vol. 92/1, pp. 295-310, https://doi.org/10.1093/toxsci/kfj203.

Tyl, R.W. et al. (1999), "Two-generation reproduction study with *para*-tert-octylphenol in rats", *Regulatory Toxicology and Pharmacology*, Vol. 30/2, pp. 81-95, https://doi.org/10.1006/rtph.1999.1343.

WHO/IPCS (2002), "Global assessment of the state-of-the-science of endocrine disrupters", Damstra, T. et al. (eds.) WHO/PCS/EDC/02.2, World Health Organization, Geneva, www.who.int/ipcs/publications/new_issues/endocrine_disruptors/en.

Table C.3.14. **Two-Generation Reproduction Toxicity Study (OECD TG 416): Guidance for scenarios of combinations of results with existing data**

This table represents possible conclusions to be drawn from assay data, and a next step which could be taken if further evidence is required about possible endocrine disrupting properties and/or effects. The guidance offered is not meant to be prescriptive, but provides science-based considerations. It encourages the use of all available data and expert judgement in a weight of evidence approach. Regional and national interpretation of results and "next steps" may vary.

The conclusions are grouped into a series of scenarios (A-R), each scenario representing a different combination of assay results, existing *in vitro* data and existing *in vivo* data. The symbol "+" indicates that the data in question represent a positive result, «-" indicates a negative result, and "Eq/0" indicates that the data are either equivocal or are not available.

Existing results: * "Mechanism (*in vitro* mechanistic data)" assumes that mechanistic data are available from estrogen receptor (ER-), androgen receptor (AR-) and steroidogenesis-based assays (Level 2). Thyroid hormone receptor (TR) and other assays concerning mechanisms of thyroid disruption may be available, but they are not in common use. In practice, data from all assays may not be available and therefore this must be taken into account when deciding on the "next step". Quantitative structure activity relationship (QSAR) predictions of estrogen and androgen binding/activation may be made for some substances.

Existing results: ** "Effects (*in vivo* effects of concern)" assumes effects have been observed in other *in vivo* screens/tests which give rise to concern that the test chemical may be an endocrine disrupter. These may be other repeated dose toxicity tests, the Uterotrophic Bioassay (UT) and Hershberger Bioassay (H), Peripubertal Assays or read-across from chemical analogues.

*** *Note:* a positive result is defined as a biologically significant change in any of the endocrine endpoints (all "apical endpoints").

Scenarios	Result of OECD TG 416 (two-generation study)	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
A	+ ***	+	+	Evidence for adverse effects via (anti)-E,A,T,S activity in OECD TG 416.		Sufficient information to conclude evidence of concern for reproductive toxicity via endocrine disruption mechanism. Note that the Extended One-Generation Reproduction Toxiticy Study (EOGRTS) provides the most information on endocrine disruption; however, for endocrine disrupting chemicals (EDCs) with a carcinogenic potential, OECD TG 451-3 may be more sensitive. Effects on apical endpoints may indicate estrogen/androgen/thyroid/steroidogenesis (E,A,T,S) modalities or other mechanisms. Consider potency of effects for existing results and whether E,A,T,S mechanism is credible for reproductive/developmental effects or whether there may be non-endocrine mechanisms. Consider route of exposures for effects data and possible implications of absorption, distribution, metabolism and excretion (ADME) characteristics of the chemical. Endocrine activity possible in lower vertebrates. Consider performing Fish Sexual Development Test (FSDT), Larval Amphibian Growth and Development Assay (LAGDA) or Medaka Extended One-Generation Reproduction Test (MEOGRT).
B	+	+	–	Evidence for adverse effects via (anti)-E,A,T,S activity in OECD TG 416.		Sufficient information to conclude evidence of concern for reproductive toxicity via endocrine disruption mechanism. Note that the EOGRTS provides the most information on endocrine disruption; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive. Effects on apical endpoints may indicate E,A,T,S modalities or other mechanisms. Consider potency of effects for existing results and whether E,A,T,S mechanism is credible for reproductive/developmental effects or whether there may be non-endocrine mechanisms. Consider route of exposures for effects data and possible implications of ADME characteristics of the chemical. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT.
C	+	+	Eq/0	Evidence for adverse effects via (anti)-E,A,T,S activity in OECD TG 416.		Sufficient information to conclude evidence of concern for reproductive toxicity via endocrine disruption mechanism. Note that the EOGRTS provides the most information on endocrine disruption; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive. Effects on apical endpoints may indicate E,A,T,S modalities or other mechanisms. Consider potency of effects for existing results and whether E,A,T,S mechanism is credible for reproductive/developmental effects or whether there may be non-endocrine mechanisms. Consider route of exposures for effects data and possible implications of ADME characteristics of the chemical. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT. Equivocal results may indicate chemical has multiple modes of action (MOA).

Scenarios	Result of OECD TG 416 (two-generation study)	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
D	+	–	+	Evidence for adverse effects in OECD TG 416 but not via E,A,T,S mechanism or may require metabolic activation for activity.	To further discern mechanism, could perform *in vitro* estrogen receptor (ER), androgen receptor (AR), thyroid hormone receptor (TR), steroidogenesis (S) assays with added metabolising system.	Sufficient information to conclude evidence of concern for reproductive toxicity via possible endocrine disruption mechanism. Note that the EOGRTS provides the most information on endocrine disruption; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive. Effects on apical endpoints may indicate E,A,T,S modalities or other mechanisms. Consider potency of effects for existing results and whether endocrine disruption mechanism is credible for reproductive/developmental effects or whether there may be non-endocrine mechanisms. Consider route of exposures for effects data and possible implications of ADME characteristics of the chemical. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT.
E	+	–	–	Evidence for adverse effects in OECD TG 416 via non-E,A,T,S/non-endocrine disruption mechanism or may require metabolic activation for activity.	To further discern mechanism, could perform *in vitro* ER, AR, TR, S assays with added metabolising system.	Sufficient information to conclude evidence of concern for reproductive toxicity via unknown mechanism. Note that the EOGRTS provides the most information on endocrine disruption; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive. Effects on apical endpoints may indicate E,A,T,S modalities or other mechanisms. Consider potency of effects for existing results and whether endocrine disruption mechanism is credible for reproductive/developmental effects or whether there may be non-endocrine mechanisms. Consider route of exposures for effects data and possible implications of ADME characteristics of the chemical. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT.
F	+	–	Eq/0	Evidence for adverse effects in OECD TG 416 via non-E,A,T,S/non-endocrine disruption mechanism or may require metabolic activation for activity.	To further discern mechanism, could perform *in vitro* ER, AR, TR, S assays with added metabolising system.	Sufficient information to conclude evidence of concern for reproductive toxicity via unknown mechanism. Note that the EOGRTS provides the most information on endocrine disruption; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive. Effects on apical endpoints may indicate E,A,T,S modalities or other mechanisms. Consider potency of effects for existing results and whether endocrine disruption mechanism is credible for reproductive/developmental effects or whether there may be non-endocrine mechanisms. Consider route of exposures for effects data and possible implications of ADME characteristics of the chemical. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT. Equivocal results may indicate chemical has multiple MOA.

Scenarios	Result of OECD TG 416 (two-generation study)	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
G	+	Eq/0	+	Evidence for adverse effects in OECD TG 416, may act via E,A,T,S mechanism and may require metabolic activation for activity.	To further discern mechanism, could perform *in vitro* ER, AR, TR, S assays with added metabolising system.	Sufficient information to conclude evidence of concern for reproductive toxicity via possible endocrine disruption mechanism. Note that the EOGRTS provides the most information on endocrine disruption; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive. Effects on apical endpoints may indicate E,A,T,S modalities or other mechanisms. Consider potency of effects for existing results and whether endocrine disruption mechanism is credible for reproductive/developmental effects or whether there may be non-endocrine mechanisms. Consider route of exposures for effects data and possible implications of ADME characteristics of the chemical. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT. Equivocal results may indicate chemical has multiple MOA.
H	+	Eq/0	–	Evidence for adverse effects in OECD TG 416 via non-E,A,T,S/non-endocrine disruption mechanism or may require metabolic activation for activity.	To further discern mechanism, could perform *in vitro* ER, AR, TR, S assays with added metabolising system.	Sufficient information to conclude evidence of concern for reproductive toxicity via unknown mechanism. Note that the EOGRTS provides the most information on endocrine disruption; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive. Effects on apical endpoints may indicate E,A,T,S modalities or other mechanisms. Consider potency of effects for existing results and whether endocrine disruption mechanism is credible for reproductive/developmental effects or whether there may be non-endocrine mechanisms. Consider route of exposures for effects data and possible implications of ADME characteristics of the chemical. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT. Equivocal results may indicate chemical has multiple MOA.
I	+	Eq/0	Eq/0	Evidence for adverse effects in OECD TG 416 via unknown mechanism.	To further discern mechanism, could perform *in vitro* ER, AR, TR, S assays with added metabolising system.	Sufficient information to conclude evidence of concern for reproductive toxicity via unknown mechanism. Consider existing results and whether endocrine disruption mechanism is credible for reproductive/developmental effects or whether there may be non-endocrine mechanisms. Consider route of exposures for effects data and possible implications of ADME characteristics of the chemical. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT. Equivocal results may indicate chemical has multiple MOA.

Scenarios	Result of OECD TG 416 (two-generation study)	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
J	–	+	+	No evidence of adverse effects on reproduction/ development/endocrine organs. Effects seen in existing (lower level) studies do not lead to adverse outcome in Level 5 assay. Metabolism or potency may explain the difference from existing *in vitro* and *in vivo* data.	If test is to current OECD TG 416 standards, maybe no further testing needed. If not, consider supplemental testing, depending on existing data. To further discern mechanism, could perform *in vitro* ER, AR, TR, S assays with added metabolising system.	If existing data are from other, adequate, apical studies, question why there are differences. Note that the EOGRTS provides the most information on endocrine disruption; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive. Consider route of exposures and possible implications for ADME characteristics of the chemical with existing studies. Further mechanistic studies with metabolism may help determine MOA.
K	–	+	–	No evidence of adverse effects on reproduction/ development/endocrine organs. Metabolism or potency may explain *in vitro*/*in vivo* differences.	If test is to current OECD TG 416 standards, maybe no further testing needed. If not, consider supplemental testing, depending on existing data. To further discern mechanism, could perform *in vitro* ER, AR, TR, S assays with added metabolising system.	If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption. Further mechanistic studies with metabolism may help determine MOA.
L	–	+	Eq/0	No evidence of adverse effects on reproduction/ development/endocrine organs. Metabolism or potency may explain *in vitro*/*in vivo* differences. Effects seen in existing (lower level) studies do not lead to adverse outcome in Level 5 assay.	If test is to current OECD TG 416 standards, maybe no further testing needed. If not, consider supplemental testing, depending on existing data. To further discern mechanism, could perform *in vitro* ER, AR, TR, S assays with added metabolising system.	There may be sufficient information to conclude absence of concern for endocrine disruption. Note that the EOGRTS provides the most information on endocrine disruption; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive. Consider route of exposures and possible implications for ADME characteristics of the chemical with existing studies. Further mechanistic studies with metabolism may help determine MOA. Equivocal results may indicate chemical has multiple MOA.

Scenarios	Result of OECD TG 416 (two-generation study)	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
M	–	–	+	No evidence of adverse effects on reproduction/ development/endocrine organs. Effects seen in existing (lower level) studies do not lead to adverse outcome in Level 5 assay.	If test is to current OECD TG 416 standards, maybe no further testing needed. If not, consider supplemental testing, depending on existing data.	If existing data are from adequate *in vivo* studies such as 28-day, 90-day, chronic/carcinogenicity studies, question why there are differences. Note that the EOGRTS provides the most information on endocrine disruption; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive. Consider route of exposures and possible implications for ADME characteristics of the chemical with existing studies. Further mechanistic studies with metabolism may help determine MOA.
N	–	–	–	No evidence of adverse effects on reproduction/ development/endocrine organs.	If test is to current OECD TG 416 standards, maybe no further testing needed. If not, consider supplemental testing, depending on existing data.	If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption. Note that the EOGRTS provides the most information on endocrine disruption; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive.
O	–	–	Eq/0	No evidence of adverse effects on reproduction/ development/endocrine organs. No evidence for (anti)-E,A,T,S activity *in vitro*.	If test is to current OECD TG 416 standards, maybe no further testing needed. If not, consider supplemental testing, depending on existing data. To further discern mechanism, could perform *in vitro* ER, AR, TR, S assays with added metabolising system.	There may be sufficient information to conclude absence of concern for endocrine disruption. Note that the EOGRTS provides the most information on endocrine disruption; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive. Further mechanistic studies with metabolism may help determine MOA. Consider route of exposures and possible implications for ADME characteristics of the chemical with existing studies. Check data on chemical analogues.
P	–	Eq/0	+	No evidence of adverse effects on reproduction/ development/endocrine organs. Effects seen in existing (lower level) studies do not lead to adverse outcome in Level 5 assay. Effects seen in existing studies are via unknown mechanism.	If test is to current OECD TG 416 standards, maybe no further testing needed. If not, consider supplemental testing, depending on existing data. To further discern mechanism, could perform *in vitro* ER, AR, TR, S assays with added metabolising system.	If existing data are from adequate *in vivo* studies such as 28-day, 90-day, chronic/carcinogenicity studies, question why there are differences. Note that the EOGRTS provides the most information on endocrine disruption; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive. Consider route of exposures and possible implications for ADME characteristics of the chemical with existing studies. Check data on chemical analogues. Equivocal results may indicate chemical has multiple MOA.

Scenarios	Result of OECD TG 416 (two-generation study)	Existing results		Possible conclusions	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
Q	–	Eq/0	–	No evidence of adverse effects on reproduction/ development/endocrine organs.	If test is to current OECD TG 416 standards, maybe no further testing needed. If not, consider supplemental testing, depending on existing data.	There may be sufficient information to conclude absence of concern for endocrine disruption. Check data on chemical analogues.
R	–	Eq/0	Eq/0	No evidence of adverse effects on reproduction/ development/endocrine organs.	If test is to current OECD TG 416 standards, maybe no further testing needed. If not, consider supplemental testing, depending on existing data.	There may be sufficient information to conclude absence of concern for endocrine disruption. Note that the EOGRTS provides the most information on endocrine disruption; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive. Further mechanistic studies may strengthen weight of evidence. Check data on chemical analogues. Equivocal results may indicate chemical has multiple MOA.

C.3.15. Extended One-Generation Reproductive Toxicity Study (EOGRTS) (OECD TG 443)

Status: Assay validated by the OECD.

966. Modalities detected: (anti)estrogen, (anti)androgen, thyroid, steroidogenesis.

Endpoints: Time to mating, male fertility, female fertility, dystocia, gestation length, number of implantations and corpora lutea, number of ovarian follicles, number of live births and post-implantation loss, viability index, litter size, sex ratio, litter/pup weight, pup survival index, placental weight, anogenital distance, presence of nipples, pup development including genitals (and presence of abnormalities), sexual maturation (age at vaginal opening and preputial separation) (F1).

Weights and/or histopathologic analysis: uterus (with oviducts and cervix), ovaries, testes, epididymides, prostate, seminal vesicles (+ coagulating glands) thyroid, adrenals, pituitary, mammary gland (P and F1).

Estrus cyclicity (P and F1).

Sperm numbers (testicular homogenization-resistant spermatids and cauda epididymal sperm reserves), sperm motility, sperm morphology (P and F1).

Hormones: T4, TSH (P and F1).

Apical endpoints from the developmental neuro- and immunotoxicity cohorts may be sensitive to endocrine modulation:

- Developmental neurotoxicity (DNT) endpoints: Auditory startle, functional observation, motor activity. Brain weight and histopathological examination.
- Developmental immunotoxicity (DIT) endpoints: Splenic lymphocyte subpopulation analysis, lymph node weight and histopathology, primary IgM antibody response to a T cell dependant antigen.

Background to the assay

967. The Extended One-Generation Reproduction Toxicity Study is an apical assay designed to evaluate specific life stages not covered by other tests and to test for effects that may occur as a result of pre- and post-natal exposure to chemicals. It is based on the proposal of Cooper et al. (2006) and includes three possible cohorts of F1 animals:

1. to assess reproductive/developmental endpoints
2. to assess effects on the developing nervous system
3. to assess effects on the developing immune system.

968. The reproductive/developmental element of the study provides a thorough evaluation of systemic, reproductive and developmental toxicity including gonadal function, the estrus cycle, epididymal sperm maturation, mating, conception, gestation, parturition, lactation, weaning, and growth and development of the offspring. The rat is the preferred

species. The assay is designed for administration of the test substance via the diet, but it can be modified for administration by other routes (drinking water, gavage, inhalation, dermal). Depending on the modules carried out in the test, effects on the developing nervous and immune systems are also assessed. These systems may also be sensitive to endocrine influences. OECD TG 443 was adopted in July 2011 (Figure 1 was corrected in October 2012 when it was found to be incorrect). The study uses fewer animals than OECD TG 416 (Two-Generation Reproduction Toxicity Study) when F2 is omitted, whilst increasing the number of pups studied in the F1 generation and the number of endpoints. Inclusion of an F2 generation may be "triggered". Triggering in other jurisdictions may depend on results obtained in the F1 generation. OECD GD 151 (2013) supports OECD TG 443, providing advice on study design including the gathering of key data on the substance to be tested, endpoints and data interpretation issues not fully covered in the test guideline (TG). Decisions on whether to assess the second generation or omit the developmental neurotoxicity or developmental immunotoxicity have to be taken on a case-by-case basis depending on existing knowledge and regulatory purpose. As endocrine activity can affect the developmental brain/immune system, inclusion of DNT and DIT cohorts in OECD TG 443 provides a more comprehensive study. The procedure and internal triggers for deciding whether a second generation should be produced are described in OECD GD 117 (2011) for those regulations under which internal triggering applies. Criteria for the study design for the EU REACH regulation is described in ECHA (2017). As the second generation is "triggered", then at present OECD TG 416 is the only OECD mammalian test that automatically covers two generations.

969. The EOGRTS was not specifically designed to detect endocrine active substances (EASs) but Cohort 1 has many endpoints relevant for the assessment of possible endocrine disruption, for example endpoints such as sexual maturation and estrous cyclicity are particularly sensitive to estrogens and androgens. Effects on the thyroid and thyroid hormones are also detected by serum T4 and TSH levels, thyroid weight and by histopathology in P and F1 generations. The assay also provides data on adverse effects related to reproduction and development which may or may not be related to endocrine disruption. Cohorts 2 and 3 also have apical endpoints that may be sensitive to endocrine modulation. The developing brain, for example, is a classical target of thyroid hormones whilst interaction of chemicals with the hypothalamic-pituitary-adrenal axis may affect both the developing immune and nervous systems (Goel et al., 2014). Motoric activity is the only sexually dimorphic behaviour included in the mandatory investigations of the DNT Cohorts 2A and 2B. Also, sexually dimorphic behaviour can be affected by exposure to compounds disrupting the hypothalamic/pituitary/gonadal (HPG) axis (Hotchkiss et al., 2002; Schantz and Widholm, 2001; Weiss, 2002).

970. Experience with of serum hormone determinations in Level 4 and 5 rodent assays has revealed that their detection/measurement in rodent studies can be challenging. A recent workshop on "Practicability of Hormonal Measurements" was organised by the BfR (Germany) and the finding from this workshop will be published (Kucheryavenko et al., 2018). The OECD Expert Group on Reproductive and Developmental Toxicity recommends that to demonstrate proficiency for thyroid hormones measurement, a laboratory should be able to show results from a separate study using a positive control substance. Laboratories may also submit their calibration curves, standard curves, as well as data on the levels of quantification and detection. This group is also establishing a historical control database with thyroid toxicant positive controls.

971. The EOGRTS is preferable for detecting endocrine disruption because it provides an evaluation of a number of endocrine endpoints in the juvenile and adult F1, which are not included in the two-generation study (OECD TG 416) adopted in 2001. For example, anogenital distance (AGD) and nipple retention, which are measured in all offspring in this guideline, are clear and sensitive markers of endocrine disruption (anti-androgenic action), especially when both endpoints are affected.

972. Anogenital distance and nipple retention are sensitive endpoints of endocrine effects; however, their utility as apical endpoints or as biological indicators of endocrine action may require further experience in their use. Increased nipple retention and reduced AGD in male offspring are hallmarks of anti-androgenicity. Nevertheless, "retained nipples/areolae" as a qualitative endpoint may have high biological variability (e.g. Melching-Kollmuss et al. [2017]), but nipple retention is a sensitive endpoint if measured quantitatively (i.e. if the number of nipples from 0 to 12 is recorded). Alteration of AGD can occur via other modes of action (e.g. Miyagawa et al. [2011]; Seifert et al. [2009]). However, current OECD guidance on these endpoints can be found in OECD GD 43 and GD 151 and it is clear that these should be considered as apical endpoints. With regard to anogenital distance, OECD GD 43 (OECD, 2008b) states, "A statistically significant change in [anogenital distance] that cannot be explained by the size of the animal indicates effects of the exposure and should be used for setting the [no observed adverse effect level (NOAEL)]". With regard to nipple retention OECD GD 151 (OECD, 2013) states "a statistically significant change in nipple retention should be evaluated similarly to an effect on anogenital distance as both endpoints indicate an adverse effect of exposure and should be considered in setting a NOAEL".

973. Dosing is continuous, prior to and during mating, and throughout production of the subsequent generation(s). The exposure of the fetus (which is a sensitive life stage for endocrine disruption effects), the long duration of dosing and the diversity of endpoints means that the EOGRTS may be considered to be the most predictive test for endocrine disruption mediated adverse effects via estrogen/androgen/thyroid/steroidogenesis (E,A,T,S) modalities. As most of the endpoints are apical, it is difficult to discern mechanism of action from this test alone. Information on mechanism of action needs to be obtained from *in vitro* E,A,T,S assays or *in vivo* lower tier tests such as a Uterotrophic Bioassay (UT) and Hershberger Bioassay (H). Beekhuijzen et al. (2016) and Moore et al. (2016) also provide recent practical guidance on this assay from specific laboratories.

When/why the assay may be used

974. The EOGRTS has replaced OECD TG 416 in many regions. As an alternative to OECD TG 416, it may form part of the package of studies required for registration of pesticides and biocides. It has now replaced OECD TG 416 as part of the standard information requirements in certain chemical legislations (e.g. REACH [ECHA, 2017]). It may also be carried out for high production volume chemicals of high concern as well as being the most comprehensive test at Level 5 of the Conceptual Framework. It is likely to have at least three dose levels and therefore may be used for hazard identification/characterisation.

975. In order to provide information relevant for assessing whether or not a chemical may fulfil the WHO/IPCS (2002) definition of an endocrine disruptor (ED), the study design has to be sufficiently robust to demonstrate the presence or absence of effects. In the dose selection, the investigator should also consider and ensure that data generated are adequate to fulfil the regulatory requirement across OECD countries as appropriate

(e.g. hazard and risk assessment and labelling, ED assessment, etc.). The top dose or concentration should be sufficiently high to give clear systemic (i.e. non endocrine-specific) toxicity in order to ensure that a wide range of exposures (high to low) is tested. However, endocrine effects observed solely in the presence of clear systemic toxicity should be interpreted with caution and may be disregarded when sufficiently justified to be caused by secondary effects which are unlikely to be due to endocrine activity. The reason for this advice is a concern that some endocrine active substance (EAS) sensitive assays are being run at doses/concentrations of EASs that are too low to trigger direct impacts on the endocrine system. This guidance document is not the place to address this issue directly, but it should be considered when EAS-sensitive test guidelines (TGs) are revised in the future. In addition, the number and spacing of dose/concentration levels should also be adequate to fulfil the objectives of the study (e.g. to demonstrate dose response relationships if this is required).

Introduction to the table of scenarios

976. Table C.3.15 gives guidance on a further step to take in the event of a positive (+) or negative (-) result and in the presence of positive (+), negative (-) or equivocal/absent (Eq/0) existing results. "Existing results" are subdivided into "mechanism" and "effects" data (third and fourth columns). The table is divided horizontally into a series of scenarios that represent all the combinations of these events.

977. The results of the EOGRTS are given in the second column. As this assay is not a screening test where a yes/no (qualitative) answer is obtained, criteria for positive results for the endocrine endpoints are not given in the TG. Results for the endpoints would be considered both individually and as a whole. It is not possible to provide guidance on all endpoints individually and therefore the endpoints have been pragmatically divided into "apical" and "indicators of hormonal activity". The terminology used has been chosen to be consistent between both the non-mammalian and mammalian tests. Both groups have similar biological importance, although the "indicators of hormonal activity" in the mammalian assays are serum hormones and are generally, but not always, more variable than "apical endpoints".

978. For this guideline "apical" endpoints are reproductive, developmental and immunological parameters, including, for example, anogenital distance, presence of nipples, genital abnormalities, sexual maturation, sperm parameters, estrous cyclicity, weights and histopathologic changes in sex organs and thyroid gland. Apical endpoints indicative of DNT and DIT are included because of their possible association with thyroid and sex hormone perturbation. "Indicators of hormonal activity" are hormones (T4, TSH).

979. Three possible outcomes for a positive result are therefore envisaged in Table C.3.15:

1. indicators of hormonal activity and apical endpoints positive
2. indicators of hormonal activity positive and apical endpoints negative
3. indicators of hormonal activity negative and apical endpoints positive.

980. A positive result for apical endpoints could be biologically significant changes in pup AGD, accompanied by treatment-related histopathologic changes in parental reproductive organs or decreased fertility. A positive result for indicators of hormonal activity could be biologically significant changes in hormone profiles. A positive result for indicators of hormonal activity alone should be considered with caution, although it is

possible that these endpoints may have detected weak effects that were not detected by the apical endpoints.

981. A negative result for the EOGRTS is taken to be the absence of changes in both endocrine-relevant indicators of hormonal activity and apical endpoints. A well-conducted study is considered to be more predictive for absence of reproductive and developmental effects and for endocrine disruptive effects mediated through E,A,T,S modalities.

982. Equivocal results for the guideline are not considered in the table, partly for brevity but also because equivocal results are by nature uncertain. A decision must eventually be reached about whether the endocrine endpoints tend to be positive or negative or whether the result must be put to one side and the test repeated or supplemented by a different test. Factors which may have interfered with the result (e.g. composition of the diet used, environmental influences) should be considered.

Existing data to be considered

983. Existing "mechanism" *in vitro* data are assumed to be available from estrogen receptor (ER-), androgen receptor (AR-) and steroidogenesis-based assays (Level 2). Assays may also be available for interference with thyroid modalities. In practice, it is possible that data from all of these assays may not be available, so judgement will need to be used to decide which assays to perform. Although the current *in vitro* test guidelines do not incorporate metabolic activation, published information on use of metabolic activation systems is available in Jacobs et al. (2008; 2013) and OECD (2008a). These methods, however, have not yet been validated.

984. Existing "effects" data refer to *in vivo* effects that may come from lower level assays (e.g. UT or H Assays) (Level 3); Peripubertal (PP) Assays or OECD TG 407 assays (Level 4), or there may be longer term studies (e.g. in the case of pesticide registration packages where 90-day and carcinogenicity studies may be available). Data may also be available on effects in mammalian and non-mammalian wildlife species, although caution should be used when extrapolating between taxa. A chemical causing endocrine effects in non-mammalian environmental species (fish, for example) may also have endocrine effects in mammals, but the physiological consequences of the effects are likely to be different.

985. When considering the results of the H assay, all available data should be used in order to reach a conclusion and a weight of evidence approach taken. This may include high throughput screening data, read-across data from structural analogues and quantitative structure activity relationship (QSAR). Several QSAR models for ER and AR binding/activation are now available (see Sections B.1.1.1 and B.1.1.2).

Scenarios: Positive and negative results combined with existing data

986. A series of scenarios (A to R) are presented in Table C.3.15 and represent all the possibilities of positive or negative results in combination with the presence or absence of existing data. The action taken will also depend on the regulatory environment, but the considerations given here are generally science based. Although rats are the preferred species for OECD TG 443, the well-conserved nature of the hormonal pathways across taxa should be a strong indication that results in this assay may be relevant to other vertebrate species. Effects in laboratory mammal tests are also highly relevant for environmental mammalian species. Wherever possible, the recommended "next step which could be taken" avoids unnecessary animal testing. However, sometimes conducting an

animal test will be indicated and then the relevance of species, strain, exposure route and species-specific metabolism should always be considered. The sensitivity and physiological function of the hormone under investigation in the test species should also be considered. In general, lower level tests should be conducted before higher level tests in order to avoid unnecessary animal usage. Further considerations specific to each scenario are given in the table.

987. Scenarios A to C represent positive results in the EOGRTS in the presence of positive *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. Each positive EOGRTS result scenario is divided into the three possible outcomes given above. A positive result in the *in vitro* assays in combination with a positive EOGRTS is strong evidence of adverse effects on reproduction/development and/or endocrine organs via E,A,T,S mechanisms. Effects on the apical endpoints or indicators of hormonal activity may assist with interpretation. In all scenarios, a robust EOGRTS should provide sufficient information to conclude evidence of concern for reproductive toxicity via an endocrine disruption mechanism. If the study is not considered to be robust, then supplemental testing could be considered. Positive results in the OECD TG 443 assay may also indicate the potential for endocrine mediated effects in lower vertebrates. These could be followed up with partial life cycle tests such as the Fish Sexual Development Test (FSDT), the Larval Amphibian Growth and Development Assay (LAGDA) or the Medaka Extended One-Generation Reproduction Test (MEOGRT) if the evidence were strong enough.

988. Scenarios D to F represent positive results in the EOGRTS in the presence of negative *in vitro* mechanistic data and positive, negative or equivocal in vivo effects data. Each positive result scenario is divided into the three possible outcomes given above. A positive result in the EOGRTS is strong evidence of adverse effects on reproduction/development and/or endocrine organs. Differential effects on the different apical endpoints or indicators of hormonal activity may assist with interpretation. In all scenarios, a robust EOGRTS should provide sufficient information to conclude evidence of concern for reproductive toxicity via an endocrine disruption mechanism. If the study is not considered to be robust, then supplemental testing could be considered. Negative results in the *in vitro* assays should be viewed with caution in case a metabolite is responsible for the positive extended one-generation study. If the metabolic profile of the test substance is not known, then performing the *in vitro* assays with addition of a metabolising system may help to understand mechanism. Positive results in the OECD TG 443 assay may also indicate the potential for endocrine mediated effects in lower vertebrates. These could be followed up with partial life cycle tests such as the FSDT, LAGDA or MEOGRT if the evidence were strong enough.

989. Scenarios G to I represent positive results in the EOGRTS in the presence of various combinations of missing or equivocal data. Each positive result scenario is divided into the three possible outcomes given above. The next step to take in these eventualities will depend on the nature of the other available data and the jurisdiction in which it is being used. In some cases, equivocal data may be viewed as positive whilst in others it may or may not contribute to the weight of evidence. The interpretation may also depend on the mode of action (MOA) in question and why the data are considered equivocal, e.g. a study that is equivocal for thyroid effects may still be of value in evaluating (anti)androgenic effects. In all three scenarios, the recommended first step is to obtain reliable mechanistic (*in vitro*) data rather than proceed further with *in vivo* testing. Equivocal and missing data are alternative scenarios and two possibilities for the next step are given in most cases, but the nature of equivocal data means that decisions need to be taken on a case-by-case basis. In all cases, the role of metabolism, route of exposure and data from structural analogues

should be considered before deciding on the next step. Positive results in the OECD TG 443 assay may also indicate the potential for endocrine mediated effects in lower vertebrates. These could be followed up with partial life cycle tests such as the FSDT, LAGDA or MEOGRT if the evidence were strong enough.

990. Scenarios J to L represent negative results in the EOGRTS in the presence of positive *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. As a negative result is taken to be negative findings for both indicators of hormonal activity and apical endpoints (unlike the situation with positive outcomes), there is only one possible negative outcome. In all scenarios, a robust EOGRTS may provide sufficient information to conclude absence of concern for reproductive toxicity via an endocrine disruption mechanism. If the study is not considered to be robust, supplemental testing could be considered. All three scenarios could fit a chemical that is positive in *in vitro* assays but is metabolised to a non-active metabolite, leading to negative results in the extended one-generation study. This possibility may be investigated to help understand mechanism. Endocrine active potency may also explain differences between *in vitro* and *in vivo* results (e.g. a weak chemical may give a positive result in vitro but may be negative *in vivo*). Positive *in vivo* effects data may involve E,A,T,S or non-E,A,T,S mechanisms (e.g. involving other receptors or endocrine axes), more sensitive endpoints or greater statistical power, but knowledge of absorption, distribution, metabolism and excretion (ADME) may help to explain differences from the EOGRTS data.

991. Scenarios M to O represent negative results in the EOGRTS in the presence of negative *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. Negative results for all tests (Scenario N) may provide sufficient information to conclude absence of concern for reproductive toxicity via an endocrine disruption mechanism. If the study is not considered to be robust, then supplemental testing could be considered. Positive *in vivo* effects data may involve E,A,T,S or non-E,A,T,S mechanisms (e.g. involving other receptors or endocrine axes), but knowledge of ADME may help to explain differences from the EOGRTS data.

992. Scenarios P to R represent negative results in the EOGRTS in the presence of various combinations of missing or equivocal data. As with the positive result scenarios above (see Paragraph 990), the next step to take in these eventualities will have to be decided on a case-by-case basis. In all cases, the role of metabolism, route of exposure and data from structural analogues should be considered before deciding on the next step.

993. In all scenarios (A to R), the next step to take to strengthen weight of evidence will depend on the existing information. Table C.3.15 is meant to provide a succinct guide and may not cover all circumstances or possibilities. The scenarios may also suggest that chemicals have simple or single MOA, when in practice they may have multiple endocrine and non-endocrine MOA. In some cases, for example, two opposite modes of simultaneous action (e.g. estrogenic and anti-estrogenic) could, depending on dose, lead to a minimisation or abolition of effects, while in others two different MOA (e.g. estrogenic and anti-androgenic) could potentially reinforce effects. Endocrine pathways interact, mixed effects are common and there are many pathways that cannot be distinguished with currently available TGs. If multiple MOA are suspected, either from the existing results or based on QSAR/read-across/integrated approaches, this should be investigated further if needed for regulatory decision making.

References

Beekhuijzen, M. et al. (2016), "Implementing the Extended One-Generation Reproductive Toxicity Study (EOGRTS): Important points to consider", *Critical Reviews in Toxicology*, Vol. 46/4, pp. 332-347, https://doi.org/10.3109/10408444.2015.1137863.

Cooper, R.L. et al. (2006), "A tiered approach to life stages testing for agricultural chemical safety assessment", *Critical Reviews in Toxicology*, Vol. 36/1, pp. 69-98, https://doi.org/10.1080/10408440500541367.

ECHA (2017), "R.7.6 Reproductive toxicity", in: *Guidance on Information Requirements and Chemical Safety Assessment: Chapter R.7a: Endpoint Specific Guidance (Version 6.0, July 2017)*, European Chemicals Agency, Helsinki, https://echa.europa.eu/documents/10162/13632/information_requirements_r7a_en.pdf.

Goel, N. et al. (2014), "Sex differences in the HPA axis", *Comprehensive Physiology*, Vol. 4/3, pp. 1121-1155, http://dx.doi.org/10.1002/cphy.c130054.

Hotchkiss, A. et al. (2002), "Androgens and environmental antiandrogens affect reproductive development and play behavior in the Sprague-Dawley rat", *Environmental Health Perspectives*, Vol. 110, Supplement 3, pp. 435-439, www.jstor.org/stable/3455400.

Jacobs, M. et al. (2013), "*In vitro* metabolism and bioavailability tests for endocrine active substances: What is needed next for regulatory purposes?", *ALTEX – Alternatives to Animal Experimentation*, Vol. 30/3, pp. 331-351.

Jacobs, M.N. et al. (2008), "The use of metabolising systems for *in vitro* testing of endocrine disrupters", *Current Drug Metabolism*, Vol. 9/8, pp. 796-826.

Kucheryavenko, O. et al. (2018), "Report from the BfR Expert Hearing on Practicability of Hormonal Measurements", *Archives of Toxicology* (in prep.).

Melching-Kollmuss, S. et al. (2017), "Comparing effect levels of regulatory studies with endpoints derived in targeted anti-androgenic studies: Example prochloraz", *Archives of Toxicology*, Vol. 91/1, pp. 143-162, https://doi.org/10.1007/s00204-016-1678-y.

Miyagawa, S. et al. (2011), "The role of sonic hedgehog-Gli2 pathway in the masculinization of external genitalia", *Endocrinology*, Vol. 152/7, pp. 2894-2903, https://doi.org/10.1210/en.2011-0263.

Moore, N.P. et al. (2016), "Guidance on the selection of cohorts for the Extended One-Generation Reproductive Toxicity Study (OECD Test Guideline 443)", *Regulatory Toxicology and Pharmacology*, Vol. 80, pp. 32-40, https://doi.org/10.1016/j.yrtph.2016.05.036.

OECD (2013), "Guidance document supporting of OECD test guideline on the Extended One-Generation Reproductive Toxicity Study", OECD Series on Testing and Assessment, No. 151, OECD, Paris, www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/JM/MONO(2013)10&doclanguage=en.

OECD (2011), "Guidance document on the current implementation of internal triggers in Test Guideline 443 for an Extended One-Generation Reproductive Toxicity Study in the United States and Canada", OECD Series on Testing and Assessment, No. 117, OECD, Paris, www.oecd.org/env/ehs/testing/48516094.pdf.

OECD (2008a), *Detailed Review Paper on the Use of Metabolising Systems for* In Vitro *Testing of Endocrine Disrupters*, OECD Series on Testing and Assessment, No. 97, OECD Publishing, Paris, https://doi.org/10.1787/9789264085497-en.

OECD (2008b), "Guidance document on mammalian reproductive toxicity testing and assessment", OECD Series on Testing and Assessment, No. 43, OECD, Paris, www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/JM/MONO(2008)16&doclanguage=en.

Schantz, S.L. and J.J. Widholm (2001), "Cognitive effects of endocrine-disrupting chemicals in animals", *Environmental Health Perspectives*, Vol. 109/12, pp. 1197-1206.

Seifert, A.W. et al. (2009), "Multiphasic and tissue-specific roles of sonic hedgehog in cloacal septation and external genitalia development", *Development*, Vol. 136/23, pp. 3949-3957, http://dx.doi.org/10.1242/dev.042291.

Weiss, B. (2002), "Sexually dimorphic nonreproductive behaviors as indicators of endocrine disruption", *Environmental Health Perspectives*, Vol. 110, Supplement 3, pp. 387-391.

WHO/IPCS (2002), "Global assessment of the state-of-the-science of endocrine disrupters", Damstra, T. et al. (eds.) WHO/PCS/EDC/02.2, World Health Organization, Geneva, www.who.int/ipcs/publications/new_issues/endocrine_disruptors/en.

Table C.3.15. **Extended One-Generation Reproductive Toxicity Study (EOGRTS) (OECD TG 443): Guidance for scenarios of combinations of results with existing data**

This table represents possible conclusions to be drawn from assay data, and a next step which could be taken if further evidence is required about possible endocrine disrupting properties and/or effects. The guidance offered is not meant to be prescriptive, but provides science-based considerations. It encourages the use of all available data and expert judgement in a weight of evidence approach. Regional and national interpretation of results and "next steps" may vary.

The conclusions are grouped into a series of scenarios (A-R), each scenario representing a different combination of assay results, existing *in vitro* data and existing *in vivo* data. The symbol "+" indicates that the data in question represent a positive result, «-" indicates a negative result, and "Eq/0" indicates that the data are either equivocal or are not available.

Existing results: * "Mechanism (*in vitro* mechanistic data)" assumes that mechanistic data are available from estrogen receptor (ER-), androgen receptor (AR-) and steroidogenesis-based assays (Level 2). Thyroid hormone receptor (TR) and other assays concerning mechanisms of thyroid disruption may be available, but they are not in common use. In practice, data from all assays may not be available and therefore this must be taken into account when deciding on the "next step". Quantitative structure activity relationship (QSAR) predictions of estrogen and androgen binding/activation may be made for some substances.

Existing results: ** "Effects (*in vivo* effects of concern)" assumes effects have been observed in other *in vivo* screens/tests which give rise to concern that the test chemical may be an endocrine disrupter. These may be other repeated dose toxicity tests, Uterotrophic Bioassays (UT) and Hershberger Bioassays (H), Peripubertal Assays or read-across from chemical analogues.

*** *Note:* three possible outcomes for a positive result are given:

1. indicators of hormonal activity and apical endpoints positive
2. indicators of hormonal activity positive and apical endpoints negative
3. indicators of hormonal activity negative and apical endpoints positive.

"Apical endpoints" are reproductive and developmental parameters (including anogenital distance, presence of nipples, genital abnormalities), sexual maturation, sperm parameters, estrous cyclicity, weights and histopathologic changes in testes, epididymides, prostate, seminal vesicles (with coagulating glands), ovary, uterus (with oviducts and cervix), thyroid. Apical endpoints from the developmental neuro- and immunotoxicity cohorts may also be sensitive to endocrine modulation.

"Indicators of hormonal activity" are hormones (T4, TSH).

Scenarios	Result of EOGRTS	Existing results		Possible conclusions: 1) Indicators of hormonal activity and apical endpoints positive 2) Indicators of hormonal activity positive and apical endpoints negative 3) Indicators of hormonal activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
A	+ ***	+	+	1) Evidence for adverse effects via (anti)-E,A,T,S activity in OECD TG 443. 2) Evidence for adverse effects via (anti)-E,A,T,S activity in OECD TG 443. Apical endpoints may be less sensitive or unaffected. 3) Evidence for adverse effects via (anti)-E,A,T,S activity in OECD TG 443. Indicators of hormonal activity may be less sensitive or unaffected.		Sufficient information to conclude evidence of concern for reproductive toxicity via endocrine disruption mechanism. Effects on indicators of hormonal activity alone may be indicative of changes not detected by apical endpoints. Effects on apical endpoints alone may indicate estrogen/androgen/thyroid/steroidogenesis (E,A,T,S) modalities or other mechanisms. Consider potency of effects for existing results and whether E,A,T,S mechanism is credible for reproductive/developmental effects or whether there may be non-endocrine mechanisms. Consider route of exposures for effects data and possible implications of absorption, distribution, metabolism and excretion (ADME) characteristics of the chemical. Hormonal activity possible in lower vertebrates. Consider performing a Fish Sexual Development Test (FSDT), Larval Amphibian Growth and Development Assay (LAGDA) or Medaka Extended One-Generation Reproduction Test (MEOGRT).
B	+	+	–	1) Evidence for adverse effects via (anti)-E,A,T,S activity in OECD TG 443. 2) Evidence for adverse effects via (anti)-E,A,T,S activity in OECD TG 443. Apical endpoints may be less sensitive or unaffected. 3) Evidence for adverse effects via (anti)-E,A,T,S activity in OECD TG 443. Indicators of hormonal activity may be less sensitive or unaffected.		Sufficient information to conclude evidence of concern for reproductive toxicity via endocrine disruption mechanism. Effects on indicators of hormonal activity alone may be indicative of changes not detected by apical endpoints. Effects on apical endpoints alone may indicate E,A,T,S modalities or other mechanisms. Consider potency of effects for existing results and whether E,A,T,S mechanism is credible for reproductive/developmental effects or whether there may be non-endocrine mechanisms. Consider route of exposures for effects data and possible implications of ADME characteristics of the chemical. Hormonal activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT.

Scenarios	Result of EOGRTS	Existing results		Possible conclusions: 1) Indicators of hormonal activity and apical endpoints positive 2) Indicators of hormonal activity positive and apical endpoints negative 3) Indicators of hormonal activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
C	+	+	Eq/0	1) Evidence for adverse effects via (anti)-E,A,T,S activity in OECD TG 443. 2) Evidence for adverse effects via (anti)-E,A,T,S activity in OECD TG 443. Apical endpoints may be less sensitive or unaffected. 3) Evidence for adverse effects via (anti)-E,A,T,S activity in OECD TG 443. Indicators of hormonal activity may be less sensitive or unaffected.		Sufficient information to conclude evidence of concern for reproductive toxicity via endocrine disruption mechanism. Effects on indicators of hormonal activity alone may be indicative of changes not detected by apical endpoints. Effects on apical endpoints alone may indicate E,A,T,S modalities or other mechanisms. Consider potency of effects for existing results and whether E,A,T,S mechanism is credible for reproductive/developmental effects or whether there may be non-endocrine mechanisms. Consider route of exposures for effects data and possible implications of ADME characteristics of the chemical. Hormonal activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT. Equivocal results may indicate chemical has multiple modes of action (MOA).
D	+	–	+	1) Evidence for adverse effects in OECD TG 443 but not via E,A,T,S mechanism or requires metabolic activation for activity. 2) Evidence for adverse effects in OECD TG 443 but not via E,A,T,S mechanism or may require metabolic activation for activity. Apical endpoints may be less sensitive or unaffected. 3) Evidence for adverse effects in OECD TG 443 but not via E,A,T,S mechanism or requires metabolic activation for activity. Indicators of hormonal activity may be less sensitive or unaffected.	To further discern mechanism, could perform *in vitro* estrogen receptor (ER-), androgen receptor (AR-), thyroid hormone receptor (TR), steroidogenesis (S) assays with added metabolising system.	Sufficient information to conclude evidence of concern for reproductive toxicity via possible endocrine disruption mechanism. Effects on indicators of hormonal activity alone may be indicative of changes not detected by apical endpoints. Effects on apical endpoints alone may indicate E,A,T,S modalities or other mechanisms.. Consider potency of effects for existing results and whether endocrine disruption mechanism is credible for reproductive/developmental effects or whether there may be non-endocrine mechanisms. Consider route of exposures for effects data and possible implications of ADME characteristics of the chemical. Hormonal activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT.

Scenarios	Result of EOGRTS	Existing results		Possible conclusions: 1) Indicators of hormonal activity and apical endpoints positive 2) Indicators of hormonal activity positive and apical endpoints negative 3) Indicators of hormonal activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
E	+	–	–	1) Evidence for adverse effects in OECD TG 443 via non-E,A,T,S/ non-endocrine disruption mechanism or may require metabolic activation for activity. 2) Evidence for adverse effects in OECD TG 443 via non-E,A,T,S/ non-endocrine disruption mechanism or may require metabolic activation for activity. Apical endpoints may be less sensitive or unaffected. 3 Evidence for adverse effects in OECD TG 443 via non-E,A,T,S/ non-endocrine disruption mechanism or may require metabolic activation for activity. Indicators of hormonal activity may be less sensitive or unaffected.	To further discern mechanism, could perform *in vitro* ER, AR, TR, S assays with added metabolising system.	Sufficient information to conclude evidence of concern for reproductive toxicity via unknown mechanism. Effects on indicators of hormonal activity alone may be indicative of changes not detected by apical endpoints. Effects on apical endpoints alone may indicate E,A,T,S modalities or other mechanisms. Consider potency of effects for existing results and whether endocrine disruption mechanism is credible for reproductive/developmental effects or whether there may be non-endocrine mechanisms. Consider route of exposures for effects data and possible implications of ADME characteristics of the chemical. Hormonal activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT.
F	+	–	Eq/0	1) Evidence for adverse effects in OECD TG 443 via non-E,A,T,S/ non-endocrine disruption mechanism or may require metabolic activation for activity. 2) Evidence for adverse effects in OECD TG 443 via non-E,A,T,S/ non-endocrine disruption mechanism or may require metabolic activation for activity. Apical endpoints may be less sensitive or unaffected. 3) Evidence for adverse effects in OECD TG 443 via non-E,A,T,S/ non-endocrine disruption mechanism or may requirs metabolic activation for activity. Indicators of hormonal activity may be less sensitive or unaffected.	To further discern mechanism, could perform *in vitro* ER, AR, TR, S assays with added metabolising system.	Sufficient information to conclude evidence of concern for reproductive toxicity via unknown mechanism. Effects on indicators of hormonal activity alone may be indicative of changes not detected by apical endpoints. Effects on apical endpoints alone may indicate E,A,T,S modalities or other mechanisms. Consider potency of effects for existing results and whether endocrine disruption mechanism is credible for reproductive/developmental effects or whether there may be non-endocrine mechanisms. Consider route of exposures for effects data and possible implications of ADME characteristics of the chemical. Hormonal activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT. Equivocal results may indicate chemical has multiple MOA.

Scenarios	Result of EOGRTS	Existing results		Possible conclusions: 1) Indicators of hormonal activity and apical endpoints positive 2) Indicators of hormonal activity positive and apical endpoints negative 3) Indicators of hormonal activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
G	+	Eq/0	+	1) Evidence for adverse effects in OECD TG 443, may act via E,A,T,S mechanism and may require metabolic activation for activity. 2) Evidence for adverse effects in OECD TG 443, may act via E,A,T,S mechanism and may require metabolic activation for activity. Apical endpoints may be less sensitive or unaffected. 3) Evidence for adverse effects in OECD TG 443, may act via E,A,T,S mechanism and may require metabolic activation for activity. Indicators of hormonal activity may be less sensitive or unaffected.	To further discern mechanism, could perform *in vitro* ER, AR, TR, S assays with added metabolising system.	Sufficient information to conclude evidence of concern for reproductive toxicity via possible endocrine disruption mechanism. Effects on indicators of hormonal activity alone may be indicative of changes not detected by apical endpoints. Effects on apical endpoints alone may indicate E,A,T,S modalities or other mechanisms. Consider potency of effects for existing results and whether endocrine disruption mechanism is credible for reproductive/developmental effects or whether there may be non-endocrine mechanisms. Consider route of exposures for effects data and possible implications of ADME characteristics of the chemical. Hormonal activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT. Equivocal results may indicate chemical has multiple MOA.
H	+	Eq/0	–	1) Evidence for adverse effects in OECD TG 443 via non-E,A,T,S/ non-endocrine disruption mechanism or requires metabolic activation for activity. 2) Evidence for adverse effects in OECD TG 443 via non-E,A,T,S/ non-endocrine disruption mechanism or may require metabolic activation for activity. Apical endpoints may be less sensitive or unaffected. 3 Evidence for adverse effects in OECD TG 443 via non-E,A,T,S/ non-endocrine disruption mechanism or may require metabolic activation for activity. Indicators of hormonal activity may be less sensitive or unaffected.	To further discern mechanism, could perform *in vitro* ER, AR, TR, S assays with added metabolising system.	Sufficient information to conclude evidence of concern for reproductive toxicity via unknown mechanism. Effects on indicators of hormonal activity alone may be indicative of changes not detected by apical endpoints. Effects on apical endpoints alone may indicate E,A,T,S modalities or other mechanisms. Consider potency of effects for existing results and whether endocrine disruption mechanism is credible for reproductive/developmental effects or whether there may be non-endocrine mechanisms. Consider route of exposures for effects data and possible implications of ADME characteristics of the chemical. Hormonal activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT. Equivocal results may indicate chemical has multiple MOA.

Scenarios	Result of EOGRTS	Existing results		Possible conclusions: 1) Indicators of hormonal activity and apical endpoints positive 2) Indicators of hormonal activity positive and apical endpoints negative 3) Indicators of hormonal activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
I	+	Eq/0	Eq/0	1) Evidence for adverse effects in OECD TG 443 via unknown mechanism. 2) Evidence for adverse effects in OECD TG 443 via unknown mechanism. Apical endpoints may be less sensitive or unaffected. 3) Evidence for adverse effects in OECD TG 443 via unknown mechanism. Indicators of hormonal activity may be less sensitive or unaffected.	To further discern mechanism, could perform *in vitro* ER, AR, TR, S assays with added metabolising system.	Sufficient information to conclude evidence of concern for reproductive toxicity via unknown mechanism. Effects on indicators of hormonal activity alone may be indicative of changes not detected by apical endpoints. Consider existing results and whether endocrine disruption mechanism is credible for reproductive/developmental effects or whether there may be non-endocrine mechanisms. Consider route of exposures for effects data and possible implications of ADME characteristics of the chemical. Hormonal activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT. Equivocal results may indicate chemical has multiple MOA.
J	–	+	+	No evidence of adverse effects on reproduction/development/endocrine organs. Effects seen in existing (lower level) studies do not lead to adverse outcome in Level 5 assay. Metabolism or potency may explain the difference from existing *in vitro*/and *in vivo* data.	If test is robust, no further testing needed. If not, consider supplemental testing, depending on existing data. To further discern mechanism, could perform *in vitro* ER, AR, TR, S assays with added metabolising system.	There may be sufficient information to conclude absence of concern for endocrine disruption. If existing data are from other, adequate, apical studies, question why there are differences. Consider route of exposures and possible implications for ADME characteristics of the chemical with existing studies. Further mechanistic studies with metabolism may help determine MOA.
K	–	+	–	No evidence of adverse effects on reproduction/development/endocrine organs. Metabolism or potency may explain *in vitro*/*in vivo* differences.	If test is robust, no further testing needed. If not, consider supplemental testing, depending on existing data. To further discern mechanism, could perform *in vitro* ER, AR, TR, S assays with added metabolising system.	If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption. Further mechanistic studies with metabolism may help determine MOA.

Scenarios	Result of EOGRTS	Existing results		Possible conclusions: 1) Indicators of hormonal activity and apical endpoints positive 2) Indicators of hormonal activity positive and apical endpoints negative 3) Indicators of hormonal activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
L	–	+	Eq/0	No evidence of adverse effects on reproduction/development/endocrine organs. Metabolism or potency may explain *in vitro*/*in vivo* differences. Effects seen in existing (lower level) studies do not lead to adverse outcome in Level 5 assay.	If test is robust, no further testing needed. If not, consider supplemental testing, depending on existing data. To further discern mechanism, could perform *in vitro* ER, AR, TR, S assays with added metabolising system.	There may be sufficient information to conclude absence of concern for endocrine disruption. Consider route of exposures and possible implications for ADME characteristics of the chemical with existing studies. Further mechanistic studies with metabolism may help determine MOA. Equivocal results may indicate chemical has multiple MOA.
M	–	–	+	No evidence of adverse effects on reproduction/development/endocrine organs. Effects seen in existing (lower level) studies do not lead to adverse outcome in Level 5 assay.	If test is robust, no further testing needed. If not, consider supplemental testing, depending on existing data.	There may be sufficient information to conclude absence of concern for endocrine disruption. Consider route of exposures and possible implications for ADME characteristics of the chemical with existing studies. Further mechanistic studies with metabolism may help determine MOA.
N	–	–	–	No evidence of adverse effects on reproduction/development/endocrine organs.	If test is robust, no further testing needed. If not, consider supplemental testing, depending on existing data.	If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption.
O	–	–	Eq/0	No evidence of adverse effects on reproduction/development/endocrine organs. No evidence for (anti)-E,A,T,S activity *in vitro*.	If test is robust, no further testing needed. If not, consider supplemental testing, depending on existing data. To further discern mechanism, could perform *in vitro* ER, AR, TR, S assays with added metabolising system.	There may be sufficient information to conclude absence of concern for endocrine disruption. Further mechanistic studies with metabolism may help determine MOA. Consider route of exposures and possible implications for ADME characteristics of the chemical with existing studies. Check data on chemical analogues.

Scenarios	Result of EOGRTS	Existing results		Possible conclusions: 1) Indicators of hormonal activity and apical endpoints positive 2) Indicators of hormonal activity positive and apical endpoints negative 3) Indicators of hormonal activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
P	–	Eq/0	+	No evidence of adverse effects on reproduction/development/endocrine organs. Effects seen in existing (lower level) studies do not lead to adverse outcome in Level 5 assay. Effects seen in existing studies are via unknown mechanism.	If test is robust, no further testing needed. If not, consider supplemental testing, depending on existing data. To further discern mechanism, could perform *in vitro* ER, AR, TR, S assays with added metabolising system.	There may be sufficient information to conclude absence of concern for endocrine disruption. Further mechanistic studies with metabolism may help determine MOA. Consider route of exposures and possible implications for ADME characteristics of the chemical with existing studies. Check data on chemical analogues. Equivocal results may indicate chemical has multiple MOA.
Q	–	Eq/0	–	No evidence of adverse effects on reproduction/development/endocrine organs.	If test is robust, no further testing needed. If not, consider supplemental testing, depending on existing data.	There may be sufficient information to conclude absence of concern for endocrine disruption. Check data on chemical analogues.
R	–	Eq/0	Eq/0	No evidence of adverse effects on reproduction/development/endocrine organs.	If test is robust, no further testing needed. If not, consider supplemental testing, depending on existing data.	There may be sufficient information to conclude absence of concern for endocrine disruption. Further mechanistic studies may strengthen weight of evidence. Check data on chemical analogues. Equivocal results may indicate chemical has multiple MOA.

Non-OECD mammalian screens and tests (Concptual Framework Levels 3-5)

C.3.16. Pubertal Development and Thyroid Function Assay in Peripubertal Male Rats (male PP assay) (US EPA OPPTS 890.1500)

Status: Assay validated at national level.

994. Modalities detected: (anti)androgen, thyroid, steroidogenesis.

Endpoints: Age and body weight at preputial separation (PPS). Weight of seminal vesicles (+ coagulating gland), ventral prostate, dorsolateral prostate, LABC (Levator ani plus bulbocavernosus muscle complex), epididymides, testes, thyroid, pituitary, adrenals. Histopathologic changes in epididymis, testis, thyroid. Serum testosterone, T4 and TSH.

Background to the assay

995. This assay is designed to identify chemicals that have the potential to interact with androgen receptor (AR-) mediated modalities, thyroid hormone mediated modalities and interference with steroidogenesis. It will also detect chemicals that alter pubertal development via changes in the hypothalamic/pituitary/gonadal (HPG) axis. It will also detect estrogen receptor (ER-) mediated effects, but the accuracy of this is unknown. The principle of the assay is that male rats are dosed with chemical during the period of sexual maturation, starting at postnatal day 23. Route of administration of test substance is via oral gavage. The prepubertal period is a very sensitive age for exposure to agents which alter the endocrine system (US EPA, 2007). Serum androgens in male rats change dramatically during puberty and reproductive organ weights grow rapidly during puberty (Stoker et al., 2000). PPS is an apical measure of the progression of puberty and has been used as the primary endpoint of puberty onset in the rat. It is an androgen-dependent event. The assay has its female counterpart in the peripubertal female rat assay. Male rats achieve sexual maturity at a later age than females (vaginal opening) and therefore the male assay is of longer duration than the female assay (31 days cf. 21 days) and this should be taken into account when comparing the severity of effects obtained in the two assays.

996. The male PP assay was designed to be one of the suite of assays comprising the United States Environmental Protection Agency's (US EPA) Endocrine Disruptor Screening Program (EDSP) "Tier 1" and has been validated in that context (US EPA, 2007). There is no OECD test guideline for the assay. The US EPA guideline (OPPTS 890.1500) was published in October 2009 (US EPA, 2009). Male and female PP assays are considered to be apical assays (i.e. they contain endpoints that may be changed by a number of different modes of action [MOA] and may not be specific to endocrine active substances [EASs]). The animals have intact hypothalamus-pituitary-gonadal/thyroid axes and therefore are a relevant model for human health, although the sensitivity of the assays for ER/AR agonists and antagonists are less than that of the Uterotrophic Bioassay (UT) and Hershberger Bioassay (H). A strength of the PP assays is that (unlike the H and UT assays) they will detect multiple MOA, although it may not be possible to isolate the mechanism of action. The male PP assay is likely to detect (anti)estrogens in addition to androgen/thyroid/steroidogenesis (ATS) modalities. The estrogen methoxychlor was

included in the validation studies of the male assay and gave a weak positive response for some endpoints. Published studies have also demonstrated that the assay responds to strong estrogens such as diethylstilbestrol (Ashby and Lefevre, 2000) and weak estrogens such as nonylphenol (Tan, Kassim and Mohd, 2003). The validation of the male PP assay indicated that sensitivity was high and although it has not been extensively investigated, it showed that the male pubertal assay can be sensitive to dose levels that are near the lowest observed effect level (LOEL) in a developmental toxicity study on the androgen antagonist vinclozolin (US EPA, 2007).

997. A limitation of the validation is that no chemical was shown to be completely negative in the assay. Chloronitrobenzene was included in the validation as a chemical that was expected to be toxic but without endocrine activity, but when tested was positive, delaying PPS, decreasing serum testosterone, decreasing growth of androgen-dependent tissues and reducing T4 levels. It is not known whether these effects were due to non-specificity of the assay or a real effect on endocrine systems. Other chemicals, however, that were positive for one endocrine system were not necessarily positive on others (e.g. perchlorate altered thyroid hormones and thyroid weight but caused no effects on any of the reproductive tract weights or puberty onset). This indicates that false positives are not always seen and helps to reinforce the specificity of the assay. Another possible limitation is the inability to detect specific aromatase inhibitors. Although more general inhibitors of steroidogenesis (including aromatase inhibition), such as ketoconazole, are detected in the assay, specific inhibitors of aromatase only, such as fadrozole, were not (Marty, Crissman and Carney, 2001).

998. Experience with of serum hormone determinations in Level 4 and Level 5 rodent assays has revealed that their detection/measurement in rodent studies can be challenging. A recent workshop on "Practicability of Hormonal Measurements" was organised by the BfR (Germany) and the finding from this workshop will be published (Kucheryavenko et al., 2018). The OECD Expert Group on Reproductive and Developmental Toxicity recommends that to demonstrate proficiency for thyroid hormones measurement, a laboratory should be able to show results from a separate study using a positive control substance. Laboratories may also submit their calibration curves, standard curves, as well as data on the levels of quantification and detection. This group is also establishing a historical control database with thyroid toxicant positive controls.

When/why the assay may be used

999. As mentioned above, the male PP assay may be used as part of the US EPA's EDSP Tier 1 screening battery as an apical assay to detect interaction with multiple endocrine systems. In this context, its use is primarily for hazard determination. It may also be used as a follow-up assay following positive results in *in vitro* assays (e.g. a positive result in the Steroidogenesis Assay). Positive results in an AR *in vitro* assay would preferably be followed by an H assay for reasons of animal welfare – H assays require fewer animals than the male PP assays and are of shorter duration. If there is a need to test in an apical assay, the PP assay may be chosen, realising the caveat that there is some uncertainty regarding its specificity. Depending on the number of doses used, the PP assay may be used for hazard identification/characterisation. The assay could potentially also be used to investigate or supplement higher tier data, possibly to clarify the MOA. One scenario could be if only limited reproductive data are available (e.g. a study not conducted to modern standards or not containing endpoints for sexual development). Data from female and male PP assays could then be used to investigate the occurrence of endocrine effects. A decision

about whether to conduct further animal tests would, however, need to consider whether sufficient supplementary data may be provided by *in vitro* tests.

1000. In order to provide information relevant for assessing whether or not a chemical may fulfil the WHO/IPCS (2002) definition of an endocrine disruptor (ED), the study design has to be sufficiently robust to demonstrate the presence or absence of effects. In the dose selection, the investigator should also consider and ensure that data generated are adequate to fulfil the regulatory requirement across OECD countries as appropriate (e.g. hazard and risk assessment and labelling, ED assessment, etc.). The top dose or concentration should be sufficiently high to give clear systemic (i.e. non endocrine-specific) toxicity in order to ensure that a wide range of exposures (high to low) is tested. However, endocrine effects observed solely in the presence of clear systemic toxicity should be interpreted with caution and may be disregarded when sufficiently justified to be caused by secondary effects which are unlikely to be due to endocrine activity. The reason for this advice is a concern that some EAS-sensitive assays are being run at doses/concentrations of EASs that are too low to trigger direct impacts on the endocrine system. This guidance document is not the place to address this issue directly, but it should be considered when EAS-sensitive test guidelines (TGs) are revised in the future. In addition, the number and spacing of dose/concentration levels should also be adequate to fulfil the objectives of the study (e.g. to demonstrate dose response relationships if this is required).

Introduction to the table of scenarios

1001. Table C.3.16 gives guidance on a further step to take in the event of a positive (+) or negative (-) result and in the presence of positive (+), negative (-) or equivocal/absent (Eq/0) existing results. “Existing results” are subdivided into “mechanism” and “effects” data (third and fourth columns). The table is divided horizontally into a series of scenarios that represent all the combinations of these events.

1002. The results of the male PP assay are given in the second column. The assay contains multiple endpoints and it is not possible to provide alternative scenarios for all combinations, therefore some discrimination has been attempted by dividing the endpoints into “apical” and “indicators of hormonal activity”. The terminology used has been chosen to be consistent between both the non-mammalian and mammalian tests. Both groups have similar biological importance, although the “indicators of hormonal activity” in the mammalian assays are serum hormones and are generally, but not always, more variable than “apical endpoints”. “Apical endpoints” are age/body weight at PPS, weights of seminal vesicles, prostate, LABC, epididymides, testes, thyroid, pituitary and adrenals; histopathologic changes in epididymis, testis, thyroid. “Indicators of hormonal activity” are hormones (testosterone, T4 and TSH).

1003. Three possible outcomes for a positive result are therefore envisaged in Table C.3.16:

1. indicators of hormonal activity and apical endpoints positive
2. indicators of hormonal activity positive and apical endpoints negative
3. indicators of hormonal activity negative and apical endpoints positive.

1004. A positive result for apical endpoints could be delayed puberty (preputial separation) or biologically significant reductions in weights of the epididymides, prostate and seminal vesicles accompanied by treatment-related histopathologic changes. A positive result for indicators of hormonal activity could be biologically significant changes in thyroid hormone profiles. The multiple endpoints in this assay mean that there is some redundancy in the

assay, but this is useful as not all chemicals may affect all endpoints associated with a mechanism of action and there may be site-specific differences in response.

1005. Single isolated changes may be indicative of spurious results, but robust dose response information may not be available as the TG only requires two dose levels. The guidance on histopathologic changes in endocrine tests (OECD, 2009) may be helpful in interpretation. Such results should be considered with caution, although it is possible that weak effects have been detected which may then be seen in longer term studies.

1006. A negative result for the male PP assay is taken to be the absence of changes in indicators of hormonal activity and apical endpoints. In the absence of other pertinent lines of evidence, negative results in this test alone cannot be taken as evidence that the substance is not an ED. Further studies will be required as confirmation.

1007. Equivocal results for the guideline are not considered in the table, partly for brevity but also because equivocal results are by nature uncertain. A decision must eventually be reached about whether the endocrine endpoints tend to be positive or negative or whether the result must be put to one side and the test repeated (using the same or a different test). Factors which may have interfered with the result (e.g. composition of the diet used, environmental influences) should be considered. Performance criteria (coefficients of variation for the test endpoints) should be checked for compliance with those in the TG. The assay does not include concurrent positive controls, but attempts have been made to mitigate this by including the performance criteria.

Existing data to be considered

1008. Existing "mechanism" *in vitro* data are assumed to be available from ER-, AR- and steroidogenesis-based assays (Level 2). Assays may also be available for interference with thyroid modalities. In practice, it is possible that data from all of these assays may not be available, so judgement will need to be used to decide which assays to perform. Although the current *in vitro* test guidelines do not incorporate metabolic activation, published information on use of metabolic activation systems is available in Jacobs et al. (2008; 2013) and OECD (2008). These methods, however, have not yet been validated.

1009. Existing "effects" data refer to *in vivo* effects that may come from H assays where a non-physiological animal model is used. In these cases, it should be remembered that these assays are specifically designed to be sensitive to EDs. Another possibility is that repeat dose oral toxicity studies, reproduction/developmental toxicity screen tests or read-across from analogues may be available. It is unlikely that the male PP assay will be performed if data from robust higher tier reproductive studies are already available as the PP assay offers no advantage over these assays. It is possible, though, that the PP assay has been performed to supplement non-robust higher tier data for the reasons given above. Data may also be available on effects in non-mammalian species, although caution should be used when extrapolating between taxa. A chemical causing endocrine effects in non-mammalian environmental species (fish, for example) may also have endocrine effects in mammals, but the physiological consequences of the effects are likely to be different.

1010. When considering the results of the male PP assay, all available data should be used in order to reach a conclusion and a weight of evidence approach taken. This may include high throughput screening data, read-across data from structural analogues and quantitative structure activity relationship (QSAR). Several QSAR models for ER and AR binding/activation are now available (see Sections B.1.1.1 and B.1.1.2).

Scenarios: Positive and negative results combined with existing data

1011. The scenarios (A to R) presented in Table C.3.16 represent all the possibilities of positive or negative results in combination with the presence or absence of existing data. The action taken will also depend on the regulatory environment, but the considerations given here are generally science based. Although the male PP assay uses rats, the well-conserved nature of the hormonal pathways across taxa should be a strong indication that results in this assay may be relevant to other vertebrate species. Effects in laboratory mammal tests are also highly relevant for environmental mammalian species. Wherever possible, the recommended "next step which could be taken" avoids unnecessary animal testing. However, sometimes conducting an animal test will be indicated and then the relevance of species, strain, exposure route and species-specific metabolism should always be considered. The sensitivity and physiological function of the hormone under investigation in the test species should also be considered. In general, lower level tests should be conducted before higher level tests in order to avoid unnecessary animal usage, unless it is apparent that a Level 5 test will be required anyway or will be needed to establish the evidence to conclude on ED properties. Information on some endocrine-related tumours may be detected more comprehensively in carcinogenicity studies (OECD TG 451/453) (Level 4); for example, detection of certain types of thyroid tumors in the absence of reproductive or developmental effects, as well as substances causing tumors in other endocrine-sensitive tissues. At Level 5, the EOGRTS (OECD TG 443) is the most sensitive reproduction assay for detecting endocrine disruption because it includes evaluation of a number of endocrine endpoints not included in the two-generation study (OECD TG 416) adopted in 2001. It is recognised however, that some jurisdictions may require a two-generation study. Further considerations specific to each scenario are given in the table.

1012. Scenarios A to C represent positive results in the male PP assay in the presence of positive *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. Each positive male PP result scenario is divided into the three possible outcomes given above. A positive result in the *in vitro* assays in combination with a positive male PP assay is moderate or strong evidence for E,A,T,S-mediated activity that may or may not be supported by the *in vivo* effects data. In the absence of robust upper-level data, the next step may be to conduct an upper-level test. In the presence of robust Level 5 data, there may be sufficient evidence to conclude concern for endocrine disruption and therefore no need for further screening. Positive results in the male PP assay may also indicate the potential for endocrine mediated effects in lower vertebrates. These could be followed up with partial life cycle tests such as the Fish Sexual Development Test (FSDT) or the Medaka Extended One-Generation Reproduction Test (MEOGRT); or the Larval Amphibian Growth and Development Assay (LAGDA) if effects are on the thyroid hormone system. *In vivo* assays/tests with negative results should be interpreted with caution as they may either indicate that the tests used do not have sufficient power to detect weak effects or, alternatively, that the effects do not present a concern for endocrine disruption. The possibility of other (non-E,A,T,S) mechanisms should also not be overlooked (e.g. involving other receptors or endocrine axes).

1013. Scenarios D to F represent positive results in the male PP assay in the presence of negative *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. Each positive male PP result scenario is divided into the three possible outcomes given above. Negative results in the *in vitro* assays should be viewed with caution in case a metabolite is responsible for the positive male PP assay. Unless the metabolic profile of the test substance is known, one option may be to conduct these assays with an added

metabolising system. If the metabolic profile is known, then a higher level *in vivo* test may be advisable. The choice of tests will depend on the available *in vivo* effects data. Positive results in the male PP assay may also indicate the potential for endocrine mediated effects in lower vertebrates. These could be followed up with partial life cycle tests such as the FSDT or MEOGRT; or the LAGDA if effects are on the thyroid hormone system. As in Scenarios A to C, *in vivo* assays/tests with negative results should be interpreted with caution as they may either indicate that the tests used do not have sufficient power to detect weak effects or, alternatively, that the effects do not present a concern for endocrine disruption.

1014. Scenarios G to I represent positive results in the male PP assay in the presence of various combinations of missing or equivocal data. Positive results in the male PP assay may also indicate the potential for endocrine mediated effects in lower vertebrates. These could be followed up with partial life cycle tests such as the FSDT or MEOGRT; or the LAGDA if effects are on the thyroid hormone system. Each positive male PP result scenario is divided into the three possible outcomes given above. The next step to take in these eventualities will depend on the nature of the other available data and the jurisdiction in which it is being used. In some cases, equivocal data may be viewed as positive whilst in others it may or may not contribute to the weight of evidence. The interpretation may also depend on the mode of action (MOA) in question and why the data are considered equivocal, e.g. a study that is equivocal for thyroid effects may still be of value in evaluating (anti)androgenic effects. In all three scenarios, the recommended first step is to obtain reliable mechanistic (*in vitro*) data rather than proceed further with *in vivo* testing. Equivocal and missing data are alternative scenarios and two possibilities for the next step are given in most cases, but the nature of equivocal data means that decisions need to be taken on a case-by-case basis. In all cases, the role of metabolism, route of exposure and data from structural analogues should be considered before deciding on the next step.

1015. Scenarios J to L represent negative results in the male PP assay in the presence of positive *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. As a negative result for the male PP is taken to be negative findings for both indicators of hormonal activity and apical endpoints (unlike the situation with positive outcomes), there is only one possible negative outcome. All three scenarios could also arise from a chemical that is positive in *in vitro* assays, but is metabolised to a non-active metabolite leading to negative results in the male PP assay. This should be considered first when investigating the next step. Endocrine active potency may also explain differences between *in vitro* and *in vivo* results (e.g. a chemical with weak endocrine activity may give a positive result *in vitro* but may be negative *in vivo*). Positive *in vivo* effects data may involve other E,A,T,S, non-E,A,T,S mechanisms (e.g. involving other receptors or endocrine axes), more sensitive endpoints, greater statistical power or life stages that are more sensitive to the substance than the young adult animals in the male PP assay.

1016. Scenarios M to O represent negative results in the male PP assay in the presence of negative *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. Negative results for all tests (Scenario N) may be sufficient to enable a conclusion of no concern for endocrine disruption. This will depend on the weight of evidence and may not be possible. Where there are positive *in vivo* effects data, there could still be an E,A,T,S-related mechanism, the effects may be related to length of exposure, route of exposure or exposure at different life stages. Other E,A,T,S or non-E,A,T,S mechanisms may also be involved.

1017. Scenarios P to R represent negative results in the male PP assay in the presence of various combinations of missing or equivocal data. As with the positive result scenarios above (see Paragraph 1 014), the next step to take in these eventualities will have to be decided on a case-by-case basis. However, the recommended first step is generally to obtain reliable mechanistic (*in vitro*) data rather than proceed further with *in vivo* testing. In all cases, the role of metabolism, route of exposure and data from structural analogues should be considered before deciding on the next step.

1018. In all scenarios (A to R), the next step to take to strengthen weight of evidence will depend on the existing information. Table C.3.16 is meant to provide a succinct guide and may not cover all circumstances or possibilities. The scenarios may also suggest that chemicals have simple or single MOA, when in practice they may have multiple endocrine and non-endocrine MOA. In some cases, for example, two opposite modes of simultaneous action (e.g. estrogenic and anti-estrogenic) could, depending on dose, lead to a minimisation or abolition of effects, while in others two different MOA (e.g. estrogenic and anti-androgenic) could potentially reinforce effects. Endocrine pathways interact, mixed effects are common and there are many pathways that cannot be distinguished with currently available TGs. If multiple MOA are suspected, either from the existing results or based on QSAR/read-across/integrated approaches, this should be investigated further if needed for regulatory decision making.

References

Ashby, J. and P.A. Lefevre (2000), "The peripubertal male rat assay as an alternative to the Hershberger castrated male rat assay for the detection of antiandrogens, oestrogens and metabolic modulators", *Journal of Applied Toxicology*, Vol. 20/1, pp. 35-47, https://doi.org/10.1002/(SICI)1099-1263(200001/02)20:1<35::AID-JAT633>3.0.CO;2-8.

Jacobs, M. et al. (2013), "*In vitro* metabolism and bioavailability tests for endocrine active substances: What is needed next for regulatory purposes?", *ALTEX – Alternatives to Animal Experimentation*, Vol. 30/3, pp. 331-351.

Jacobs, M.N. et al. (2008), "The use of metabolising systems for *in vitro* testing of endocrine disrupters", *Current Drug Metabolism*, Vol. 9/8, pp. 796-826.

Kucheryavenko, O. et al. (2018), "Report from the BfR Expert Hearing on Practicability of Hormonal Measurements", *Archives of Toxicology* (in prep.).

Marty, M.S., J.W. Crissman and E.W. Carney (2001), "Evaluation of the male pubertal onset assay to detect testosterone and steroid biosynthesis inhibitors in CD rats", *Toxicological Sciences*, Vol. 60/1, pp. 285-295, http://dx.doi.org/10.1093/toxsci/60.1.63.

OECD (2009), "Guidance document for histologic evaluation of endocrine and reproductive tests in rodents", OECD Series on Testing and Assessment, No. 106, OECD, Paris, www.oecd.org/env/ehs/testing/43411534.pdf.

OECD (2008), *Detailed Review Paper on the Use of Metabolising Systems for* In Vitro *Testing of Endocrine Disrupters*, OECD Series on Testing and Assessment, No. 97, OECD Publishing, Paris, https://doi.org/10.1787/9789264085497-en.

Stoker, T.E. et al. (2000), "Endocrine disrupting chemicals: Prepubertal exposures and effects on sexual maturation and thyroid function in the male rat. A focus on the EDSTAC recommendations", *Critical Reviews in Toxicology*, Vol. 30/2, pp. 197-252, https://doi.org/10.1080/10408440091159194.

Tan, B.L., N.M. Kassim and M.A. Mohd (2003), "Assessment of pubertal development in juvenile male rats after sub-acute exposure to bisphenol A and nonylphenol", *Toxicology Letters*, Vol. 143/3, pp. 261-270, https://doi.org/10.1016/S0378-4274(03)00172-3.

US EPA (2009), "Endocrine Disrupter Screening Program Test Guidelines OPPTS 890.1500: Pubertal development and thyroid function in intact juvenile/peripubertal male rats", Environmental Protection Agency, Washington, DC, https://www.regulations.gov/document?D=EPA-HQ-OPPT-2009-0576-0010.

US EPA (2007), "Integrated summary report for validation of a test method for assessment of pubertal development and thyroid function in juvenile male rats as a potential screen in the Endocrine Disrupter Screening Program Tier-1 battery", Environmental Protection Agency, Washington, DC.

WHO/IPCS (2002), "Global assessment of the state-of-the-science of endocrine disrupters", Damstra, T. et al. (eds.) WHO/PCS/EDC/02.2, World Health Organization, Geneva, www.who.int/ipcs/publications/new_issues/endocrine_disruptors/en.

Table C.3.16. **Pubertal Development and Thyroid Function Assay in Peripubertal Male Rats (male PP assay) (OPPTS 890.1500): Guidance for scenarios of combinations of results with existing data**

This table represents possible conclusions to be drawn from assay data, and a next step which could be taken if further evidence is required about possible endocrine disrupting properties and/or effects. The guidance offered is not meant to be prescriptive, but provides science-based considerations. It encourages the use of all available data and expert judgement in a weight of evidence approach. Regional and national interpretation of results and "next steps" may vary.

The conclusions are grouped into a series of scenarios (A-R), each scenario representing a different combination of assay results, existing *in vitro* data and existing *in vivo* data. The symbol "+" indicates that the data in question represent a positive result, «-" indicates a negative result, and "Eq/0" indicates that the data are either equivocal or are not available.

Existing results: * "Mechanism (*in vitro* mechanistic data)" assumes that mechanistic data are available from estrogen receptor (ER-), androgen receptor (AR-) and steroidogenesis-based assays (Level 2). Thyroid hormone receptor (TR) and other assays concerning mechanisms of thyroid disruption may be available, but they are not in common use. In practice, data from all assays may not be available and therefore this must be taken into account when deciding on the "next step". Quantitative structure activity relationship (QSAR) predictions of estrogen and androgen binding/activation may be made for some substances.

Existing results: ** "Effects (*in vivo* effects of concern)" assumes effects have been observed in other *in vivo* screens/tests which give rise to concern that the test chemical may be an endocrine disrupter. These may be repeated dose toxicity tests (e.g. OECD TG 407, TG 408 28- and 90-day studies) or read-across from chemical analogues.

*** *Note:* three possible outcomes for a positive result are given:

1. indicators of hormonal activity and apical endpoints positive
2. indicators of hormonal activity positive and apical endpoints negative
3. indicators of hormonal activity negative and apical endpoints positive.

"Apical endpoints" are age/body weight at PPS; weights of seminal vesicles, prostate, LABC, epididymides, testes, thyroid, pituitary and adrenals; histopathologic changes in epididymis, testis, thyroid.

"Indicators of hormonal activity" are hormones (testosterone, T4 and TSH).

Scenarios	Result of male PP assay	Existing results		Possible conclusions: 1) Indicators of hormonal activity and apical endpoints positive 2) Indicators of hormonal activity positive and apical endpoints negative 3) Indicators of hormonal activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
A	+ ***	+	+	1) Increased evidence of (anti)-A,T,S activity (weak, moderate or strong). 2) Possible evidence of (anti)-A,T,S activity, apical endpoints may be less sensitive or unaffected. Increased evidence of (anti)-A,T,S activity. 3) Moderate or strong (anti)-A,T,S activity, indicators of hormonal activity may be less sensitive or unaffected. Increased evidence of (anti)-A,T,S activity.	Perform assay from Level 5 (e.g. Extended One-Generation Reproduction Toxiticy Study [EOGRTS] or two-generation assay).	If existing data are from a Level 5 assay, there may be sufficient information to conclude evidence of concern for endocrine disruption (the EOGRTS provides the most information; however, for endocrine disrupting chemicals [EDCs] with a carcinogenic potential, OECD TG 451-3 may be more sensitive). Effects on indicators of hormonal activity alone may be indicative of changes not detected by apical endpoints. Effects on apical endpoints alone may indicate E,A,T,S modalities or other mechanisms. Possible effects on estrogen modality should also be considered. Consider route of exposures for effects data and possible implications of absorption, distribution, metabolism and excretion (ADME) characteristics of the chemical. Hormonal activity possible in lower vertebrates. Consider performing a Fish Sexual Development Test (FSDT), Larval Amphibian Growth and Development Assay (LAGDA) or Medaka Extended One-Generation Reproduction Test (MEOGRT).
B	+	+	–	1) Increased evidence of (anti)-A,T,S activity (weak, moderate or strong). 2) Possible evidence of (anti)-A,T,S activity, apical endpoints may be less sensitive or unaffected. Increased evidence of (anti)-A,T,S activity. 3) Moderate or strong (anti)-A,T,S activity, indicators of hormonal activity may be less sensitive or unaffected. Increased evidence of (anti)-A,T,S activity.	Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	Question why there are differences from existing data. Consider route of exposures and possible implications of ADME characteristics of the chemical. If existing data are from a less sensitive assay, a higher level test may be required. Effects on indicators of hormonal activity alone may be indicative of changes not detected by apical endpoints. Effects on apical endpoints alone may indicate A,T,S modalities or other mechanisms. Possible effects on estrogen modality should also be considered. Hormonal activity possible in lower vertebrates. Consider performing a FSDT, LAGDA or MEOGRT.

Scenarios	Result of male PP assay	Existing results		Possible conclusions: 1) Indicators of hormonal activity and apical endpoints positive 2) Indicators of hormonal activity positive and apical endpoints negative 3) Indicators of hormonal activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
C	+	+	Eq/0	1) Increased evidence of (anti)-A,T,S activity (weak, moderate or strong). 2) Possible evidence of (anti)-A,T,S activity, apical endpoints may be less sensitive or unaffected. Increased evidence of (anti)-A,T,S activity. 3) Moderate or strong (anti)-A,T,S activity, indicators of hormonal activity may be less sensitive or unaffected. Increased evidence of (anti)-A,T,S activity.	Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	Check data on chemical analogues. Effects on indicators of hormonal activity alone may be indicative of changes not detected by apical endpoints. Effects on apical endpoints alone may indicate A,T,S modalities or other mechanisms. Possible effects on estrogen modality should also be considered. Consider route of exposure for female Peripubertal (PP) Assay and follow-up assay. Possible implications of ADME characteristics of the chemical. Hormonal activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT. Equivocal results may indicate chemical has multiple modes of action (MOA).
D	+	–	+	1) Increased evidence of (anti)-A,T,S activity (weak, moderate or strong). Acts via non-AR, TR, S mechanism or may require metabolic activation for activity. 2) Possible evidence of (anti)-A,T,S activity, apical endpoints may be less sensitive or unaffected. Acts via non-AR, TR, S mechanism or may require metabolic activation for activity. 3) Moderate or strong (anti)-A,T,S activity, indicators of hormonal activity may be less sensitive or unaffected. Acts via non-AR, TR, S mechanism or may require metabolic activation for activity.	Perform *in vitro* estrogen receptor (ER-), androgen receptor (AR-), thyroid hormone receptor (TR-), steroidogenesis (S) assays with added metabolising system.	If existing data are from an adequate Level 5 assay, there is sufficient information to conclude evidence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). Effects on indicators of hormonal activity alone may be indicative of changes not detected by apical endpoints. Effects on apical endpoints alone may indicate A,T,S modalities or other mechanisms. Possible effects on estrogen modality should also be considered. Consider route of exposures for effects data and possible implications of ADME characteristics of the chemical. Further mechanistic studies may help determine MOA. Hormonal activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT.

Scenarios	Result of male PP assay	Existing results		Possible conclusions: 1) Indicators of hormonal activity and apical endpoints positive 2) Indicators of hormonal activity positive and apical endpoints negative 3) Indicators of hormonal activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
E	+	–	–	1) Increased evidence of (anti)-A,T,S activity (weak, moderate or strong). Acts via non-AR, TR, S mechanism or may require metabolic activation for activity. Route of exposure may account for the differences from existing data. 2) Possible evidence of (anti)-A,T,S activity, apical endpoints may be less sensitive or unaffected. Acts via non-AR, TR, S mechanism or may require metabolic activation for activity. Route of exposure may account for the differences from existing data. 3) Possible evidence of (anti)-A,T,S activity, indicators of hormonal activity may be less sensitive or unaffected. Acts via non-AR, TR, S mechanism or may require metabolic activation for activity. Route of exposure may account for the differences from existing data.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system OR Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	Question why there are differences from existing data. Consider route of exposures and possible implications of ADME characteristics of the chemical. If existing data are from a less sensitive assay, a higher level test may be required. Effects on indicators of hormonal activity alone may be indicative of changes not detected by apical endpoints. Effects on apical endpoints alone may indicate other mechanisms. Hormonal activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT.
F	+	–	Eq/0	1) Increased evidence of (anti)-A,T,S activity (weak, moderate or strong). Acts via non-AR, TR, S mechanism or may require metabolic activation for activity. 2) Possible evidence of (anti)-A,T,S activity, apical endpoints may be less sensitive or unaffected. Acts via non-AR,TR, S mechanism or may require metabolic activation for activity. 3) Moderate (anti)-E,A,T,S activity, indicators of hormonal activity may be less sensitive or unaffected. Acts via non-AR, TR, S mechanism or may require metabolic activation for activity.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system OR Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	Effects on indicators of hormonal activity alone may be indicative of changes not detected by apical endpoints. Effects on apical endpoints alone may indicate other mechanisms. Check data on chemical analogues. Further mechanistic studies may help determine MOA. If existing data are from an adequate Level 5 assay, question why there are differences. If existing data are from a less sensitive assay, a higher level test may be required. Hormonal activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT. Equivocal results may indicate chemical has multiple MOA.

Scenarios	Result of male PP assay	Existing results		Possible conclusions: 1) Indicators of hormonal activity and apical endpoints positive 2) Indicators of hormonal activity positive and apical endpoints negative 3) Indicators of hormonal activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
G	+	Eq/0	+	1) Increased evidence of (anti)-A,T,S activity (weak, moderate or strong). May act via AR, TR, S mechanism (metabolic activation may be needed). 2) Possible evidence of (anti)-A,T,S activity, apical endpoints may be less sensitive or unaffected. May act via AR, TR, S mechanism (metabolic activation may be needed). 3) Moderate or strong (anti)-A,T,S activity, indicators of hormonal activity may be less sensitive or unaffected. Increased evidence of (anti)-A,T,S activity. May act via AR, TR, S mechanism (metabolic activation may be needed).	Perform *in vitro* ER, AR, TR, S assays with added metabolising system (for the "0" scenario, otherwise Eq result available).	If existing data are from a Level 5 asssay, there may be sufficient information to conclude evidence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). Effects on indicators of hormonal activity alone may be indicative of changes not detected by apical endpoints. Effects on apical endpoints alone may indicate A,T,S modalities or other mechanisms. Possible effects on estrogen modality should also be considered. Check data on chemical analogues. Further mechanistic studies may help determine MOA. Consider route of exposures for effects data and possible implications of ADME characteristics of the chemical. Equivocal results may indicate chemical has multiple MOA.
H	+	Eq/0	–	1) Increased evidence of (anti)-A,T,S activity (weak, moderate or strong). Acts via unknown mechanism or may require metabolic activation for activity. Route of exposure may account for the differences from existing data. 2) Possible evidence of (anti)-A,T,S activity, apical endpoints may be less sensitive or unaffected. Acts via unknown mechanism. Route of exposure may account for the differences from existing data. 3) Moderate (anti)- A,T,S activity, indicators of hormonal activity may be less sensitive or unaffected. Acts via unknown mechanism. Route of exposure may account for the differences from existing data.	For the "0" scenario, perform *in vitro* ER, AR, TR, S assays with added metabolising system. (otherwise Eq result available).	Question why there are differences from existing data. Consider route of exposures and possible implications of ADME characteristics of the chemical. If existing data are from a less sensitive assay, a higher level test may be required. Effects on indicators of hormonal activity alone may be indicative of subtle changes not detected by apical endpoints. Effects on apical endpoints alone may indicate A,T,S modalities or other mechanisms. Possible effects on estrogen modality should also be considered. Check data on chemical analogues. Further mechanistic studies may help determine MOA. Hormonal activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT. Equivocal results may indicate chemical has multiple MOA.

Scenarios	Result of male PP assay	Existing results		Possible conclusions: 1) Indicators of hormonal activity and apical endpoints positive 2) Indicators of hormonal activity positive and apical endpoints negative 3) Indicators of hormonal activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
I	+	Eq/0	Eq/0	1) Increased evidence of (anti)-A,T,S activity (weak, moderate or strong). Acts via unknown mechanism. Unknown potential for adverse effects. 2) Possible evidence of (anti)-E,A,T,S activity, apical endpoints may be less sensitive or unaffected. Acts via unknown mechanism. Unknown potential for adverse effects. 3) Moderate or strong (anti)-A,T,S activity, indicators of hormonal activity may be less sensitive or unaffected. Acts via unknown mechanism. Unknown potential for adverse effects. There may be a need for metabolic activation.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system OR Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	Effects on indicators of hormonal activity alone may be indicative of changes not detected by apical endpoints. Effects on apical endpoints alone may indicate A,T,S modalities or other mechanisms. Possible effects on estrogen modality should also be considered. Check data on chemical analogues. Further mechanistic studies may help determine MOA. Hormonal activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT. Equivocal results may indicate chemical has multiple MOA.
J	–	+	+	No evidence for A,T,S activity in male PP assay. Metabolism or potency explains the difference from existing *in vitro* and *in vivo* data. Effects seen in existing studies are via non-A,T,S mechanism.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system OR Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	If existing data are from an adequate Level 5 assay, question why there are differences. If data are from a Hershberger Bioassay (H), this may be more sensitive than male Peripubertal (PP) Assay. Effects seen in existing studies may be in a more sensitive life stage. Consider route of exposures and possible implications of ADME characteristics of the chemical. Further mechanistic studies may help determine MOA.
K	–	+	–	No evidence for A,T,S activity in male PP assay. Metabolism or potency explains *in vitro*/*in vivo* differences.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system OR Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from a less sensitive assay, a higher level test may be required. If data are from H assay, need to conduct higher tier assay to conclude absence of concern for endocrine disruption. Further mechanistic studies with metabolism may help determine MOA.

Scenarios	Result of male PP assay	Existing results		Possible conclusions: 1) Indicators of hormonal activity and apical endpoints positive 2) Indicators of hormonal activity positive and apical endpoints negative 3) Indicators of hormonal activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
L	–	+	Eq/0	No evidence for A,T,S activity in male PP assay. Metabolism or potency explains *in vitro*/*in vivo* differences. Unknown potential for adverse effects.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system OR Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	Metabolic deactivation of chemical may occur *in vivo* so that possible *in vitro* activity is not realised. Consider possible routes of exposure implications of metabolism. Equivocal results may indicate chemical has multiple MOA.
M	–	–	+	No evidence for A,T,S activity in male PP assay. Effects seen in existing studies are via non-A,T,S mechanism.	Perform assay from Level 5 (e.g. EOGRTS or two-generation assay).	If existing data are from an adequate Level 5 assay, question why there are differences (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If data are from H assay, this may be more sensitive than male PP assay. Effects seen in existing studies may be in a more sensitive life stage. Consider route of exposures and possible implications of ADME characteristics of the chemical.
N	–	–	–	No evidence for A,T,S activity in male PP assay. No evidence for (anti)-A,T,S activity *in vitro*. No evidence of adverse effects.	Possibly no need for further testing. If there is uncertainty, may perform assay from Level 5 (e.g. EOGRTS or two-generation) assay.	If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive).
O	–	–	Eq/0	No evidence for A,T,S activity in male PP assay. No evidence for (anti)-A,T,S activity *in vitro*. Unknown potential for adverse effects.	Perform assay from Level 5 (e.g. EOGRTS or two-generation) assay.	Consider route of exposures and possible implications for ADME characteristics of the chemical in follow-up assay.
P	–	Eq/0	+	No evidence for A,T,S activity in female PP assay. Potential for adverse effects via unknown mechanism.	Perform *in vitro* ER, AR,TR, S assays.	Consider route of exposure and possible implications for differences from existing assay. If data are from H assay, this may be more sensitive than male PP assay. Effects seen in existing studies may be in a more sensitive life stage. Equivocal results may indicate chemical has multiple MOA.

Scenarios	Result of male PP assay	Existing results		Possible conclusions: 1) Indicators of hormonal activity and apical endpoints positive 2) Indicators of hormonal activity positive and apical endpoints negative 3) Indicators of hormonal activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
Q	–	Eq/0	–	No evidence for A,T,S activity in male PP assay. No evidence of adverse effects.	Perform *in vitro* ER, AR,TR, S assays.	If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). Further mechanistic studies may strengthen weight of evidence.
R	–	Eq/0	Eq/0	No evidence for A,T,S activity in male PP assay.	Perform *in vitro* ER, AR,TR, S assays, otherwise Eq result available.	Further mechanistic studies may strengthen weight of evidence. Check data on chemical analogues. Equivocal results may indicate chemical has multiple MOA.

C.3.17. Pubertal Development and Thyroid Function Assay in Peripubertal Female Rats (female PP assay) (US EPA OPPTS 890.1450)

Status: Assay validated at national level.

1019. Modalities detected: (anti)estrogen, thyroid, steroidogenesis.

Endpoints: Age and body weight at vaginal opening (VO). Weight of ovaries, uterus, thyroid, pituitary, adrenals. Histopathologic changes in ovaries, uterus, thyroid. Serum T4 and TSH. Age at first vaginal estrus after VO, estrus cyclicity parameters.

Background to the assay

1020. This assay is designed to identify chemicals that have the potential to interact with estrogen receptor (ER-) mediated modalities, thyroid hormone mediated modalities and interference with steroidogenesis. It will also detect chemicals that alter pubertal development via changes in the hypothalamic/pituitary/gonadal (HPG) axis. The principle of the assay is that female rats are dosed with chemical during period of sexual maturation, starting at postnatal day 22. Route of administration of test substance is via oral gavage. The prepubertal period is a very sensitive age for exposure to agents which alter the endocrine system (Goldman et al., 2000). Sexual maturation is determined in females as VO (or patency) and is an estrogen-dependent event that follows the first period of ovarian follicular growth (Goldman et al., 2000). The assay has its male counterpart in the peripubertal (PP) male rat assay. Female rats achieve sexual maturity at an earlier age than males (preputial separation) and therefore the female assay is of shorter duration than the male assay (21 days cf. 31 days) and this should be taken into account when comparing the severity of effects obtained in the two assays.

1021. The female PP assay was designed to be one of the suite of assays comprising the United States Environmental Protection Agency's Endocrine Disruptor Screening Program "Tier 1" and has been validated in that context (US EPA, 2007). There is no OECD test guideline for the assay. The US EPA guideline (OPPTS 890.1450) was published in October 2009 (US EPA, 2009). Male and female PP assays are considered to be apical assays (i.e. they contain endpoints that may be changed by a number of different modes of action [MOA] and may not be specific to endocrine active substances [EASs]). The animals have intact hypothalamus-pituitary-gonadal/thyroid axes and therefore are a relevant model for human health, although the sensitivity of the assays for estrogen receptor/androgen receptor (ER/AR) agonists and antagonists are less than that of the Uterotrophic Bioassay (UT) and Hersgberger Bioassay (H). A strength of the PP assays is that (unlike the H and UT assays) they will detect multiple MOA, although it may not be possible to isolate the mechanism of action. The female PP assay is likely to detect (anti)androgens in addition to E,T,S modalities, although androgens and anti-androgens were not included in the validation studies of the female assay. The validation of the female PP assay indicated that sensitivity was high and although it has not been extensively investigated, it appeared to provide a good estimate of the no-observed-effect-concentration/lowest-observed-effect-

concentration (NOEL/LOELs) obtained in studies of similar or longer duration (e.g. the LOAEL for ethinylestradiol in the female PP assay was similar to that for reproductive effects in a multigenerational study) (US EPA, 2007).

1022. A limitation of the validation is that no chemical was shown to be completely negative in the assay. Chloronitrobenzene was included in the validation as a chemical that was expected to be toxic but without endocrine activity, but when tested was positive in the assay, delaying VO, reducing uterine weight, reducing T4 levels and increasing TSH levels. It is not known whether these effects were due to non-specificity of the assay or a real effect on endocrine systems. Other chemicals, however, that were positive for one endocrine system were not necessarily positive on others (e.g. propylthiouracil altered thyroid hormones and thyroid weight but caused no effects on any of the reproductive tract weights or puberty onset). This indicates that false positives are not always seen and helps to reinforce the specificity of the assay.

1023. Experience with serum hormone determinations in Level 4 and Level 5 rodent assays has revealed that their detection/measurement in rodent studies can be challenging. A recent workshop on "Practicability of Hormonal Measurements" was organised by the BfR (Germany) and the finding from this workshop will be published (Kucheryavenko et al., 2018). The OECD Expert Group on Reproductive and Developmental Toxicity recommends that to demonstrate proficiency for thyroid hormones measurement, a laboratory should be able to show results from a separate study using a positive control substance. Laboratories may also submit their calibration curves, standard curves, as well as data on the levels of quantification and detection. This group is also establishing a historical control database with thyroid toxicant positive controls.

When/why the assay may be used

1024. As mentioned above, the female PP assay may be used as part of the US EPA's Tier 1 screening battery as an apical assay to detect interaction with multiple endocrine systems. In this context, its use is primarily for hazard determination. It may also be used as a follow-up assay following positive results in *in vitro* assays (e.g. a positive result in the Steroidogenesis Assay). Positive results in an ER *in vitro* assay would preferably be followed by a UT assay for reasons of animal welfare – UT assays require fewer animals than the female PP assays and are of shorter duration. If there is a need to test in an apical assay, then the PP assay may be chosen, realising the caveat that there is some uncertainty regarding the specificity of the PP assay. Depending on the number of doses used, the PP assay may be used for hazard identification/characterisation. The assay could potentially also be used to investigate or supplement higher tier data, possibly to clarify the MOA. One scenario could be if only limited reproductive data are available (e.g. a study not conducted to modern standards or not containing endpoints for sexual development). Data from female and male PP assays could then be used to investigate the occurrence of endocrine effects. A decision about whether to conduct further animal tests would, however, need to consider whether sufficient supplementary data may be provided by *in vitro* tests.

1025. In order to provide information relevant for assessing whether or not a chemical may fulfil the WHO/IPCS (2002) definition of an endocrine disruptor (ED), the study design has to be sufficiently robust to demonstrate the presence or absence of effects. In the dose selection, the investigator should also consider and ensure that data generated are adequate to fulfil the regulatory requirement across OECD countries as appropriate (e.g. hazard and risk assessment and labelling, ED assessment, etc.). The top dose or concentration should be sufficiently high to give clear systemic (i.e. non endocrine-specific) toxicity in order

to ensure that a wide range of exposures (high to low) is tested. However, endocrine effects observed solely in the presence of clear systemic toxicity should be interpreted with caution and may be disregarded when sufficiently justified to be caused by secondary effects which are unlikely to be due to endocrine activity. The reason for this advice is a concern that some EAS-sensitive assays are being run at doses/concentrations of EASs that are too low to trigger direct impacts on the endocrine system. This guidance document is not the place to address this issue directly, but it should be considered when EAS-sensitive test guidelines (TGs) are revised in the future. In addition, the number and spacing of dose/concentration levels should also be adequate to fulfil the objectives of the study (e.g. to demonstrate dose response relationships if this is required).

Introduction to the table of scenarios

1026. Table C.3.17 gives guidance on a further step to take in the event of a positive (+) or negative (-) result and in the presence of positive (+), negative (-) or equivocal/absent (Eq/0) existing results. "Existing results" are subdivided into "mechanism" and "effects" data (third and fourth columns). The table is divided horizontally into a series of scenarios that represent all the combinations of these events.

1027. The results of the female PP assay are given in the second column. The assay contains multiple endpoints and it is not possible to provide alternative scenarios for all combinations, therefore some discrimination has been attempted by dividing the endpoints into "apical" and "indicators of hormonal activity". The terminology used has been chosen to be consistent between both the non-mammalian and mammalian tests. Both groups have similar biological importance, although the "indicators of hormonal activity" in the mammalian assays are serum hormones and are generally, but not always, more variable than "apical endpoints". "Apical endpoints" are age/body weight at VO, estrus cyclicity parameters, weights of ovaries, uterus, thyroid, pituitary and adrenals; histopathologic changes in ovaries and uterus. "Indicators of hormonal activity" are hormones (T4 and TSH).

1028. Three possible outcomes for a positive result are therefore envisaged in Table C.3.17:

1. indicators of hormonal activity and apical endpoints positive
2. indicators of hormonal activity positive and apical endpoints negative
3. indicators of hormonal activity negative and apical endpoints positive.

1029. A positive result for apical endpoints could be delayed puberty (VO) or biologically significant reductions in uterine weights, accompanied by treatment-related histopathologic changes. A positive result for indicators of hormonal activity could be biologically significant changes in hormone profiles. The multiple endpoints in this assay mean that there is some redundancy in the assay, but this is useful as not all chemicals may affect all endpoints associated with a mechanism of action and there may be site-specific differences in response.

1030. Single isolated changes may be indicative of spurious results, but robust dose response information may not be available, as the TG only requires two dose levels. The guidance on histopathologic changes in endocrine tests (OECD, 2009) may be helpful in interpretation. Such results should be considered with caution, although it is possible that these endpoints may have detected weak effects that were not detected by the apical endpoints in this study but may then be detected in longer term studies.

1031. A negative result for the female PP assay is taken to be the absence of changes in both endocrine relevant indicators of hormonal activity and apical endpoints. In the absence of other pertinent lines of evidence, negative results in this test alone cannot be taken as evidence that the substance is not an ED. Further studies will be required as confirmation.

1032. Equivocal results for the guideline are not considered in the table, partly for brevity, but also because equivocal results are by nature uncertain. A decision must eventually be reached about whether the endocrine endpoints tend to be positive or negative or whether the result must be put to one side and the test repeated (using the same or a different test). Factors which may have interfered with the result (e.g. composition of the diet used, environmental influences) should be considered. Performance criteria (coefficients of variation for the test endpoints) should be checked for compliance with those in the TG. The assay does not include concurrent positive controls, but attempts have been made to mitigate this by including the performance criteria.

Existing data to be considered

1033. Existing "mechanism" *in vitro* data are assumed to be available from ER-, AR- and steroidogenesis-based assays (Level 2). Assays may also be available for interference with thyroid modalities. In practice, it is possible that data from all of these assays may not be available, so judgement will need to be used to decide which assays to perform. Although the current *in vitro* test guidelines do not incorporate metabolic activation, published information on use of metabolic activation systems is available in Jacobs et al. (2008; 2013) and OECD (2008). These methods, however, have not yet been validated.

1034. Existing "effects" data refer to *in vivo* effects that may come from UT assays where a non-physiological animal model is used. In these cases, it should be remembered that these assays are specifically designed to be sensitive to EDs. The immature rodent UT assay is also sensitive to activities other than ER (ant)agonism, including changes resulting from energy intake (Odum et al., 2004). Another possibility is that repeat dose oral toxicity studies, reproduction/developmental toxicity screen tests or read-across from analogues may be available. It is unlikely that the female PP assay will be performed if data from robust higher tier reproductive studies are already available, as the PP assay offers no advantage over these assays. It is possible, though, that the PP assay has been performed to supplement non-robust higher tier data for the reasons given above. Data may also be available on effects in non-mammalian species, although caution should be used when extrapolating between taxa. A chemical causing endocrine effects non-mammalian environmental species (fish, for example) may also have endocrine effects in mammals, but the physiological consequences of the effects are likely to be different.

1035. When considering the results of the female PP assay, all available data should be used in order to reach a conclusion and a weight of evidence approach taken. This may include high throughput screening data, read-across data from structural analogues and quantitative structure activity relationship (QSAR). Several QSAR models for ER and AR binding/activation are now available (see Sections B.1.1.1 and B.1.1.2).

Scenarios: Positive and negative results combined with existing data

1036. The scenarios (A to R) presented in Table C.3.17 represent all the possibilities of positive or negative results in combination with the presence or absence of existing data. The action taken will also depend on the regulatory environment, but the considerations given here are generally science based. Although the female PP assay uses rats, the

well-conserved nature of the hormonal pathways across taxa should be a strong indication that results in this assay may be relevant to other vertebrate species. Effects in laboratory mammal tests are also highly relevant for environmental mammalian species. Wherever possible, the recommended "next step which could be taken" avoids unnecessary animal testing. However, sometimes conducting an animal test will be indicated and then the relevance of species, strain, exposure route route and species-specific metabolism should always be considered. The sensitivity and physiological function of the hormone under investigation in the test species should also be considered. In general, lower level tests should be conducted before higher level tests in order to avoid unnecessary animal usage, unless it is apparent that a Level 5 test will be required anyway or will be needed to establish the evidence to conclude on ED properties. Information on some endocrine-related tumours may be detected more comprehensively in carcinogenicity studies (OECD TG 451/453) (Level 4); for example, detection of certain types of thyroid tumors in the absence of reproductive or developmental effects, as well as substances causing tumors in other endocrine-sensitive tissues. At Level 5, the Extended One-Generation Reproductive Toxicity Study (EOGRTS – OECD TG 443) is the most sensitive reproduction assay for detecting endocrine disruption because it includes evaluation of a number of endocrine endpoints not included in the two-generation study (OECD TG 416) adopted in 2001. It is recognised, however, that some jurisdictions may require a two-generation study. Further considerations specific to each scenario are given in the table.

1037. Scenarios A to C represent positive results in the female PP assay in the presence of positive *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. Each positive female PP result scenario is divided into the three possible outcomes given above. A positive result in the *in vitro* assays in combination with a positive female PP assay is moderate or strong evidence for estrogen/androgen/thyroid/steroidogenesis (E,A,T,S-) mediated activity that may or may not be supported by the *in vivo* effects data. In the absence of robust upper-level data, the next step may be to conduct an upper-level test. In the presence of robust Level 5 data, there may be sufficient evidence to conclude concern for endocrine disruption and therefore no need for further screening. Positive results in the female PP assay may also indicate the potential for endocrine mediated effects in lower vertebrates. These could be followed up with partial life cycle tests such as the Fish Sexual Development Test (FSDT) or the Medaka Extended One-Generation Reproduction Test (MEOGRT); or the Larval Amphibian Growth and Development Assay (LAGDA) if effects are on the thyroid hormone system. *In vivo* assays/tests with negative results should be interpreted with caution as they may either indicate that the tests used do not have sufficient power to detect weak effects or, alternatively, that the effects do not present a concern for endocrine disruption. The possibility of other (non-E,A,T,S) mechanisms should also not be overlooked (e.g. involving other receptors or endocrine axes).

1038. Scenarios D to F represent positive results in the female PP assay in the presence of negative *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. Each positive female PP result scenario is divided into the three possible outcomes given above. Negative results in the *in vitro* assays should be viewed with caution in case a metabolite is responsible for the positive female PP assay. Unless the metabolic profile of the test substance is known, one option may be to conduct these assays with an added metabolising system. If the metabolic profile is known, then a higher level *in vivo* test may be advisable. The choice of tests will depend on the available *in vivo* effects data. Positive results in the female PP assay may also indicate the potential for endocrine mediated effects in lower vertebrates. These could be followed up with partial life cycle tests such as the

FSDT or MEOGRT; or the LAGDA if effects are on the thyroid hormone system. As in Scenarios A to C, *in vivo* assays/tests with negative results should be interpreted with caution as they may either indicate that the tests used do not have sufficient power to detect weak effects or, alternatively, that the effects do not present a concern for endocrine disruption.

1039. Scenarios G to I represent positive results in the female PP assay in the presence of various combinations of missing or equivocal data. Positive results in the female PP assay may also indicate the potential for endocrine mediated effects in lower vertebrates. These could be followed up with partial life cycle tests such as the FSDT or MEOGRT; or the LAGDA if effects are on the thyroid hormone system. Each positive female PP result scenario is divided into the three possible outcomes given above. The next step to take in these eventualities will depend on the nature of the other available data and the jurisdiction in which it is being used. In some cases, equivocal data may be viewed as positive whilst in others it may or may not contribute to the weight of evidence. The interpretation may also depend on the MOA in question and why the data are considered equivocal, e.g. a study that is equivocal for thyroid effects may still be of value in evaluating (anti)androgenic effects. In all three scenarios, the recommended first step is to obtain reliable mechanistic (*in vitro*) data rather than proceed further with *in vivo* testing. Equivocal and missing data are alternative scenarios and two possibilities for the next step are given in most cases, but the nature of equivocal data means that decisions need to be taken on a case-by-case basis. In all cases, the role of metabolism, route of exposure and data from structural analogues should be considered before deciding on the next step.

1040. Scenarios J to L represent negative results in the female PP assay in the presence of positive *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. As a negative result for the female PP is taken to be negative findings for both indicators of hormonal activity and apical endpoints (unlike the situation with positive outcomes), there is only one possible negative outcome. All three scenarios could also arise from a chemical that is positive in *in vitro* assays, but is metabolised to a non-active metabolite leading to negative results in the female PP assay. This should be considered first when investigating the next step. Endocrine active potency may also explain differences between *in vitro* and *in vivo* results (e.g. a weak chemical may give a positive result *in vitro* but may be negative *in vivo*). Positive *in vivo* effects data may involve other E,A,T,S, non-E,A,T,S mechanisms (e.g. involving other receptors or endocrine axes), more sensitive endpoints, greater statistical power or life stages that are more sensitive to the substance than the young adult animals in the female PP assay.

1041. Scenarios M to O represent negative results in the female PP assay in the presence of negative *in vitro* mechanistic data and positive, negative or equivocal *in vivo* effects data. Negative results for all tests (Scenario N) may be sufficient to enable a conclusion of no concern for endocrine disruption. This will depend on the weight of evidence and may not be possible. Where there are positive *in vivo* effects data, there could still be an E,A,T,S-related mechanism, the effects may be related to length of exposure, route of exposure or exposure at different life stages. Other E,A,T,S or non-E,A,T,S mechanisms may also be involved.

1042. Scenarios P to R represent negative results in the female PP assay in the presence of various combinations of missing or equivocal data. As with the positive result scenarios above (see Paragraph 1 039), the next step to take in these eventualities will have to be decided on a case-by-case basis. However, the recommended first step is generally to obtain reliable mechanistic (*in vitro*) data rather than proceed further with *in vivo* testing. In all

cases, the role of metabolism, route of exposure and data from structural analogues should be considered before deciding on the next step.

1043. In all scenarios (A to R), the next step to take to strengthen weight of evidence will depend on the existing information. Table C.3.17 is meant to provide a succinct guide and may not cover all circumstances or possibilities. The scenarios may also suggest that chemicals have simple or single MOA, when in practice they may have multiple endocrine and non-endocrine MOA. In some cases, for example, two opposite modes of simultaneous action (e.g. estrogenic and anti-estrogenic) could, depending on dose, lead to a minimisation or abolition of effects, while in others two different MOA (e.g. estrogenic and anti-androgenic) could potentially reinforce effects. Endocrine pathways interact, mixed effects are common and there are many pathways that cannot be distinguished with currently available TGs. If multiple MOA are suspected, either from the existing results or based on QSAR/read-across/integrated approaches, this should be investigated further if needed for regulatory decision making.

References

Goldman, J.M. et al. (2000), "Endocrine-disrupting chemicals: Prepubertal exposures and effects on sexual maturation and thyroid activity in the female rat. A focus on the EDSTAC recommendations", *Critical Reviews in Toxicology*, Vol. 30/2, pp. 135-196, https://doi.org/10.1080/10408440091159185.

Jacobs, M. et al. (2013), "*In vitro* metabolism and bioavailability tests for endocrine active substances: What is needed next for regulatory purposes?", *ALTEX – Alternatives to Animal Experimentation*, Vol. 30/3, pp. 331-351.

Jacobs, M.N. et al. (2008), "The use of metabolising systems for *in vitro* testing of endocrine disrupters", *Current Drug Metabolism*, Vol. 9/8, pp. 796-826.

Kucheryavenko, O. et al. (2018), "Report from the BfR Expert Hearing on Practicability of Hormonal Measurements", *Archives of Toxicology* (in prep.).

Odum, J. et al. (2004), "Cumulative dietary energy intake determines the onset of puberty in female rats", *Environmental Health Perspectives*, Vol. 112/15, pp. 1472-1480, http://dx.doi.org/10.1289/ehp.7039.

OECD (2008), *Detailed Review Paper on the Use of Metabolising Systems for* In Vitro *Testing of Endocrine Disrupters*, OECD Series on Testing and Assessment, No. 97, OECD Publishing, Paris, https://doi.org/10.1787/9789264085497-en.

OECD (2009), "Guidance document for histologic evaluation of endocrine and reproductive tests in rodents", OECD Series on Testing and Assessment, No. 106, OECD, Paris, www.oecd.org/env/ehs/testing/43411534.pdf.

US EPA (2009), "Endocrine Disrupter Screening Program Test Guidelines OPPTS 890.1450: Pubertal development and thyroid function in intact juvenile/peripubertal female rats", Environmental Protection Agency, Washington, DC.

US EPA (2007), "Integrated summary report for validation of a test method for assessment of pubertal development and thyroid function in juvenile female rats as a potential screen in the Endocrine Disrupter Screening Program Tier-1 battery", Environmental Protection Agency, Washington, DC.

WHO/IPCS (2002), "Global assessment of the state-of-the-science of endocrine disrupters", Damstra, T. et al. (eds.) WHO/PCS/EDC/02.2, World Health Organization, Geneva, www.who.int/ipcs/publications/new_issues/endocrine_disruptors/en.

Table C.3.17. **Pubertal Development and Thyroid Function Assay in Peripubertal Female Rats (female PP assay) (US EPA OPPTS 890.1450): Guidance for scenarios of combinations of results with existing data**

This table represents possible conclusions to be drawn from assay data, and a next step which could be taken if further evidence is required about possible endocrine disrupting properties and/or effects. The guidance offered is not meant to be prescriptive, but provides science-based considerations. It encourages the use of all available data and expert judgement in a weight of evidence approach. Regional and national interpretation of results and "next steps" may vary.

The conclusions are grouped into a series of scenarios (A-R), each scenario representing a different combination of assay results, existing *in vitro* data and existing *in vivo* data. The symbol "+" indicates that the data in question represent a positive result, «-" indicates a negative result, and "Eq/0" indicates that the data are either equivocal or are not available.

Existing results: * "Mechanism (*in vitro* mechanistic data)" assumes that mechanistic data are available from estrogen receptor (ER-), androgen receptor (AR-) and steroidogenesis-based assays (Level 2). Thyroid hormone receptor (TR) and other assays concerning mechanisms of thyroid disruption may be available, but they are not in common use. In practice, data from all assays may not be available and therefore this must be taken into account when deciding on the "next step". Quantitative structure activity relationship (QSAR) predictions of estrogen and androgen binding/activation may be made for some substances.

Existing results: ** "Effects (*in vivo* effects of concern)" assumes effects have been observed in other *in vivo* screens/tests which give rise to concern that the test chemical may be an endocrine disrupter. These may be repeated dose toxicity tests (e.g. OECD TG 407, TG 408 28- and 90-day studies) or read-across from chemical analogues.

*** *Note:* three possible outcomes for a positive result are given:

1. indicators of hormonal activity and apical endpoints positive
2. indicators of hormonal activity positive and apical endpoints negative
3. indicators of hormonal activity negative and apical endpoints positive.

"Apical endpoints" are age/body weight at vaginal opening, estrus cyclicity parameters, weights of ovaries, uterus, thyroid, pituitary and adrenals; histopathologic changes in ovaries and uterus.

"Indicators of hormonal activity" are hormones (T4 and TSH).

Scenarios	Result of female PP assay	Existing results		Possible conclusions: 1) Indicators of hormonal activity and apical endpoints positive 2) Indicators of hormonal activity positive and apical endpoints negative 3) Indicators of hormonal activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
A	+ ***	+	+	1) Increased evidence of (anti)-E,T,S activity (weak, moderate or strong). 2) Possible evidence of (anti)-E,T,S activity, apical endpoints may be less sensitive or unaffected. Increased evidence of (anti)-E,T,S activity. 3) Moderate or strong (anti)-E,T,S activity, indicators of hormonal activity may be less sensitive or unaffected. Increased evidence of (anti)-E,T,S activity.	Perform assay from Level 5 (e.g. Extended One-Generation Reproduction Toxiticy Study [EOGRTS] or two-generation) assay.	If existing data are from a Level 5 assay, there may be sufficient information to conclude evidence of concern for endocrine disruption (the EOGRTS provides the most information; however, for endocrine disrupting chemicals [EDCs] with a carcinogenic potential, OECD TG 451-3 may be more sensitive). Effects on indicators of hormonal activity alone may be indicative of changes not detected by apical endpoints. Effects on apical endpoints alone may indicate estrogen/androgen/thyroid/steroidogenesis (E,A,T,S) modalities or other mechanisms. Possible effects on androgen modality should also be considered. Consider route of exposures for effects data and possible implications of absorption, distribution, metabolism and excretion (ADME) characteristics of the chemical. Endocrine activity possible in lower vertebrates. Consider performing a Fish Sexual Development Test (FSDT), Larval Amphibian Growth and Development Assay (LAGDA) or Medaka Extended One-Generation Reproduction Test (MEOGRT).
B	+	+	–	1) Increased evidence of (anti)-E,T,S activity (weak, moderate or strong). 2) Possible evidence of (anti)-E,T,S activity, apical endpoints may be less sensitive or unaffected. Increased evidence of (anti)-E,T,S activity. 3) Moderate or strong (anti)-E,T,S activity, indicators of hormonal activity may be less sensitive or unaffected. Increased evidence of (anti)-E,T,S activity.	Perform assay from Level 5 (e.g. EOGRTS or two-generation) assay.	Question why there are differences from existing data. Consider route of exposures and possible implications of ADME characteristics of the chemical. If existing data are from a less sensitive assay,n a higher level test may be required. Effects on indicators of hormonal activity alone may be indicative of changes not detected by apical endpoints. Effects on apical endpoints alone may indicate E,T,S modalities or other mechanisms. Possible effects on androgen modality should also be considered. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT.

Scenarios	Result of female PP assay	Existing results		Possible conclusions: 1) Indicators of hormonal activity and apical endpoints positive 2) Indicators of hormonal activity positive and apical endpoints negative 3) Indicators of hormonal activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
C	+	+	Eq/0	1) Increased evidence of (anti)-E,T,S activity (weak, moderate or strong). 2) Possible evidence of (anti)-E,T,S activity, apical endpoints may be less sensitive or unaffected. Increased evidence of (anti)-E,T,S activity. 3) Moderate or strong (anti)-E,T,S activity, indicators of hormonal activity may be less sensitive or unaffected. Increased evidence of (anti)-E,T,S activity.	Perform assay from Level 5 (e.g. EOGRTS or two-generation) assay.	Check data on chemical analogues. Effects on indicators of hormonal activity alone may be indicative of changes not detected by apical endpoints. Effects on apical endpoints alone may indicate E,T,S modalities or other mechanisms. Possible effects on androgen modality should also be considered. Consider route of exposure for female Peripubertal (PP) Assay and follow-up assay. Possible implications of ADME characteristics of the chemical. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT. Equivocal results may indicate chemical has multiple modes of action (MOA).
D	+	–	+	1) Increased evidence of (anti)-E,T,S activity (weak, moderate or strong). Acts via non-endocrine receptor (ER), thyroid hormone receptor (TR), steroidogenesis (S) mechanism or requires metabolic activation for activity. 2) Possible evidence of (anti)-E,T,S activity, apical endpoints may be less sensitive or unaffected. Acts via non-ER, TR, S mechanism or may require metabolic activation for activity. 3) Moderate or strong (anti)-E,T,S activity, indicators of hormonal activity may be less sensitive or unaffected. Acts via non-ER, TR, S mechanism or may require metabolic activation for activity.	Perform *in vitro* ER, androgen receptor (AR), TR, S assays with added metabolising system.	If existing data are from an adequate Level 5 assay, there is sufficient information to conclude evidence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). Effects on indicators of hormonal activity alone may be indicative of changes not detected by apical endpoints. Effects on apical endpoints alone may indicate E,T,S modalities or other mechanisms. Possible effects on androgen modality should also be considered. Consider route of exposures for effects data and possible implications of ADME characteristics of the chemical. Further mechanistic studies may help determine MOA. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT.

Scenarios	Result of female PP assay	Existing results		Possible conclusions: 1) Indicators of hormonal activity and apical endpoints positive 2) Indicators of hormonal activity positive and apical endpoints negative 3) Indicators of hormonal activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
E	+	–	–	1) Increased evidence of (anti)-E,T,S activity (weak, moderate or strong). Acts via non-ER, TR, S mechanism or may require metabolic activation for activity. Route of exposure may account for the differences from existing data. 2) Possible evidence of (anti)-E,T,S activity, apical endpoints may be less sensitive or unaffected. Acts via non-ER, TR, S mechanism or may require metabolic activation for activity. Route of exposure may account for the differences from existing data. 3) Possible evidence of (anti)-E,T,S activity, indicators of hormonal activity may be less sensitive or unaffected. Acts via non-ER, TR, S mechanism or may require metabolic activation for activity. Route of exposure may account for the differences from existing data.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system OR Perform assay from Level 5 (e.g. EOGRTS or two-generation) assay.	Question why there are differences from existing data. Consider route of exposures and possible implications of ADME characteristics of the chemical. If existing data are from a less sensitive assay, a higher level test may be required. Effects on indicators of hormonal activity alone may be indicative of changes not detected by apical endpoints. Effects on apical endpoints alone may indicate other mechanisms. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT.
F	+	–	Eq/0	1) Increased evidence of (anti)-E,T,S activity (weak, moderate or strong). Acts via non-ER, TR, S mechanism or may require metabolic activation for activity. 2) Possible evidence of (anti)-E,T,S activity, apical endpoints may be less sensitive or unaffected. Acts via non-ER,TR, S mechanism or may require metabolic activation for activity. 3) Moderate (anti)-E,A,T,S activity, indicators of hormonal activity may be less sensitive or unaffected. Acts via non-ER, TR, S mechanism or may require metabolic activation for activity.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system OR Perform assay from Level 5 (e.g. EOGRTS or two-generation) assay.	Effects on indicators of hormonal activity alone may be indicative of changes not detected by apical endpoints. Effects on apical endpoints alone may indicate other mechanisms. Check data on chemical analogues. Further mechanistic studies may help determine MOA. If existing data are from an adequate Level 5 assay, question there are why differences. If existing data are from a less sensitive assay, a higher level test may be required. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT. Equivocal results may indicate chemical has multiple MOA.

Scenarios	Result of female PP assay	Existing results		Possible conclusions: 1) Indicators of hormonal activity and apical endpoints positive 2) Indicators of hormonal activity positive and apical endpoints negative 3) Indicators of hormonal activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
G	+	Eq/0	+	1) Increased evidence of (anti)-E,T,S activity (weak, moderate or strong). May act via ER, TR, S mechanism (metabolic activation may be needed). 2) Possible evidence of (anti)-E,T,S activity, apical endpoints may be less sensitive or unaffected. May act via ER, TR, S mechanism (metabolic activation may be needed). 3) Moderate or strong (anti)-E,A,T,S activity, indicators of hormonal activity may be less sensitive or unaffected. Increased evidence of (anti)-E,T,S activity. May act via ER, TR, S mechanism (metabolic activation may be needed).	Perform *in vitro* ER, AR, TR, S assays (for the "0" scenario, otherwise Eq result available) OR Perform *in vitro* ER, AR, TR, S assays with added metabolising system.	If existing data are from a Level 5 assay, there may be sufficient information to conclude evidence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). Effects on indicators of hormonal activity alone may be indicative of changes not detected by apical endpoints. Effects on apical endpoints alone may indicate E,T,S modalities or other mechanisms. Possible effects on A modality should also be considered. Check data on chemical analogues. Further mechanistic studies may help determine MOA. Consider route of exposures for effects data and possible implications of ADME characteristics of the chemical. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT. Equivocal results may indicate chemical has multiple MOA.
H	+	Eq/0	–	1) Increased evidence of (anti)-E,T,S activity (weak, moderate or strong). Acts via unknown mechanism or may require metabolic activation for activity. Route of exposure may account for the differences from existing data. 2) Possible evidence of (anti)-E,T,S activity, apical endpoints may be less sensitive or unaffected. Acts via unknown mechanism. Weak activity does not result in adverse effects. 3) Moderate (anti)-E,T,S activity, indicators of hormonal activity may be less sensitive or unaffected. Acts via unknown mechanism. Weak activity does not result in adverse effects.	For the "0" scenario, perform *in vitro* ER, AR, TR, S assays with added metabolising system (otherwise Eq result available).	Question why there are differences from existing data. Consider route of exposures and possible implications of ADME characteristics of the chemical. If existing data are from a less sensitive assay, then a higher level test may be required. Effects on indicators of hormonal activity alone may be indicative of subtle changes not detected by apical endpoints. Effects on apical endpoints alone may indicate E,T,S modalities or other mechanisms. Possible effects on A modality should also be considered. Check data on chemical analogues. Further mechanistic studies may help determine MOA. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT. Equivocal results may indicate chemical has multiple MOA.

Scenarios	Result of female PP assay	Existing results		Possible conclusions: 1) Indicators of hormonal activity and apical endpoints positive 2) Indicators of hormonal activity positive and apical endpoints negative 3) Indicators of hormonal activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
I	+	Eq/0	Eq/0	1) Increased evidence of (anti)-E,T,S activity (weak, moderate or strong). Acts via unknown mechanism. Unknown potential for adverse effects. 2) Possible evidence of (anti)-E,T,S activity, apical endpoints may be less sensitive or unaffected. Acts via unknown mechanism. Unknown potential for adverse effects. 3) Moderate or strong (anti)-E,T,S activity, indicators of hormonal activity may be less sensitive or unaffected. Acts via unknown mechanism. Unknown potential for adverse effects. There may be a need for metabolic activation.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system OR Perform assay from Level 5 (e.g. EOGRTS or two-generation) assay.	Effects on indicators of hormonal activity alone may be indicative of changes not detected by apical endpoints. Effects on apical endpoints alone may indicate E,T,S modalities or other mechanisms. Possible effects on androgen modality should also be considered. Check data on chemical analogues. Further mechanistic studies may help determine MOA. Endocrine activity possible in lower vertebrates. Consider performing an FSDT, LAGDA or MEOGRT. Equivocal results may indicate chemical has multiple MOA.
J	–	+	+	No evidence for E,T,S activity in female PP assay. Metabolism or potency may explains the difference from existing *in vitro* and *in vivo* data. Effects seen in existing studies are via non-E,T,S mechanism.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system OR Perform assay from Level 5 (e.g. EOGRTS or two-generation) assay.	If existing data are from an adequate Level 5 assay, question why there are differences. If data are from Uterotrophic Bioassays (UT) then this may be more sensitive than female PP assay. Effects seen in existing studies may be in a more sensitive life stage. Consider route of exposures and possible implications of ADME characteristics of the chemical. Further mechanistic studies may help determine MOA.
K	–	+	–	No evidence for E,T,S activity in female PP assay. Metabolism or may potency explain *in vitro*/*in vivo* differences.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system OR Perform assay from Level 5 (e.g. EOGRTS or two-generation) assay.	If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). If existing data are from a less sensitive assay, then a higher level test may be required. If data are from UT assay, then need to conduct higher tier assay to conclude absence of concern for endocrine disruption. Further mechanistic studies with metabolism may help determine MOA.

Scenarios	Result of female PP assay	Existing results		Possible conclusions: 1) Indicators of hormonal activity and apical endpoints positive 2) Indicators of hormonal activity positive and apical endpoints negative 3) Indicators of hormonal activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
L	–	+	Eq/0	No evidence for E,T,S activity in female PP assay. Metabolism or potency may explain *in vitro*/*in vivo* differences. Unknown potential for adverse effects.	Perform *in vitro* ER, AR, TR, S assays with added metabolising system OR Perform assay from Level 5 (e.g. EOGRTS or two-generation) assay.	Metabolic deactivation of chemical may occur *in vivo* so that potential *in vitro* activity is not realised. Consider possible routes of exposure implications of metabolism. Equivocal results may indicate chemical has multiple MOA.
M	–	–	+	No evidence for E,T,S activity in female PP assay. Effects seen in existing studies are via non-E,T,S mechanism.	Perform assay from Level 5 (e.g. EOGRTS or two-generation) assay.	If existing data are from an adequate Level 5 assay, question why there are differences. If data are from UT assay, then this may be more sensitive than female PP assay. Effects seen in existing studies may be in a more sensitive life stage. Consider route of exposures and possible implications of ADME characteristics of the chemical.
N	–	–	–	No evidence for E,T,S activity in female PP assay. No evidence for (anti)-E,T,S activity *in vitro*. No evidence of adverse effects.	Possibly no need for further testing. If there is uncertainty, may perform assay from Level 5 (e.g. EOGRTS or two-generation) assay.	If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive).
O	–	–	Eq/0	No evidence for E,T,S activity in female PP assay. No evidence for (anti)-E,T,S activity *in vitro*. Unknown potential for adverse effects.	Perform assay from Level 5 (e.g. EOGRTS or two-generation) assay.	Consider route of exposures and possible implications for ADME characteristics of the chemical in follow-up assay.
P	–	Eq/0	+	No evidence for E,T,S activity in female PP assay. Potential for adverse effects via unknown mechanism.	Perform *in vitro* ER, AR,TR, S assays.	Consider route of exposure and possible implications for differences from existing assay. If data are from Hershberger Bioassay (H), then this may be more sensitive than female PP assay. Effects seen in existing studies may be in a more sensitive life stage. Equivocal results may indicate chemical has multiple MOA.

Scenarios	Result of female PP assay	Existing results		Possible conclusions: 1) Indicators of hormonal activity and apical endpoints positive 2) Indicators of hormonal activity positive and apical endpoints negative 3) Indicators of hormonal activity negative and apical endpoints positive	Next step which could be taken to strengthen weight of evidence if necessary	Other considerations
		Mechanism (*in vitro* mechanistic data)*	Effects (*in vivo* effects of concern)**			
Q	–	Eq/0	–	No evidence for E,T,S activity in female PP assay. No evidence of adverse effects.	Perform *in vitro* ER, AR,TR, S assays.	If existing data are from an adequate Level 5 assay, there may be sufficient information to conclude absence of concern for endocrine disruption (the EOGRTS provides the most information; however, for EDCs with a carcinogenic potential, OECD TG 451-3 may be more sensitive). Further mechanistic studies may strengthen weight of evidence.
R	–	Eq/0	Eq/0	No evidence for E,T,S activity in female PP assay.	Perform *in vitro* ER, AR, TR, S assays, otherwise Eq result available.	Further mechanistic studies may strengthen weight of evidence. Check data on chemical analogues. Equivocal results may indicate chemical has multiple MOA.

Glossary

Apical endpoints	Results of an *in vivo* assay which describe a response by the organism as a whole (e.g. fecundity or growth) which have possible implications for its biological fitness, rather than a response of the endocrine system alone (including physiological changes dependent on the endocrine system, such as vitellogenin induction). Apical responses may or may not result from endocrine changes (e.g. fecundity may be affected both by some endocrine disrupters and by some non-endocrine disrupters).
Assay	An experimental system that can be used to obtain a range of information from chemical properties through the adverse effects of a substance. The terms "assay" and "test method" may be used interchangeably for non-mammalian as well as for mammalian studies (OECD, 2005).
EASZY	Detection of Substances Acting through Estrogen Receptors using Transgenic cyp19a1b GFP Zebrafish Embryos.
Endocrine active substance	A substance having the inherent ability to interact or interfere with one or more components of the endocrine system resulting in a biological effect, but need not necessarily cause adverse effects (EFSA, 2013).
Endocrine disruption	"An [endocrine disrupter] is an exogenous substance or mixture that alters function(s) of the endocrine system and consequently causes adverse health effects in an intact organism, or its progeny, or (sub)populations." "A potential [endocrine disrupter] is an exogenous substance or mixture that possesses properties that might be expected to lead to endocrine disruption in an intact organism, or its progeny, or (sub)populations." (WHO, 2002)
Epigenesis	Inherited changes in phenotype or gene expression caused by mechanisms other than alteration in gene sequences (e.g. DNA methylation).
Hazard characterisation	The qualitative and, wherever possible, quantitative description of the inherent property of an agent or situation having the potential to cause adverse effects. This should, where possible, include a dose-response assessment and its attendant uncertainties (IPCS/WHO, 2004).

Hazard identification	The identification of the type and nature of adverse effects that an agent has an inherent capacity to cause in an organism, system, or (sub)population (IPCS/WHO, 2004).
***In vitro* assay**	Assay where whole live animals are not used. Systems used may include cell lines or subcellular preparations from untreated animals.
***In vivo* assay**	Assay where a whole live animal is treated. This may be a mammalian assay where individual animals are treated or a non-mammalian assay where a population of animals is treated.
Indicators of hormonal activity	These are endpoints in an *in vivo* assay which show whether or not the endocrine system has been stimulated, and often provide information of mechanistic value. In other words, they are not apical endpoints (see definition above). It is possible in some cases for indicators of hormonal activity to respond to a test chemical while apical endpoints do not respond, while in other cases, both types of endpoint give a response or only apical endpoints respond.
Mode of action	A set of key events and processes starting with the interaction of an agent with a cell, through physiological and tissue or organ changes, resulting in an adverse outcome (Dellarco and Fenner-Crisp, 2012).
Non-mammalian	In this context (non-mammalian screens and tests), the test species are fish, amphibians and birds.
Screen	*In vitro* or *in vivo* assays which provide information on an endocrine disruption mechanism, but not generally information on adverse effects, for use in hazard identification/characterisation. Screens are generally rapid, simple test methods and may have a truncated response range.
Test	*In vivo* assays which can provide evidence to support a conclusion that a chemical is an endocrine disrupter that can cause adverse effects in an intact organism.
Validated assay (also equivalent to validated test method)	A test method for which validation studies have been completed to determine the relevance (including accuracy) and reliability for a specific purpose. It is important to note that a validated test method may not have sufficient performance in terms of accuracy and reliability to be found acceptable for the proposed purpose (OECD, 2005).
Validation	The process by which the reliability and relevance of a particular approach, method, process or assessment is established for a defined purpose (OECD, 2005).

Weight of evidence	In the context of this document, this implies that all relevant data from the test being evaluated, and from other tests on the chemical in question, should be considered before taking decisions about interpretation of the new data and the possible need for additional testing. Each datum is not necessarily given equal weight – such weighting will depend on the type and reliability of the datum. Although it is possible to provide guidelines for weight of evidence assessment, its effective use will always depend to some extent on the application of expert judgement.

About the OECD

The Organisation for Economic Co-operation and Development (OECD) is an intergovernmental organisation in which representatives of 34 industrialised countries in North and South America, Europe, and the Asia and Pacific region, as well as the European Commission, meet to co-ordinate and harmonise policies, discuss issues of mutual concern, and work together to respond to international problems. Most of the OECD's work is carried out by more than 200 specialised committees and working groups composed of member country delegates. Observers from several countries with special status at the OECD, and from interested international organisations, attend many of the OECD's workshops and other meetings. Committees and working groups are served by the OECD Secretariat, located in Paris, France, which is organised into directorates and divisions.

The Environment, Health and Safety Division publishes free-of-charge documents in 12 different series: Testing and Assessment; Good Laboratory Practice and Compliance Monitoring; Pesticides; Biocides; Risk Management; Harmonisation of Regulatory Oversight in Biotechnology; Safety of Novel Foods and Feeds; Chemical Accidents; Pollutant Release and Transfer Registers; Emission Scenario Documents; Safety of Manufactured Nanomaterials; and Adverse Outcome Pathways. More information about the Environment, Health and Safety Programme and EHS publications is available on the OECD's website (www.oecd.org/chemicalsafety).

This publication was developed in the IOMC context. The contents do not necessarily reflect the views or stated policies of individual IOMC participating organisations.

The Inter-Organisation Programme for the Sound Management of Chemicals (IOMC) was established in 1995 following recommendations made by the 1992 United Nations Conference on Environment and Development to strengthen co-operation and increase international co-ordination in the field of chemical safety. The participating organisations are the Food and Agriculture Organization, the International Labour Organization, the United Nations Environment Programme, the United Nations Industrial Development Organization, the United Nations Institute for Training and Research, the World Health Organization, the World Bank and the OECD. The United Nations Development Programme is an observer. The purpose of the IOMC is to promote co-ordination of the policies and activities pursued by the participating organisations, jointly or separately, to achieve the sound management of chemicals in relation to human health and the environment.

This publication is available electronically, at no charge.

For this and many other Environment, Health and Safety publications, consult the OECD's website: www.oecd.org/ehs

or contact:

OECD Environment Directorate,
Environment, Health and Safety Division
2 rue André Pascal
75775 Paris Cedex 16
France

Fax: (33-1) 44 30 61 80

E-mail: ehscont@oecd.org

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

The OECD is a unique forum where governments work together to address the economic, social and environmental challenges of globalisation. The OECD is also at the forefront of efforts to understand and to help governments respond to new developments and concerns, such as corporate governance, the information economy and the challenges of an ageing population. The Organisation provides a setting where governments can compare policy experiences, seek answers to common problems, identify good practice and work to co-ordinate domestic and international policies.

The OECD member countries are: Australia, Austria, Belgium, Canada, Chile, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Korea, Latvia, Lithuania, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, the Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Turkey, the United Kingdom and the United States. The European Union takes part in the work of the OECD.

OECD Publishing disseminates widely the results of the Organisation's statistics gathering and research on economic, social and environmental issues, as well as the conventions, guidelines and standards agreed by its members.

OECD Series on Testing and Assessment

Revised Guidance Document 150 on Standardised Test Guidelines for Evaluating Chemicals for Endocrine Disruption

This guidance document was originally published in 2012 and updated in 2018 to reflect new and updated OECD test guidelines, as well as reflect on scientific advances in the use of test methods and assessment of the endocrine activity of chemicals. The document is intended to provide guidance for evaluating chemical using standardised test guidelines. Specific objectives include providing a description of the OECD conceptual framework for evaluating chemicals for endocrine disruption, background on the standardised test methods used, and guidance for interpreting the outcome of individual tests. The general approach taken by the document is primarily to provide guidance on how test results might be interpreted based on the outcome of standardised assays. Key questions addressed in the document concern likely mechanisms of endocrine action and any resulting apical effects that can be attributed to such action. The document is not proscriptive but provides suggestions for possible next steps in testing (if any) which might be appropriate for a regulatory authority to take, given the various data scenarios. The guidance document is focused primarily on endocrine modalities included in the conceptual framework; estrogen, androgen, and thyroid mediated endocrine disruption and chemicals that interfer with steroidogenesis.

Consult this publication on line at ***https://doi.org/10.1787/9789264304741-en***.

This work is published on the OECD iLibrary, which gathers all OECD books, periodicals and statistical databases.
Visit ***www.oecd-ilibrary.org*** for more information.

OECD publishing
www.oecd.org/publishing

ISBN 978-92-64-30474-1
97 2018 40 1 E