Feasibility study for minor enhancements of TG 414 (Prenatal Developmental Toxicity Study) with ED-relevant endpoints
December, 2017 Draft (for 2nd WNT comments)

Sofie Christiansen & Ulla Hass
Division of Diet, Disease prevention and Toxicology, National Food Institute,
Technical University of Denmark
Table of contents

Terms of reference ............................................................................................................................................ 4
Aim ..................................................................................................................................................................... 4
Background and expected regulatory need/data requirement that will be met by the proposed outcome of the project ......................................................................................................................................................... 4
Anogenital distance (AGD) ................................................................................................................................ 7
Methodology ................................................................................................................................................. 7
Data analysis, sensitivity and power ............................................................................................................. 7
Data collection ........................................................................................................................................... 8
Results ........................................................................................................................................................... 9
Human relevance ......................................................................................................................................... 12
Animal welfare ............................................................................................................................................. 13
Inclusion of AGD in TG 414 .......................................................................................................................... 13
Testosterone levels in male fetuses ................................................................................................................ 15
Method ........................................................................................................................................................ 15
Data analysis, sensitivity/power .................................................................................................................. 16
Human relevance ......................................................................................................................................... 17
Animal welfare ............................................................................................................................................. 18
Inclusion of Testosterone in TG 414 ............................................................................................................ 18
Thyroid hormones ........................................................................................................................................... 18
Method ........................................................................................................................................................ 18
Data analysis, sensitivity/power .................................................................................................................. 19
Human relevance ......................................................................................................................................... 19
Animal welfare ............................................................................................................................................. 20
Inclusion of thyroid hormones in TG 414 .................................................................................................... 20
Abnormalities of external genital organs ........................................................................................................ 21
Method ........................................................................................................................................................ 21
Data analysis, sensitivity/power .................................................................................................................. 21
Human relevance ......................................................................................................................................... 21
Animal welfare ............................................................................................................................................. 22
Inclusion abnormalities of external genital organs in TG 414 ..................................................................... 22
Overall discussion and conclusions ................................................................................................................. 23
References ....................................................................................................................................................... 24
Appendix 1 Text changes suggestions for TG 414 shown with track changes ................................................. 29
Terms of reference

1. This feasibility report has initially been prepared by the National Food Institute, Technical University of Denmark, DK that is leading the project in OECD. The report gives input for discussions in the OECD expert group on reproductive toxicity involved in the project Feasibility study for minor enhancements of TG 414 (Prenatal Developmental Toxicity Study) with ED-relevant endpoints. Subsequently, the report will be revised based on input after EDTA meetings, WNT commenting rounds in OECD and discussions in the OECD expert group on reproductive toxicity.

Aim

2. The aim of this project is to do a Feasibility study for minor enhancements of TG 414 (Prenatal Developmental Toxicity Study) with ED-relevant endpoints. This review addresses scientific and technical concerns regarding inclusion of additional ED related endpoints in TG 414. The endpoints considered include anogenital distance (AGD) in all fetuses, Testosterone in male fetuses and thyroid hormones and guidance for genital malformations in all fetuses. Moreover measurement of thyroid hormones in the dams will be considered to be included.

For these endpoints, the scientific and technical questions considered include:

• Are standardized methods available?
• Is the sensitivity sufficient with the number of litters per group?
• Are the endpoints of relevance for humans?
• Are there animal welfare concerns?
• Is the enhancement possible without changes or with only minor changes in study design?

Background and expected regulatory need/data requirement that will be met by the proposed outcome of the project

3. A scientific approach will be used to give input to the existing TG 414 (Prenatal Developmental Toxicity Study) in relation to the feasibility of inclusion of sensitive endpoints in all fetuses and dams for detection of chemicals with endocrine disrupting properties.

4. The specific purpose of this project is to consider the relevance and feasibility of enhancement of the OECD 414 (OECD, 2001). The TG 414 provides information on adverse effects on prenatal development and is used in various regulatory frameworks (such as REACH and several pesticide regulations) to generate information for risk assessment of chemicals.
5. OECD TG 414 is included in Level 4 (OECD conceptual framework) as the TG involves repeated dosing of pregnant females and therefore potential exposure of the developing fetus. The assay includes some endpoints that may detect endocrine disruption (e.g. abnormalities of male and female genitalia) (OECD GD 150, 2012a, currently under revision).

6. However, an update in relation to inclusion of more endpoints relevant for endocrine disruption would increase the possibility for detecting effects of endocrine disrupting substances. Assessments of testosterone levels in serum and anogenital distance (AGD) in male fetuses have been used in several published studies and appear to be sensitive endpoints for detecting effects of endocrine disrupters with anti-androgenic properties. Thus, inclusion of these endpoints in TG 414 in dams or fetuses at the time of caesarean section could be a significant enhancement with regards to detection of effects of endocrine disrupting substances. Also, an enhancement with some additional text giving guidance on evaluation of abnormalities of male and female genitalia would be relevant to include.

7. After the inclusion on the workplan in OECD (as project number 4.100) the possibility to include thyroid hormones in the dams and/or fetuses has been mentioned in the OECD expert group on reproductive toxicity as well as at international meetings in 2017 (Thyroid workshop report, 2017; Priority setting workshop report, 2017) and EDTA meetings in May and October 2017.

8. The expert group on reproductive toxicity (EG) from the update on TG 421/422 (Project 4.71 on OECD Workplan) has been convened as the scientific discussion is similar. Additional experts have been invited to the EG via the WNT NCs. These EG and the EDTA meetings in 2017 (May and October) have provided guidance on which endpoints to be considered and how this can be done (e.g. timing and logistics) based on a proposal from the lead.

9. DK has undertaken the examination of available existing data. Data have been received from OECD countries and also peer reviewed scientific relevant papers have been included to make a proposal to the EG on whether or not it is relevant to include the ED related endpoints in a proposal for revision of OECD TG 414.

10. It will also be considered whether certain slight adaptions of the test design of the test guideline may be warranted to include for consideration if other ED related endpoints are suggested by the EG for this project. However, the timing of the OECD TG 414 study cannot be changed as was the case in TGs 421/422 screening studies (which was terminated later due to assessment of Nipple retention).

11. The results of this project may contribute to an improved sensitivity for identification of developmental toxicants in mammalian species at an early stage in the regulatory testing schemes for industrial chemicals (e.g. REACH) as information from TG 414 is already required in such regulatory testing schemes.

12. If these endpoints are implemented in TG 414 it will enhance the international harmonization of hazard assessment with regard to developmental toxicity effects.
13. An important point is that the ability for detection of EDs should be enhanced without increasing the number of experimental animals used.

14. TG 414 is designed to provide general information concerning the effects of prenatal exposure 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 fetus. The proposed update of TG 414 must not impair the ability to fulfil the purpose of TG 414.

15. Assessment of AGD in both sexes is mandatory in TG 443 and TGs 421/422 and this report will elucidate whether this endpoint could also be included in TG 414 at the day of caesarean section. See below for more information.

16. TG 414 was revised in 2001 but not with regard to inclusion of ED relevant endpoints. It seems relevant to include some ED relevant endpoints in TG 414 as the exposure periods cover some of the sensitive periods for sexual differentiation (prenatal period). Evaluation for serum testosterone during this time frame would allow for evaluation of androgenic chemical activity and the potential for masculinization of female embryos. The proposed endpoints are described below.

17. The OECD TG 407 (Repeated dose 28-day oral toxicity study in rodents) has been updated in 2008. The assay has been validated for some endocrine endpoints but the sensitivity of the assay is not sufficient to identify all EATS-mediated EDs. The validation of the assay (OECD, 2006) showed that it identified strong and moderate EDs acting through the ER and AR; and EDs weakly and strongly affecting thyroid function. It was relatively insensitive to weak EDs acting through the ER and AR. This assay also has some optional endpoints such as uterine and ovary weight, vaginal cytology (to provide a marker for physiological estrogenicity that would guide the histological interpretation of the ovary and estrogen-sensitive tissues), histopathologic changes in mammary gland histopathology (mandatory in females optional in males) as well as serum T3, T4, TSH as well as thyroid weight which can be examined if there is additional concern.

18. The extended one-generation reproductive toxicity study (EOGRTS) (OECD, 2012b) includes more endpoints sensitive to endocrine disruption than OECD TG 416 and, as it also uses reduced animal numbers if conducted without F2. It is expected that it will often replace OECD TG 416 for mammalian reproductive toxicity testing (OECD GD 150, 2012a). Endpoints sensitive to endocrine disruption, not specified in OECD TG 416, include anogenital distance at birth, areola/nipple retention, measurement of thyroid hormones and TSH levels. Effects on the developing nervous and immune systems are also assessed by the DNT and DIT cohorts. These systems may also be sensitive to endocrine influences. This test is also expected to have greater sensitivity than OECD TG 416 as it requires an increased number of pups to be examined. In summary, the new 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 2-generation study (OECD TG 416) adopted in 2001.
19. In 2015 and 2016, the OECD 421 (Reproduction/Developmental Toxicity Screening Test) and OECD 422 (Combined Repeated Dose Toxicity Study with the Reproduction/Developmental Toxicity Screening Test) guidelines were revised to include estrogen, androgen, thyroid, and steroidogenesis (EATS) relevant endpoints (OECD, 2016a; OECD 2016b; OECD, 2015).

20. In April 2015, OECD launched this feasibility study for the enhancement of OECD 414 (Prenatal Developmental Toxicity Study) (project 4.100 on the OECD workplan) with selected parameters intended to increase the detection of EATS disrupting potential. In the autumn 2015 the lead and OECD secretariat requested data (control litter means + SD) from OECD member states to enhance TG 414 with endocrine disrupter relevant endpoints.

**Anogenital distance (AGD)**

**Methodology**

21. New-born male rats have no scrotum, and the external genitalia are undeveloped, and only a genital tubercle is apparent for both sexes. The AGD is the distance from the anus to the insertion of this tubercle, the developing genital bud. The AGD is androgen dependent, and studies show that the AGD is normally about twice as long in male as in female rats. Similarly, in new-born humans the AGD measure was about two-fold greater in males than in females (Salazar-Martinez et al. 2004). At caesarean section, the distance between the proximal end of the anus and the genital tubercle of all fetuses must be measured. Anogenital distance is a non-invasive measure of in utero androgenicity (Wolf et. al. 2005). The distance will be measured from the base of the genital tubercle to the proximal end of the anal opening using a dissecting microscope with a micrometer eyepiece or another sensitive method. The fetal weight will be measured to derive anogenital index.

22. In TG 414 the AGD will be measured in both male and female fetuses in all litters at caesarean section one day prior to the expected day of delivery (ref. para. 22 in TG 414).

**Data analysis, sensitivity and power**

23. Important parameters when evaluating the sensitivity and power of the data are the standard deviation, SD (σ) and coefficient of variation

24. The standard deviation is an expression of how much the value disperses from the population mean (μ). Coefficient of variation is expressed as σ/μ and is often evaluated as a percentage and therefore expresses the standard deviation as the percentage of the population mean, μ.

25. In order to take into account the size of the rat when evaluating the AGD, the AGD was divided by the cubic root of the body weight, i.e. \((\text{AGD [mm]})/\sqrt[3]{(\text{body weight [g]})}\) resulted in the anogenital distance index (hereafter AGDI).
26. The rats grows rapidly, growth occurs in three dimensions so that body weight can be viewed as a cubic measure. In contrast, AGD is a purely linear metric. Therefore the relationship between AGD and body weight should be more properly evaluated using the cube root of body weight as with AGDI or using the bodyweight as covariate in statistics (OECD, 2015).

Data collection

27. The evaluation of both AGD and plasma testosterone levels as endpoints is based on data from studies at DTU National food institute (only published), external (non-published data from the expert group) and published studies primarily from Saillenfait et al. from The French Research and Safety Institute for the prevention of occupational accidents and diseases (INRS) (Saillenfait, et al. 2009; Saillenfait et al. 2010; Saillenfait et al., 2011a and 2011b; Saillenfait, et al. 2013). The references for the data from DTU data is given as (Taxvig, et al. 2013; Hansen, et al. 2009; Kristensen, et al. 2011; Taxvig et al. 2007; Taxvig et al. 2008).

28. Unfortunately, the data sent from the expert group (unpublished) could not be included in the analysis. One dataset missed the information about litter affiliation and thereby group means based on litter means could not be obtained. The other dataset did not give any details on the strain of rodent and could not be compared with others.

29. Furthermore, a literature search was conducted to increase the amount of data. The parameters for this search were based on the dosing period requirements in TG 414. “Normally, the test substance should be administered daily from implantation (e.g., day 5 post mating) to the day prior to scheduled caesarean section […] Females should be killed one day prior to the expected day of delivery.” (OECD, 2001). The dosing period for the rats used in the DTU group (in house data) was GD 7 to 21 and GD 6-21 for the French group. GD 0 is understood as the day of mating and GD 21 is 3 weeks after (our Wistar or SD rats give birth day 22 or 23).

30. In general, only few other published papers satisfied this dosing period requirement, either because the administration of the compound was initiated later on (e.g. GD 12-GD 17), or because the caesarean section was conducted several days before the expected day of delivery, most often at GD 19 (Parks et al., 2000; Ema et al., 2003; Fisher et al., 2003; Hotchkiss et al., 2004; Thompson, Ross and Gaido, 2004; Liu et al., 2005; Saillenfait et al., 2016). A reason for this deviation could be the interest in the developmental stages in the embryotic phase or in the testosterone peak levels during the fetal development.

31. Another important criterion was the presence of data in the papers that could actually be used for this analysis. Hence, data presentation in the form of graphs or bar charts was excluded because of inaccuracies when transforming these and therefore only data in the form of tables was included.
Results

32. To evaluate whether a specific endpoint is sensitive enough to be included in a test guideline, it is important to consider the control values. If large variations are present in these samples, the identification of any effect is hampered.

Table 1. Overview of male AGD in control data GD 21, mean ± SD. Averages for all control group means, control group standard deviations and coefficients of variation for the control groups are given. In relation to all calculated averages, the standard deviation is depicted. The total number of studies and the range of the number of litters for the control groups are also given. Importantly, the controls are also split with regards to the rat strain used. The data AGDI mean is a group mean based of litter means in each study.

<table>
<thead>
<tr>
<th></th>
<th>AGDI mean [mm/g]</th>
<th>AGDI standard deviation [mm/g]</th>
<th>AGDI Coefficient of variation [%]</th>
<th>Control experiments (range of litters)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All control experiments</td>
<td>2.00 ± 0.36</td>
<td>0.08 ± 0.04</td>
<td>4.03 ± 1.67</td>
<td>17 (3-25)</td>
</tr>
<tr>
<td>Control experiments with Wistar rats</td>
<td>2.31 ± 0.14</td>
<td>0.10 ± 0.04</td>
<td>4.49 ± 1.69</td>
<td>9 (3-18)</td>
</tr>
<tr>
<td>Control experiments with Sprague-Dawley rats</td>
<td>1.72 ± 0.21</td>
<td>0.07 ± 0.02</td>
<td>3.99 ± 1.09</td>
<td>8 (6-25)</td>
</tr>
</tbody>
</table>

33. An overview of all control data of AGD at GD 21 including the overall mean, standard deviation and coefficient of variation are given in Table 1.
As seen in Table 1, the combined Coefficient of Variation in male control fetuses is low (overall 4.03±1.67) even though most studies have much fewer litters than in TG 414. However, it is also observed (Table 1, Figure 1) that Sprague-Dawley rats in the French studies have a lower mean AGDI, standard deviation and coefficient of variation compared to Wistar rats.

Figure 1. Mean AGDI +SD at GD 21 male fetuses, i.e. the mean with corresponding SD for each control group for the studies included. Blue bars indicate Wistar rats (N=3-18 litters) and red bars indicate Sprague-Dawley rats (6-25 litters). The DTU group in general uses Wistar rats (only one study with Sprague Dawley), and the French research group uses Sprague-Dawley rats only.
35. To evaluate whether the prenatal AGD measurement is sensitive enough it is also relevant to compare to postnatal data. This endpoint is not feasible to include if the prenatal measurements has a much higher variation than postnatal measurement. The measurements must be consistent with standard deviations and coefficients of variation not greatly influenced by the gestation day compared to postnatal measurements. Table 2 and figure 2 outlines such a comparison between coefficients of variation for AGDI measured pre- and postnatally.

36. As seen in table 2 the CoV for prenatal AGDI is 4.49 (9 studies) whereas in new born males it is 4.00 (23 studies). This indicates that the sensitivity/power for detecting effect on AGD is rather similar in GD 21 fetuses and new born male pups. Thus the power analysis performed in relation to inclusion of AGD in new-born males in TG 421/422 is also useful for this TG. Most of the studies included to derive CoV for AGD in this report used fewer litters than in TG 414.

Table 2. Comparison of coefficients of variation for AGDI measured prenatally and postnatally. The given coefficients of variation are group means based on litter means of all control group coefficients, mean ± SD. The total number of studies/experiments and the range of number of litters for the control groups are given. Only data on Wistar rats (from table 1) is included from DTU, DK to be sure that data is from same laboratory.

<table>
<thead>
<tr>
<th></th>
<th>AGDI Coefficient of variation</th>
<th>Control experiments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prenatal AGDI</td>
<td>4.49 ± 1.69</td>
<td>9 (3-18)</td>
</tr>
<tr>
<td>Postnatal AGDI</td>
<td>4.00 ± 1.50</td>
<td>23 (3-21)</td>
</tr>
</tbody>
</table>
Figure 2. Mean coefficient of variation (group means based on litter means in each study) for prenatal studies: 4.49 ±1.69 % (9) and postnatal studies: 4.00 ±1.50 % (23 studies). For the prenatal studies 2 CoVs have the same value and therefore it looks like there are only 8 studies.

Human relevance

37. In rats, both AGD and nipple retention have been shown to be highly predictive of or correlated to adverse effects of the male reproductive system including increased incidence of genital malformations (dysgenesis or hypospadias), and or altered reproductive organ weight changes (Bowman et al. 2003, Christiansen et al. 2008, van den Driesche et al. 2011, Welsh et al. 2008). Some studies have reported changes in female AGD. In two novel studies slight reductions in female pup AGD following butylparaben exposure (Boberg et al. 2016) and bisphenol A exposure (Christiansen et al., 2014) have been reported. In contrast, exposure to ethinyl estradiol has induced increased AGD and number of retained nipples of female offspring (Mandrup et al., 2013) and prochloraz exposure in utero has increased female AGD in several studies (Laier et al., 2006; Melching-Kollmuss et al., 2017). Therefore AGD in both sexes is suggested included in this TG.

38. A shorter anogenital distance in humans has been shown to be associated with phthalate exposure (Bornehag et al. 2015). Recent studies reported that male infants and boys with adverse effects such as hypospadias or undescended testis also had reduced AGD (Hsieh et al. 2012; Hsieh et al. 2008; Jain and Singal 2013; Thankamony et al. 2013). Moreover, a shorter AGD in adult men has been related to decreased fertility (Eisenberg et al. 2011), impaired semen quality (Mendiola et al. 2011) and decreased serum testosterone levels (Eisenberg et al. 2012). Shortened AGD has also been suggested as a biomarker of testicular dysgenesis syndrome (Sharpe 2005).

39. As AGD in both sexes is included as an endpoint in OECD TG 443 and TG 421/422 it can therefore be considered as an endpoint evaluated to be of human relevance. Moreover EC-
HA have in a newer evaluation stated that: “The findings in AGD, nipple retention and foetal T, suggest an anti-androgenic mode of action (androgen deficiency) and may be considered as relevant findings and predictors of potential adverse effect during human development.” (ECHA 2013). The OECD GD 43 and GD 151 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 considered in setting the NOAEL” (OECD 2008; OECD 2013). As the NOAEL can be used as the point of departure for setting safe exposure levels for humans this further supports that effects on AGD are of human relevance. Last, but not least the observations of similar effects in experimental animals and in humans support that effects on AGD in experimental animals are relevant for humans.

**Animal welfare**

40. An important point to remember is that this TG can be enhanced with the ability for detection of EDs can be enhanced without increasing the number of experimental animals used.

41. Assessment of AGD in fetus at caesarean section does not increase the number of experimental animals used, but requires marginally more handling of the fetuses. This assessment can be done very gently just before killing and is therefore not expected to lead to any animal welfare concerns.

**Inclusion of AGD in TG 414**

42. There are standardized OECD test methods for assessing AGD and AGD measured at birth (e.g. PD 1-4) have been included in several OECD TGs. Moreover, several studies (see table 1 and 2) have shown ability to measure AGD at caesarean section at GD 20 or 21 (one day prior to the expected day of delivery). These studies have used a dissecting microscope with a micrometer eyepiece.

43. The current report have shown that the power for assessment of AGD is almost equal in GD 21 fetuses versus in new-borns as the analysis showed similar CoV in these two time points. Therefore this endpoint can be included in TG 414 at GD 21 without any modification of the overall test design. Several studies have also measured AGD with success in rats at GD 19 and 20.

44. At the TC in EG statistical experts from the USEPA noted that it is important to ensure the CoV results analysed are statistically interpretable. The lead have therefore sent detailed data on 9 studies in order for USEPA to calculate the CoVs by using a mixed effects model to distinguish between variability between litters from variability between studies and to better address the sensitivity of the AGD endpoint. These results will be discussed in EG early 2018.

45. AGD is an endpoint of high human relevance and there are no concerns for animal welfare related to the assessment of this endpoint (OECD, 2015).
46. This all supports that assessment of AGD in all fetuses can be included in TG 414 (Appendix 1).
Testosterone levels in male fetuses

Method

47. Changes in testosterone levels in sensitive time windows can cause permanent/long lasting reproductive changes in laboratory rats. Studies have shown that exposure during gestation to e.g. some phthalates, can show effects on testosterone synthesis, fetal testicular content or ex vivo testicular testosterone production in males fetuses leading to effects observed postnatally e.g. anogenital distance and reproductive organ weight changes (Borch et al. 2004, Saillenfait et al. 2013, Borch et al 2006).

48. In general, only few other published papers satisfied the dosing period used in TG 414, either because the administration of the compound was initiated later on (e.g. GD 12), or because the caesarean section was conducted several days before the expected day of delivery, most often at GD 19 (Parks et al., 2000; Fisher et al., 2003; Hotchkiss et al., 2004; Lehmann et al., 2004; Thompson, Ross and Gaido, 2004; Saillenfait et al., 2016). A reason for this deviance could be the interest in the developmental stages in the embryotic phase or in the testosterone peak levels during the foetal development.

49. In relation to the feasibility of inclusion of testosterone hormone measurements the lead and expert group have discussed whether serum testosterone levels, testosterone synthesis (testicular) or ex vivo testosterone production in foetal testes would be the most suitable method.

50. Only measurement of testosterone hormones in the serum would be feasible as other methods might impair the ability to fulfil the purpose of TG 414 that is designed to provide general information concerning the effects of prenatal exposure 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 fetus.

51. Blood sampling in TG 414 (also other hormones) can be done by:

1. Cardiac puncture, can only be done for fetuses selected for skeletal assessment

2. Decapitation, will impact all fetal assessments

3. The umbilical cord, will not impact fetal assessments but only gives a very small amount of blood (so blood of most fetuses/litter should be pooled)
   i. This latter method is not relevant for testosterone as this should be only pooled by sex and the amount of blood might be too small.

52. Fetal measurements of T4, T3 and TSH can be pooled, however testosterone should not be pooled from both sexes.
**Data analysis, sensitivity/power**

53. Weisz and Ward (1980) reported significantly higher serum testosterone levels in male rat fetuses with a peak at GD 18 and this finding was confirmed by Lichtensteiger and Schlumpf (1981, 1985). Due to this testosterone peak in male fetuses at GD18, this stage has been examined when the focus is on sexual dimorphism of sex steroid regulation. As already stated this TG 414 performs caesarean section the day before expected delivery and therefore it is not feasible to assess testosterone level in serum from other time-windows.

54. As stated above (para. 46), plasma testosterone levels have also been found to be affected by chemicals at subsequent fetal and neonatal stages, and have been related to developmental disturbances. This is why this project suggests determining plasma testosterone at around GD 21 in TG414.

55. In many studies intratesticular testosterone levels rather than plasma levels were reported (Parks et al., 2000; Fisher et al., 2003; Hotchkiss et al., 2004; Lehmann et al., 2004; Thompson, Ross and Gaido, 2004). This endpoint is however not feasible to include in TG 414 due to the purpose of the TG 414 (see para 49).

56. In Table 3, the overall mean and standard deviation for the individually measured mean plasma testosterone levels, the standard deviations and the coefficient of variation for each control experiment are presented. Furthermore, the number of control experiments included in the calculations and the range of the numbers of litters in these are stated.

57. It is seen at table 3 that a large overall coefficient of variation of around 43% in controls is obtained indicating that the possibility for detecting effects is low. However, it is also seen that the number of litters included is relatively low compared the number of litters in TG 414, i.e. 3-9 litters per group compared to 20 litters per group in TG 414. Therefore the power for detecting effects is expected to be clearly higher in the TG 414 than in those studies.

58. Moreover, in two of the 7 studies (see figure 3), the coefficient of variation is actually around 20% in spite of the low number of litters per group. This indicates that the power for detecting effect with around 20 litters per group may be sufficient.

59. Nevertheless, as also mentioned in this report several studies on EDs have shown significant effects on testosterone measurement in males fetuses.
Table 3. Overall mean ± SD for the mean level of plasma testosterone at GD 21, the standard deviation and the derived coefficient of variation for each study (7). Additionally, the number of studies/experiments and the range of number of litters in these are included.

<table>
<thead>
<tr>
<th>Control experiments with Wistar rats</th>
<th>Plasma testosterone Mean [nM]</th>
<th>Plasma testosterone Standard deviation [nM]</th>
<th>Plasma testosterone Coefficient of variation [%]</th>
<th>Control experiments (range of litters)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.30 ± 0.10</td>
<td>0.13 ± 0.08</td>
<td>42.99 ± 21.14</td>
<td>7 (3-9)</td>
</tr>
</tbody>
</table>

**Figure 3.** The mean plasma (based on litter means) testosterone level +SD in each control group for the 7 studies included in this feasibility report. These studies included fewer animals’ pr. litter than in the TG 414.

**Human relevance**

60. Testosterone and dihydrotestosterone are two of the key hormone players in sex differentiation of a fetus and are both classified as sex steroid hormones. Therefore by measuring testosterone in TG 414 as an indication of endocrine disruption could indeed be relevant for human risk assessment.
Animal welfare

61. Blood samples for assessment of testosterone or other steroid hormones in fetus are taken at caesarian section by termination of the study. This leads to no concern for animal welfare, as the blood samples will be collected at the time of sacrifice.

Inclusion of Testosterone in TG 414

62. There are standardized test methods for assessing testosterone hormones in serum. The performed power analysis showed coefficient of variation of around 43% in controls indicating that the possibility for detecting effects is low.
63. Nevertheless, several studies on EDs have shown effects on testosterone measurement in male fetuses.
64. The feasibility study demonstrates large variations in sample measurements and the variation in controls for plasma testosterone is high. These data fail to demonstrate that this measurement is sensitive enough to be included in a test guideline.
65. The inclusion of hormone measurement (testosterone and thyroid hormones) in fetuses was discussed in the EG after 1st commenting round. A representative from BIAC described recent research evaluating technical feasibility of collecting the additional endpoints (AGD, blood samples) from fetuses. To preserve the integrity of other endpoints in TG 414, blood samples were collected by cardiac puncture of fetuses selected for skeletal examination. However, this added about 20-25 minutes per litter and fetuses were alive for up to 35 minutes between Caesarean section and necropsy, raising concerns on variation in processing time for animals within a litter which could lead to high variability in data. BIAC noted that in many cases, blood volumes collected from fetuses were limited and needed to be pooled for analyses.
66. The EG recognised that there is large CoV in fetal testosterone due to the gestational surge which occurs before GD 20, and thus this measurement has limited reliability. Overall, the group felt that there is not enough confidence in fetal testosterone to support including the testosterone measure in the revised TG.

Thyroid hormones

Method

67. Thyroid hormones were included in the TGs 421/422 as blood samples from the day 13 pups and assessments of thyroid hormones (T4) in the adult males was required. In contrast,
further assessment of T4 in blood samples from the dams and day 4 pups is to be done if relevant. Moreover the TGs also include an option for other hormones.

68. Due to the circadian rhythm of thyroid hormones, sample collection should occur at approximately the same time of day and be randomized across dosage groups, preferably in the morning hours at which time basal values should be present (Döhler et al., 1979).

69. Blood samples for evaluation of triiodothyronine (T3), thyroxine (T4), and TSH should be collected immediately following sacrifice.

70. Hormonal analyses should be conducted on GD 21 fetuses and dams in this TG 414.

71. If there are inadequate fetuses in a litter to obtain sufficient blood for the hormonal measures, the measurements of T4 and TSH would be a priority, with less emphasis placed on T3 measures (ref. EPA Dev Thyroid Protocol).

72. Prior to sacrifice, every effort should be made to avoid inducing stress that could affect hormone concentrations (Döhler et al., 1979).

73. When the inclusion of thyroid hormones in the dams was discussed in the EG it was suggested that the revised guideline specify that blood should be collected from dams within a short timeframe (i.e. two hours) on the morning of the day of necropsy to reduce variability in thyroid hormone levels. Fasting is not necessary for the blood sampling. It was also noted that only non-pregnant dams should be excluded from analyses, and though all blood samples should be collected, Japan representative recommended that initially high dose groups are compared to controls and if differences are observed, hormone levels in samples from other dose groups can be analysed as well.

Data analysis, sensitivity/power

74. The number of animals included in TG 414 is similar to the number of animals in TG 443, where assessment of thyroid hormones is included. Thus, specific data analysis of power related to number of animals is not needed here.

75. Moreover, the feasibility study from TGs 421/422 (OECD, 2015) included intensive power simulations for TH measurements with fewer animals.

76. However, blood samples will be taken in dams in TG 414 compared to non-pregnant adult animals in TG 443 and TG 421/422 (males). This may affect the sensitivity and therefore needs to be addressed.

Human relevance

77. Thyroid hormones (TH) are needed for proper nerve cell differentiation and proliferation, and normal status of these hormones during early development is therefore crucial. In humans even moderate and transient reductions in maternal T4 levels during pregnancy, may adversely affect the child’s neurological development (OECD, 2015, Thyroid report, 2017).
78. This indicates that by measuring thyroid hormones (T4, T3 and TSH) in dams and fetuses in TG 414 as an indication of thyroid disruption could indeed be relevant for human risk assessment as also described in the feasibility study for TGs 421/422 (OECD, 2015).

Animal welfare

79. Blood samples for assessment of thyroid hormones in fetus could be taken at caesarean section by termination of the study. This leads to no concern for animal welfare, as the blood samples will be collected at the time of sacrifice.
80. Blood samples will be collected from all dams at termination for mandatory assessment of thyroid hormones T4 and T3 or TSH (within a short timeframe (i.e. two hours) on the morning of the day of necropsy).

Inclusion of thyroid hormones in TG 414

81. There are standardized OECD test methods for assessing thyroid hormones. The performed power analysis in TGs 421/422 made in the feasibility study supported that assessment of thyroid hormones could also be included in these TGs.
82. Thyroid measurements (T3, T4 and TSH) will be included as mandatory endpoints in the updated TG 408 (90 days study).
83. Therefore the assessment of Thyroid hormones in TG 414 dams and fetuses is sufficiently sensitive to provide relevant data.
84. Due to the adverse effects seen in humans after developmental hypothyroidism, this endpoint is of high human relevance and there are no concerns for animal welfare related to the assessment of this endpoint as long as blood sampling is done in animals that are being sacrificed anyway.
85. The revised TG 414 will include mandatory measures of T4, T3, and TSH from the dams. As suggested by EG the revised guideline will specify that blood should be collected from dams within a short timeframe (i.e. two hours) on the morning of the day of necropsy to reduce variability in thyroid hormone levels. Fasting is not necessary for the blood sampling.
86. The EG discussed proposed inclusion of optional T4, T3, and TSH measures in male and female fetal blood in the revised TG 414 after 1st commenting round (see para 65). The EG discussed possible criteria to trigger the optional endpoints, but no considerations were agreed upon by the group. Several experts noted that if other indications of potential thyroid impairment were observed (e.g. decrease T4 among dams in TG 414), it may be more helpful to request data from guidelines that include more specific information and adverse responses (e.g. neurodevelopmental endpoints). There was no clear support for including optional endpoints in TG 414 revision without considerations for when these should be included. It was then decided in the EG to not include serum hormone measurements in fetuses.
Abnormalities of external genital organs

Method

87. In TG 414 the reproductive tract is examined for signs of altered development. This project will result in more guidance on evaluation of abnormalities of external genitalia in fetuses such as hypospadias (Hsieh et al. 2007). However, Hsieh et al. (2007) used histopathological examination of the genital tubercle to detect hypospadias while the guidance in TG 414 will be to pay attention to the external reproductive genitals which should be examined for signs of altered development.

88. In the current TG 414 it is mentioned (para. 29) that each fetus should be examined for external alterations. The text has now been modified to take also abnormalities of external genital organs into account.

Data analysis, sensitivity/power

89. The feasibility study for TG 421/422 (OECD, 2015) included a calculation of the effect size needed for finding significant effect for abnormalities/malformations.

90. The data in this feasibility study strongly supported that all male pups in TG 421/422 needs to be evaluated, similarly as in OECD TG 414.

91. This limited sensitivity for detecting significant effects on rare adverse outcomes is generally recognized for malformations. Thus, the occurrence of a few similar rare genital malformations may generally be considered toxicologically relevant although the finding is not statistically significant (OECD, 2015).

Human relevance

92. Hypospadias in humans is one of the most common urogenital congenital anomalies affecting boys (Harris 1990). Prevalence estimates in Europe range from 4 to 24 per 10,000 births, depending on definition (Dolk et al. 2004) with higher rates of about 5% reported in a Danish study (Boisen et al. 2005). Little is known about the aetiology of hypospadias, but a role for EDs has been proposed, and especially the anti-androgenic chemicals (Baskin et al. 2001).

93. Exposure during critical developmental phases such as in utero and in the early postnatal period may lead to adverse effects on both reproductive development and neurodevelopment. The fact that many of the basic mechanisms underlying this developmental process are similar in many known species of mammals indicates that chemicals that have adverse effects on reproductive development in rodents should be considered as potential human reproductive toxicants as well (Gray 1992).
Animal welfare

94. Assessment of abnormalities of external genital organs requires slightly more handling of fetuses. This assessment can be done very gently and is therefore not expected to lead to any animal welfare concerns. However, as the assessment of abnormalities of external genital organs is done after termination of the fetuses in TG 414, there will obviously be no concern for animal welfare.

Inclusion abnormalities of external genital organs in TG 414

95. Assessment of abnormalities is already included in TG 414. However, no details with regard to assessment of abnormalities of external genitals organs are included in the current version. The proposed text to be added in the revised TG 414 in relation to abnormalities of external genitals is observational neither a scoring system nor histopathology and is modified from para 29 in OECD TG 414.
Overall discussion and conclusions

96. The aim of this project was to do a feasibility study for minor enhancements of TG 414 with ED-relevant endpoints. The endpoints considered for inclusion were AGD, Testosterone levels in fetuses, thyroid hormones in the dams and fetuses and guidance for genital abnormalities.

97. For all endpoints, OECD test methods are available for assessing these. Power analyses have been done showing sufficient sensitivity to get relevant data with the number of litters per group in the TGs 421/422 and TG 414 uses 20 litters per group. All four endpoints are of relevance for humans as described in this review. All four of them are mandatory to assess in some OECD Test guidelines used for human risk assessment of chemicals. The overall animal welfare considerations will not increase by the assessments of the 4 endpoints. Inclusion of all four endpoints in TG 414 does not trigger any animal welfare concerns.

98. The EG have decided to not include any serum hormone measurements in fetuses.

99. The Test Guideline has been updated with specific text proposals. No changes in study design and only few text changes are necessary to include the assessment of anogenital distance (AGD), Thyroid hormone measurements (in the dams only) and inclusion of short guidance on abnormalities of external genital organs.

100. In conclusion, it is feasible to make the proposed minor enhancements of TG 414 with ED-relevant endpoints: anogenital distance (AGD), thyroid hormones (in dams) and inclusion of guidance on abnormalities of external genital organs in male fetuses.

101. This report conclude after telephone conference with EG that the proposals for the update of TG 414 are in the 2nd WNT commenting round:

1. AGD mandatory – all fetuses
2. T4, T3 and TSH dams – mandatory
3. Inclusion of guidance on abnormalities of external genital organs in male fetuses.
References (to be updated)


Boisen KA, Chellakooty M, Schmidt IM, Kai CM, Damgaard IN, Suomi AM, Toppari J, Skakkebaek NE, Main KM. 2005. Hypospadias in a cohort of 1072 Danish newborn boys: prevalence and relationship to placental weight, anthropometrical measurements at birth, and reproductive hormone levels at three months of age. The Journal of Clinical Endocrinology & Metabolism 90:4041-4046.


**Priority setting workshop report, 2017**


Thyroid workshop report, 2017


(Wolf et. al. 2005).
Appendix 1 Text changes suggestions for TG 414 shown with track changes

The TG 414 shown with proposed text suggestions (for 2nd WNT commenting) as track changes are on the next pages
OECD GUIDELINE FOR THE TESTING OF CHEMICALS

PROPOSAL FOR UPDATING GUIDELINE 414

Prenatal Developmental Toxicity Study

INTRODUCTION

1. OECD Guidelines for the Testing of Chemicals are periodically reviewed in the light of scientific progress. In Copenhagen in June 1995, an OECD Working Group on Reproduction and Developmental Toxicity discussed the need to update existing OECD Test Guidelines for reproduction and developmental toxicity and the development of new Guidelines for endpoints not yet covered. The Working Group recommended that the Guideline for Developmental Toxicity should be revised, based on a proposal received from the US (1). The Working Group reached agreement on all major elements of the revised version of this Guideline. This update was adopted in 2001.

1a. This test guideline (TG) has been updated with endocrine disruptor relevant endpoints, as a follow up to the high-priority activity initiated at OECD in 1998 to revise existing Test Guidelines and to develop new Test Guidelines for the screening and testing of potential endocrine disruptors (OECD 1998). In this context TG 407 (Repeated Dose 28-Day Oral Toxicity Study in Rodents) was enhanced in 2008 by parameters suitable to detect endocrine activity of test chemicals. In 2015 and 2016 TGs 421 and 422 (Reproduction/Developmental Toxicity Screening Test and Combined Repeated Dose Toxicity Study with the Reproduction/Developmental Toxicity Screening Test) were updated to include some endocrine disruptor relevant endpoints in screening TGs where the exposure periods cover some of the sensitive periods during development (pre- or early postnatal periods) (ref. feasibility report TG 422/421).

1b. The selected additional endocrine disrupter relevant endpoints, were included in TG 414 based on a feasibility study addressing scientific and technical questions related to their inclusion, as well as possible changes in the methods needed for their inclusion (reference to feasibility report TG 414).

1c. The proposed additional endpoints are not intended to be mandatory in TG 414 if they have been or is planned to be measured in TG 443 (FOGRTS).

1d. The objective of this project is not to develop two separate TGs but to include rat specific requirements in the TG and the selected additional changes is only suggested for rats and not for rabbits.

INITIAL CONSIDERATIONS

2. This guideline for developmental toxicity testing is designed to provide general information concerning the effects of prenatal exposure 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. Functional deficits, although an important part of development, are not a part of this Guideline. They may be tested for in a separate study or as an adjunct to this study using the Guideline for developmental neurotoxicity. For information on testing for functional deficiencies and other postnatal effects the Guidelines 443, 416, 421/422 and 426 for the two-generation reproductive toxicity study and the developmental neurotoxicity study should be consulted.

3. This guideline may require specific adaptation in individual cases on the basis of specific knowledge on e.g. physicochemical or toxicological properties of the test substance. Such adaptation is acceptable, when convincing scientific evidence suggests that the adaptation will lead to a more informative test. In such a case, this scientific evidence should be carefully documented in the study report. In conducting the study, the guiding principles and considerations outlined in the OECD Guidance Document n° 19 on the recognition, assessment, and use of clinical signs as humane endpoints for experimental animals used in safety evaluations (5) should be followed.

4. Definitions used are given in the Annex.

PRINCIPLE OF THE TEST

5. Normally, the test substance is administered to pregnant animals at least from implantation to one day prior to the day of scheduled kill/sacrifice, which should be as close as possible to the normal day of delivery without risking loss of data resulting from early delivery. The guideline is not intended to examine solely the period of organogenesis, (e.g. days 5-15 in the rodent, and days 6-18 in the rabbit) but also effects from preimplantation, when appropriate, through the entire period of gestation to the day before caesarean section. Shortly before caesarean section, the females are killed, the uterine contents are examined, and the foetuses are evaluated for soft tissue and skeletal changes.

PREPARATION FOR THE TEST

Selection of animal species

6. It is recommended that testing be performed in the most relevant species, and that laboratory species and strains which are commonly used in prenatal developmental toxicity testing be employed. The preferred rodent species is the rat and the preferred non-rodent species is the rabbit. Justification should be provided if another species is used.

Housing and feeding conditions

7. The temperature in the experimental animal room should be (22 ± 3)°C for rodents and (18 ± 3)°C rabbits. Although the relative humidity should be at least 30% and preferably not exceed 70% other than during room cleaning, the aim should be 50-60%. Lighting should be artificial, the sequence being 12 hours light, 12 hours dark. For feeding, conventional laboratory diets may be used with an unlimited supply of drinking water.

7a. Animals should be group housed in small groups of the same sex; animals may be housed individually if scientifically justified. For group caging, no more than five animals should be housed per cage. Mating
procedures should be carried out in cages suitable for the purpose. Pregnant females should be caged individually and provided with nesting materials. Lactating females will be caged individually with their offspring.

8. Mating procedures should be carried out in cages suitable for the purpose. While individual housing of mated animals is preferred, group housing in small numbers is also acceptable.

**Preparation of the animals**

9. Healthy animals, which have been acclimated to laboratory conditions for at least 5 days and have not been subjected to previous experimental procedures, should be used. The test animals should be characterised as to species, strain, source, sex, weight and/or age. The animals of all test groups should, as nearly as practicable, be of uniform weight and age. Young adult nulliparous female animals should be used at each dose level. The females should be mated with males of the same species and strain, and the mating of siblings should be avoided. For rodents day 0 of gestation is the day on which a vaginal plug and/or sperm are observed; for rabbits day 0 is usually the day of coitus or of artificial insemination, if this technique is used. Mated females should be assigned in an unbiased manner to the control and treatment groups. Cages should be arranged in such a way that possible effects due to cage placement are minimised. Each animal should be assigned a unique identification number. Mated females should be assigned in an unbiased manner to the control and treatment groups, and if the females are mated in batches, the animals in each batch should be evenly distributed across the groups. Similarly, females inseminated by the same male should be evenly distributed across the groups.

**PROCEDURE**

**Number and sex of animals**

10. Each test and control group should contain a sufficient number of females to result in approximately 20 female animals with implantation sites at necropsy. Groups with fewer than 16 animals with implantation sites may be inappropriate. Maternal mortality does not necessarily invalidate the study providing it does not exceed approximately 10 percent.

**Preparation of doses**

11. Using similar doses in the reproductive toxicity studies as in the repeated dose toxicity studies, will allow interpretation of any potential effects on fertility in context with general systemic toxicity. If a vehicle or other additive is used to facilitate dosing, consideration should be given to the following characteristics: effects on the absorption, distribution, metabolism, and retention or excretion of the test substance; effects on the chemical properties of the test substance which may alter its toxic characteristics; and effects on the food or water consumption or the nutritional status of the animals. It is recommended that control animals be dosed with the vehicle at the same dosing regimen as test group animals. The vehicle should neither be developmentally toxic nor have effects on reproduction.
Dosage

12. Normally, the test substance should be administered daily from implantation (e.g., day 5 post mating) to the day prior to scheduled caesarean section. If preliminary studies, when available, do not indicate a high potential for preimplantation loss, treatment may be extended to include the entire period of gestation, from mating to the day prior to scheduled kill. It is well known that inappropriate handling or stress during pregnancy can result in prenatal loss. To guard against foetal loss from factors which are not treatment-related, unnecessary handling of pregnant animals as well as stress from outside factors such as noise should be avoided.

13. At least three dose levels (if possible) and a concurrent control should be included in each study. If dose levels are based on toxicity, the highest dose should be chosen with the aim to induce some systemic toxicity (but not mortality or severe suffering of the animals). If a vehicle is used, the control group should receive the vehicle in the highest volume used for treated groups. Healthy animals should be assigned in an unbiased manner to the control and treatment groups. The dose levels should be spaced to produce a gradation of toxic effects. Unless limited by the physical/chemical nature or biological properties of the test substance, the highest dose should be chosen with the aim to induce some developmental and/or maternal toxicity (clinical signs or a decrease in body weight) but not death or severe suffering. At least one intermediate dose level should produce minimal observable toxic effects. The lowest dose level should not produce any evidence of either maternal or developmental toxicity. A descending sequence of dose levels should be selected with a view to demonstrating any dosage-related response and no-observed-adverse-effect level (NOAEL) or doses near the limit of detection that would allow the determination of a benchmark dose. Two- to four-fold intervals are frequently optimal for setting the descending dose levels, and the addition of a fourth test group is often preferable to using very large intervals (e.g. more than a factor of 10) between dosages. Although establishment of a maternal NOAEL is the goal, studies which do not establish such a level may also be acceptable (2).

14. Dose levels should be selected taking into account any existing toxicity data as well as additional information on metabolism and toxicokinetics of the test substance or related materials. This information will also assist in demonstrating the adequacy of the dosing regimen.

Limit test

16. If a test at one dose level of at least 1000 mg/kg body weight/day by oral administration, using the procedures described for this study, produces no observable toxicity and if an effect would not be expected based upon existing data (e.g., from structurally and/or metabolically related compounds), then a full study using three dose levels may not be considered necessary. Expected human exposure may indicate the need for a higher oral dose level to be used in the limit test. For other types of administration, such as inhalation or dermal application, the physical chemical properties of the test substance often may indicate the maximum attainable level of exposure (for example, dermal application should not cause severe localised toxicity).
Administration of doses

17. The test substance or vehicle is usually administered orally by intubation. If another route of administration is used, the tester should provide justification and reasoning for its selection, and appropriate modifications may be necessary (3)(4)(5). The test substance should be administered at approximately the same time each day.

18. The dose to each animal should normally be based on the most recent individual body weight determination. However, caution should be exercised when adjusting the dose during the last trimester of pregnancy. Existing data should be used for dose selection to prevent excess maternal toxicity. However, if excess toxicity is noted in the treated dams, those animals should be humanely killed / sacrificed. If several pregnant animals show signs of excess toxicity, consideration should be given to terminating that dose group. When the substance is administered by gavage, this should preferably be given as a single dose to the animals using a stomach tube or a suitable intubation canula. The maximum volume of liquid that can be administered at one time depends on the size of the test animal. The volume should not exceed 1 ml/100 g body weight, except in the case of aqueous solutions where 2 ml/100 g body weight may be used. When corn oil is used as a vehicle, the volume should not exceed 0.4 ml/100 g body weight. Variability in test volume should be minimised by adjusting the concentrations to ensure a constant volume across all dose levels.

Observations of the dams

19. Clinical observations should be made and recorded at least once a day, preferably at the same time(s) each day taking into consideration the peak period of anticipated effects after dosing. The condition of the animals should be recorded including mortality, moribundity, pertinent behavioural changes, and all signs of overt toxicity.

Body weight and food consumption.

20. Animals should be weighed on day 0 or no later than day 3 if time-mated animals are supplied by an outside breeder, on the first day of dosing, at least every 3 days during the dosing period and on the day of scheduled kill / sacrifice.

21. Food consumption should be recorded at three-day intervals and should coincide with days of body weight determination.

Post-mortem examination

22. Females should be killed / sacrificed one day prior to the expected day of delivery. Females showing signs of abortion or premature delivery prior to scheduled kill / sacrifice should be killed and subjected to a thorough macroscopic examination.

23. At the time of termination or death during the study, the dam should be examined macroscopically for any structural abnormalities or histopathology assessment should be taken from every species / gender / dose to observe pathological changes. Evaluation of the dams during caesarean section and
subsequent foetal analyses should be conducted preferably without knowledge of treatment group in order to minimise bias.

**Examination of uterine contents**

24. Immediately after termination or as soon as possible after death, the uteri should be removed and the pregnancy status of the animals ascertained. Uteri that appear non-gravid should be further examined (e.g. by ammonium sulphide staining for rodents and Salewski staining or a suitable alternative method for rabbits) to confirm the non-pregnant status (6).

25. Gravid uteri including the cervix should be weighed. Gravid uterine weights should not be obtained from animals found dead during the study.

26. The number of corpora lutea should be determined for pregnant animals.

27. The uterine contents should be examined for numbers of embryonic or foetal deaths and viable foetuses. The degree of resorption should be described (early, late) in order to estimate the relative time of death of the conceptus (see Annex for definitions).

**Examination of foetuses**

28. The sex and body weight of each foetus should be determined. The anogenital distance (AGD) should be measured in all live fetuses.

29. Each foetus should be examined for external alterations including those of the oral cavity (7). Particular attention should be paid to the external reproductive genitals which should be examined for signs of altered development.

30. Foetuses should be examined for skeletal and soft tissue alterations (e.g. variations and malformations or anomalies) (8)(9)(10)(11)(12)(13)(14)(15)(16)(17)(18)(19)(20)(21)(22)(23)(24)(25). Categorisation of foetal alterations is preferable but not required. When categorisation is done, the criteria for defining each category should be clearly stated. Particular attention should be paid to the reproductive tract which should be examined for signs of altered development.

31. For rodents, approximately one-half of each litter should be prepared and examined for skeletal alterations. The remainder should be prepared and examined for soft tissue alterations, using accepted or appropriate serial sectioning methods or careful gross dissection techniques.

32. For non-rodents, e.g. rabbits, all foetuses should be examined for both soft tissue and skeletal alterations. The bodies of these foetuses are evaluated by careful dissection for soft tissue alterations, which may include procedures to further evaluate internal cardiac structure (26). The heads of one-half of the foetuses examined in this manner should be removed and processed for evaluation of soft tissue alterations (including eyes, brain, nasal passages and tongue), using standard serial sectioning methods (27) or an equally sensitive method. The bodies of these foetuses and the remaining intact foetuses should be processed and examined for skeletal alterations, utilising the same methods as described for rodents.

**Clinical Chemistry**
32.a All blood samples are stored under appropriate conditions. Blood samples are taken based on the following schedule:
- From all dams at termination for mandatory assessment of thyroid hormones T4 and T3 or TSH (within a short timeframe (i.e., two hours) on the morning of the day of necropsy).
- Further assessment of other hormones may be measured in serum from dams (optional).
- For quality control it is proposed that historical control data are collected and that for numerical data coefficients of variation are calculated, especially for the parameters linked with endocrine disrupter detection. These data can be used for comparison purposes when actual studies are evaluated.

**DATA AND REPORTING**

**Data**

33. Individual animal data should be provided. Additionally, all data should be summarised in tabular form, showing for each test group the number of animals at the start of the test, the number of animals found dead during the test or killed for humane reasons, the time of any death or humane kill, the number of fertile animals, the number of pregnant females, the number of animals showing signs of toxicity, a description of the signs of toxicity observed, including time of onset, duration, and severity of any toxic effects, the types of histopathological changes, and all relevant litter data. A tabular summary report format that has proven to be very useful for the evaluation of reproductive/developmental effect is given in Annex 2. Data shall be reported individually and summarised in tabular form, showing for each test group and each generation the number of animals at the start of the test, the number of animals found dead during the test or killed for humane reasons, the time of any death or humane kill, the number of pregnant females, the number of animals showing signs of toxicity, a description of the signs of toxicity observed, including time of onset, duration, and severity of any toxic effects, the types of foetal observations, and all relevant litter data.

34. Numerical results should be evaluated by an appropriate statistical method using the litter as the unit for data analysis. A generally accepted statistical method should be used; the statistical methods should be selected as part of the design of the study. Data from animals that do not survive to the scheduled kill/sacrificed should also be reported. These data may be included in group means where relevant. Relevance of the data from such an animal, and therefore inclusion or exclusion from any group mean(s), should be judged on an individual basis.

**Evaluation of Results**

35. The findings of the Prenatal Developmental Toxicity Study should be evaluated in terms of the observed effects. The evaluation will include the following information:

- maternal and foetal test results, including an evaluation of the relationship, or lack thereof, between the exposure of the animals to the test substance and the incidence and severity of all findings;
- criteria used for categorising foetal external, soft tissue, and skeletal alterations if categorisation has been done;
- when appropriate, historical control data to enhance interpretation of study results;
- the numbers used in calculating all percentages or indices;
adequate statistical analysis of the study findings, when appropriate, which should include sufficient information on the method of analysis, so that an independent reviewer/statistician can re-evaluate and reconstruct the analysis.

- a requirement to compare increases of similar malformations which are not statistically significant with the historical control data.

36. In any study which demonstrates an absence of toxic effects, further investigation to establish absorption and bioavailability of the test substance should be considered.

Test report

37. The test report or study records should include the following specific information:

Test chemical substance:

- source, lot number, limit date for use, if available
- stability of the test chemical, if known
- physical nature and, where relevant, physicochemical properties
- identification including CAS number if known/established
- purity

Mono-constituent substance:

- physical appearance, water solubility, and additional relevant physicochemical properties
- chemical identification, such as IUPAC or CAS name, CAS number, SMILES or InChI code, structural formula, purity, chemical identity of impurities as appropriate and practically feasible, etc.

Multi-constituent substance, UVCBs and mixtures:

- characterised as far as possible by chemical identity (see above), quantitative occurrence and relevant physicochemical properties of the constituents.

Vehicle (if appropriate):

- justification for choice of vehicle, if other than water.

Test animals:

- species and strain used;
- number and age of animals;
- source, housing conditions, diet, etc.;
- individual weights of animals at the start of the test.

Test conditions:

- rationale for dose level selection;
- details of test substance formulation/diet preparation, achieved concentration, stability and homogeneity of the preparation;
- details of the administration of the test substance;
- conversion from diet/drinking water test substance concentration (ppm) to the actual dose (mg/kg body weight/day), if applicable;
- environmental conditions;
- details of food and water quality.

Results:

- Maternal toxic response data by dose, including but not limited to:
  - the number of animals at the start of the test, the number of animals surviving, the number pregnant, and the number aborting, number of animals delivering early;
  - day of death during the study or whether animals survived to termination;
  - data from animals that do not survive to the scheduled kill/sacrifice should be reported but not included in the inter-group statistical comparisons;
  - day of observation of each abnormal clinical sign and its subsequent course;
  - body weight, body weight change and gravid uterine weight, including, optionally, body weight change corrected for gravid uterine weight;
  - food consumption and, if measured, water consumption;
  - thyroid hormone thyroid-stimulating hormone (TSH), T4, T3, and Free T4 (if possible) (calibration curves, LOQ, LOD, etc.) and other hormone levels (if measured);
  - necropsy findings, including uterine weight;
  - NOAEL values for maternal and developmental effects should be reported.

Developmental endpoints by dose for litters with implants, including:

- number of corpora lutea;
- number of implantations, number and percent of live and dead foetuses and resorptions;
- number and percent of pre- and post-implantation losses.

Developmental endpoints by dose for litters with live foetuses, including:

- number and percent of live offspring;
- sex ratio;
- foetal body weight, preferably by sex and with sexes combined;
- Anogenital distance of all foetuses (statistically evaluated by sex/gender)
- Thyroid hormone and other hormone levels (if measured)
- external, soft tissue, and skeletal malformations and other relevant alterations;
- criteria for categorisation if appropriate;
- total number and percent of foetuses and litters with any external, soft tissue, or skeletal alteration, as well as the types and incidences of individual anomalies and other relevant alterations (including in external reproductive genitals).
Discussion of results.

Conclusions.

**Interpretation of Results**

38. A prenatal developmental toxicity study will provide information on the effects of repeated oral exposure to a substance during pregnancy. The results of the study should be interpreted in conjunction with the findings of subchronic, reproduction, toxicokinetic and other studies. Since emphasis is placed on both general toxicity and developmental toxicity endpoints, the results of the study will allow for the discrimination between developmental effects occurring in the absence of general toxicity and those which are only expressed at levels that are also toxic to the maternal animal (28).

**LITERATURE** *(Should be updated in final version, may be useful to include some of the relevant references included in the feasibility study to support the additional endocrine-related endpoints.)*


ANNEX

Developmental toxicology: the study of adverse effects on the developing organism that may result from exposure prior to conception, during prenatal development, or postnatally to the time of sexual maturation. The major manifestations of developmental toxicity include 1) death of the organism, 2) structural abnormality, 3) altered growth, and 4) functional deficiency. Developmental toxicology was formerly often referred to as teratology.

Adverse effect: any treatment-related alteration from baseline that diminishes an organism’s ability to survive, reproduce or adapt to the environment. Concerning developmental toxicology, taken in its widest sense it includes any effect which interferes with normal development of the conceptus, both before and after birth.

Altered growth: an alteration in offspring organ or body weight or size.

Alterations (anomalies): structural alterations in development that include both malformations and variations (29):

- **Malformation/Major Abnormality:** Structural change considered detrimental to the animal (may also be lethal) and is usually rare.
- **Variation/Minor Abnormality:** Structural change considered to have little or no detrimental effect on the animal; may be transient and may occur relatively frequently in the control population.

Conceptus: the sum of derivatives of a fertilised ovum at any stage of development from fertilisation until birth including the extra-embryonic membranes as well as the embryo or foetus.

Implantation (nidation): attachment of the blastocyst to the epithelial lining of the uterus, including its penetration through the uterine epithelium, and its embedding in the endometrium.

Embryo: the early or developing stage of any organism, especially the developing product of fertilisation of an egg after the long axis appears and until all major structures are present.

Embryotoxicity: detrimental to the normal structure, development, growth, and/or viability of an embryo.

Foetus: the unborn offspring in the post-embryonic period.

Foetotoxicity: detrimental to the normal structure, development, growth, and/or viability of a foetus.

Abortion: the premature expulsion from the uterus of the products of conception: of the embryo or of a nonviable foetus.

Resorption: a conceptus which, having implanted in the uterus, subsequently died and is being, or has been resorbed:

- **Early resorption:** evidence of implantation without recognisable embryo/foetus.
- **Late resorption:** dead embryo or foetus with external degenerative changes.

NOAEL: abbreviation for no-observed-adverse-effect level and is the highest dose level where no adverse treatment-related findings are observed.