Draft Guidance for describing characterising non-guideline in vitro test methods
Guidance for characterising non-guideline *in vitro* test methods to facilitate their consideration in regulatory applications

(Version 97.0/JRC)

INTRODUCTION

Purpose

Regulators are increasingly faced with the need to include data from non-guideline studies, especially from novel in vitro assays, *in particular for*, in priority setting of substances that require testing but also in the context of their safety assessments, or other regulatory activities. The rapid timelines for development of such new assays and the lack of transferability of certain technologies do not make these assays easily amenable to Test Guideline development and use. This guidance outlines the information that ideally should be provided by developers to describe non-guideline *in vitro* methods in order to facilitate an assessment of the quality of data produced and the potential utility in regulatory applications. It has been developed in the context of the work of the OECD Advisory Group on Molecular Screening and Toxicogenomics, and its programme on the development of Adverse Outcome Pathways (AOPs).

The purpose of this guidance is to harmonise the way non-guideline *in vitro* methods are described and thereby facilitate an assessment of the quality of data produced, irrespective of whether these tests are based on manual protocols or assay protocols adapted for use on automated platforms or high-throughput screening systems (HTS). This guidance outlines the elements considered relevant for providing a comprehensive description of an *in vitro* method to facilitate the interpretation of results and support scientifically defensible fit-for-purpose applications, so as to be suitable for their use in decision making processes by regulators in particular or by the scientific community and end users in general.

This guidance is *not* intended to be prescriptive nor does it endorse a particular structure for reporting the information. On the contrary, the structure should be both flexible and transparent and the example given below merely illustrates the type of information that could be helpful to characterise a particular method. The completeness of the information provided will vary depending on the level of development of the method under consideration and this in turn will impact its use in different regulatory applications (see Section 5). Note the structure here outlines the information about a given method, not how the results generated from such methods should be structured and recorded. Work is
additionally underway to develop an OECD Harmonised Template for capturing results from *in vitro* assays, so named OHT 201.

**Background**

In the past, the development of alternative methods and in particular *in vitro* alternative methods have been largely focused on modelling *downstream health effects* observable *in situ* (so-called “apical endpoints”), be it by one method only (=stand-alone alternative) or by a group of assays used either in conjunction (“test battery”) or in a strategic sequence (“testing strategy”). Traditional *in vitro* methods have been largely focused towards the development of alternatives or replacements to animal testing for apical endpoints, be it by one method only (=stand-alone alternative) or by a group of assays used either in conjunction (“test battery”) or in a strategic sequence (“testing strategy”). Examples include methods to evaluate skin corrosion and irritation potential (stand-alone test methods) or alternatives for ocular irritation assessment (used within strategic approaches), which have resulted in specific OECD Test Guidelines. The validation process that such methods have followed is described in more detail in OECD Guidance Document 34 (guidance document on the validation and international acceptance of new or updated test methods for hazard assessment), although the principles and scientific elements of validation are by no means intrinsically linked to the test outcome/predictions of assays.

However since the publication of the NRC Report in 2007, initiatives such as Tox21 and the US EPA's Toxcast have sought to exploit emerging technologies such as high throughput screening (HTS) and high content screening (HCS) assays to generate data. Such assays could be termed “next generation” *in vitro* assays. There are currently a number of challenges in using HTS and HCS data in regulatory activities. A major challenge is the lack of development of standard guidelines based on the traditional validation paradigm, which might not be appropriate or relevant for such methods, not least since the assays might not be easily transferable to third laboratories due to very specific equipment (i.e. robotic platforms). In such cases assessment of transferability and between-laboratory as foreseen in regular validation (GD 34) would be less relevant. Indeed, the evolution of some of these assays might be too rapid to justify a full and lengthy formal validation process in view of the standardization and acceptance as OECD Test Guidelines. However, data generated by these methods, as well as by many other *in vitro* methods that may be some methods may be early in their development stage or rapidly developing, not measuring an established regulatory endpoint, or of narrow applicability for a given hazard/biological endpoint, can nevertheless be of practical value, e.g. to facilitate a better understanding of the mechanisms of action of a substance and thus could be used to a greater extent in a regulatory context. Indeed the rapid development of new technologies, and their use in generating large amounts of data, has made the lengthy formal validation, the standardization and acceptance as OECD Test Guidelines extremely cumbersome. Additionally, many other *in vitro* methods may be in early development stages, may not measure an established regulatory endpoint, or may be of narrow applicability thus making them less attractive for Test Guidelines development.
The availability of novel technologies and new in vitro methods at early stages of development, including these “next-generation” in vitro methods, raise questions - how can these methods be practically utilised whilst ensuring that the resulting data are scientifically robust and interpreted in the appropriate context? It is critical to have some type of framework in place to help evaluating the uncertainties and interpreting such in vitro methods to assure scientific confidence. Anchoring such methods in the appropriate biological context (e.g. Adverse Outcome Pathway (AOP)) supports the examination of the scientific uncertainties and should in turn help to inform their use in integrated approaches to testing and assessment (IATA) as well as chemical grouping and prioritisation. These types of regulatory applications are being described in more detail in another guidance that is currently undergoing development (ref to Guidance on IATA for skin sensitization, in preparation).

This guidance provides a framework for describing existing knowledge about a method, termed Method Description, and abbreviated to MD. The MD structure below outlines the type and level of information that would be ideally captured. It is based on principles outlined in the IOM Framework (2010) [see http://nap.edu/catalog.php?record_id=12869] and uses the same information elements for in vitro methods indicated in OECD Guidance 34 and in an existing Streamlined Summary Document (OECD Series on Testing and Assessment Nr. 180; Streamlined Summary Document Related to the Fluorescein Leakage (FL) Test Method for Identification of Ocular Corrosives and Severe Irritants (TG 460)). It also mirrors many of the same information elements that form the basis of the Joint Research Centre’s (JRC) Data Base service on ALternative Methods (DB-ALM) [see http://ecvam-dbalm.jrc.ec.europa.eu/] as well as characteristics defined by the OECD Thyroid Scoping Group (under the OECD Validation Management Group for Non-Animal Testing (VMG_NA)). This guidance is not intended to duplicate OECD Guidance Document 34 (GD34), rather it is complementary.

Ultimately any completed MD should be stored in a publicly available repository/library of methods. This is beyond the scope of this guidance, although existing repositories such as JRC’s DB-ALM may be helpful.

In the MD example below, some of the information elements may not be relevant for a given assay, or are not known, or simply cannot be provided. In such cases, annotating with “no information available” will still be helpful for the end user to understand when evaluating the overall assay. It is also worth noting that some elements may be redundant or may be more readily populated based on the way in which an electronic repository has been structured e.g. particular fields in Section 1 are of note.
1. General information

1.1 Assay Name (title): Provide a short and descriptive title for the assay.

1.2 Summary: Provide a summary of the assay features. E.g. “the effects of a chemical irritant on the opacity and permeability of a freshly isolated bovine cornea can be used to measure eye irritancy.”

1.3 Date of MD: Report the date of the first version of the MD (day/month/year). Example: “5 August 2013”.

1.4 MD author(s) and contact details: Indicate the name(s) and the contact details of the author(s) of the MD (first version of the MD).

1.5 Date of MD update(s) and contacts:
- Date: Indicate the date (day/month/year) of any update of the MD. The MD can be updated for a number of reasons such as additions of new information (e.g. addition of new validation studies in section 7) or corrections to the existing information. Summarise briefly which information has been updated.
- Contact: Indicate the name and the contact details of the author(s) of the updates if different from that in 1.4.

1.6 Assay developer(s)/Laboratory and contact details: Indicate the name of the assay developer(s)/author(s)/laboratory(ies), and the corresponding contact details.

1.7 Date of assay development and/or publication: Report the year of initial release/publication of the assay described in the current MD.

1.8 Reference(s) to main scientific papers: List the main bibliographic references to original paper(s) explaining the assay development. Any other reference such as references to validation datasets or prediction model development can be reported in field 6.0 “Bibliography”.

1.9 Availability of information about the assay in relation to proprietary elements: Indicate whether the assay is proprietary or non-proprietary (to what extent is the assay method transferable or contains proprietary elements) and specify (if possible) what kind of information about the assay cannot be disclosed or is not available (e.g., chemical reference sets (training or test sets), prediction model). Example: “The assay is non-proprietary but the training and test sets are not available”; “The model is proprietary and the prediction model and associated data sets are confidential”; “Company XYZ’s in vitro model for sensitisation may be transferable but the prediction model is confidential”. Obviously ensuring scientific confidence in an assay is facilitated if all information is transparent and can be disclosed.
1.10 Information about the throughput of the assay: Indicate the throughput of the assay to provide an indication of likely resource intensity e.g. low (manual assay, one chemical tested at a time), low-moderate, moderate, moderate-high, high throughput (e.g. in 96/384 well-plate and higher), and justify with experience gained so far. Qualify with e.g. approximate number of chemicals/concentrations per run. If appropriate indicate whether a manual assay could be run in a higher throughput mode.

1.11 Status of method development and uses: Compile information for the following sections if appropriate. Considerations could include:

i) Development status: Indicate if the assay is still under development, and the estimated timeline for completion as far as possible

ii) Known uses: Summarise the current and/or past use of the assay by different laboratories

iii) Evaluation study: Summarise the main conclusions or refer to individual protocol if available

iv) Validation study: Indicate participation in a formal validation study/studies and summarise the conclusions and their outcomes or refer to the individual protocol if available

v) Regulatory use: Provide details of any regulatory application

1.12 Abbreviation and Definitions

If any terms were abbreviated or need specific definitions, they should be provided in this section.

2. Test Method Definition

2.1 Purpose of the test method

The claimed purpose and rationale for intended use of the method (e.g. alternative to an existing method, screening, provision of novel information in regulatory decision-making, mechanistic information, adjunct test, replacement, etc.) should be explicitly described and documented. If the biological endpoint refers to a key event or molecular initiating event (MIE), provide a short description indicating what key event within an existing or developing AOP, or in relation to a mode of action, the assay is aiming to characterise. This will be useful to understand where the assay might fit in the context of an existing regulatory endpoint. E.g. the GSH assay measures reactivity which characterises covalent binding, which is the MIE within the AOP for skin sensitisation; the relative binding affinity as measured in an estrogen receptor (ER) binding assay characterises the molecular initiating event (MIE) in an AOP for ER-mediated reproductive impairment. For definitions of an AOP and guidance in developing an AOP, see the Guidance Document No. 184, Guidance Document on Developing And Assessing Adverse Outcome Pathways.

In the absence of any AOP, provide an indication of the plausible linkage between the mechanism(s) the assay is measuring and the overall endpoint. More specific information can be captured in section 2.2 (Scientific principle of the method), e.g. “This assay addresses peptide reactivity as the MIE within the AOP for skin sensitisation by covalent binding by measuring depletion of synthetic heptapeptides containing either cysteine or lysine following a 24 hour incubation period.”

Comment [GA4]: Question from the US: “an AOP refers to a published AOP, one from a list of pre-approved AOPs, or one that is being proposed, in which case extensive detail would need also to be provided around that proposed AOP.”

Comment [JRAS]: This extra information should be handled as an attachment to the document and only a summary provided in the MD. See also section 3.
2.2 Scientific principle of the method: Provide the scientific rationale, supported by and bibliographic scientific references to articles, for the development of the assay. A summary description of the scientific principle including the biological/physiological basis and relevance (e.g. modeling of a specific organ) and/or mechanistic basis (e.g. modeling a particular mechanism by biochemical parameters) should be described. If possible, indicate what the anchor point is within an AOP.

2.3 Tissue, cells or extracts utilised in the assay and the species source: Indicate the experimental system for the endpoint being measured. e.g. “freshly isolated bovine cornea”; “Porcine TOP”; “Bovine lactoperoxidase”; “hER from extracts of human breast cancer cells”; “mERs from mouse uterine membranes”. Provide information on whether materials are readily available commercially or whether materials are developed in the laboratory (e.g. cell suspensions from tissue).

2.4 Description of the experimental system exposure regime: Provide a summary description of the essential information pertaining to the exposure regime (dosage and exposure time including observation frequency) of the test compounds to the experimental system including information on metabolic competence if appropriate; number of doses/concentrations tested or testing range, number of replicates. Also, describe any specialized equipment needed to perform the assay and measure the endpoint.

2.5 Endpoint and Endpoint Measurement: Endpoint here makes reference to any biological effect, process or response that can be measured. Specify precisely and describe the endpoint and its measurement, e.g. corneal opacity measured using an opacimeter; half maximal activity concentration (AC50) derived from a competitive binding assay in human estrogen receptor assay or from the up-regulation of the proinflammatory antiangiogenic chemokine CXCL10.

Specify the precise endpoint assay investigated as applicable e.g. “IC50”; “half maximal activity concentration - AC50”, “in vitro irritation score (IVIS) = mean opacity value + (15*mean permeability OD490 value) is used to evaluate the irritation potential of a test material”, and how it is calculated (e.g. IC50 using modified Hill equation).

2.6 Data analysis/interpretation: Comment on the endpoint value in terms of a boundary or range to provide a context for interpretation. E.g. putting into context what a negative value or >100% value might represent in a binding inhibition assay.

2.7 Quality / Acceptance criteria: As appropriate, provide information on the availability of acceptance criteria and quality assurance as it pertains to the:
- Experimental data (storage/archiving)
- Experimental system(s) used
- Equipment used
- Availability of internal standards (e.g. positive and negative controls, reference chemicals, performance benchmarks)
- Criteria to accept or reject experimental data
2.8. Known technical limitations and strengths: Specify any known technical limitations or strengths in running the assay. E.g. “Obtaining thyroid glands as a source of TPO may be an obstacle to the use of this assay in other laboratories” (see section 4.4.6 for the domain of applicability of the assay).

2.9 Other related assays that characterise the same event as in 2.1: Identify any related assays if known that may characterise the same key event as described in 2.1 e.g. GSH assay that measures reactivity and provide an indication of whether a MD has been prepared if possible. Make references to it, as appropriate.

3. Data interpretation and prediction model

3.1 Assay endpoints captured in the prediction model: Identify the endpoint(s) from the given assay(s) that form(s) the basis of the prediction model. Note the endpoint could be the same as specified in 2.2; if so, state “see 2.2”.

3.2 Explicit prediction model: Describe the prediction model. If the prediction model is too long or complex and cannot be practically reported here, include in this field a reference to a paper or a document where the prediction model is already described. This material can also be attached as supporting information in field 7.0.

3.3 Software name and version for algorithm/prediction model generation: Specify the name and the version of software applications used to derive the prediction model or to undertake the statistical processing.

4. Test Method Performance
4.1 Method robustness data

This section should provide information on the robustness of the method and how it performs. If possible and as appropriate, it should include the rationale for the selection of reference substances and a discussion on any unexpected outcome. Information on:

- within-laboratory repeatability and reproducibility,
- between laboratory transferability and reproducibility,

should be reported if known, or any other information relevant for the evaluation of the method depending on the method’s purpose.

4.2 Reference chemicals, rationale for their selection and other available information: Indicate whether the results for the reference chemicals (i.e. the “training set” chemicals used in the development and evaluation of the assay and its associated prediction model) are free and publicly available in some form (e.g., published in a paper or a regulatory document, stored in a database) or appended to the current MD as supporting information (see field 7.0). If it is not available, explain why: e.g. “They are available and attached”; “They are available but not attached”; “They are not available because the data set is proprietary”.

If the training set is being appended as supporting information (see field 7.0), indicate what information exists for the reference chemicals including the rationale for their selection if available; e.g. a) Endpoint values; b) Chemical names (common names and/or IUPAC names); b) CAS numbers; c) SMILES; d) MOL files; e) Structural formula; f) Availability from commercial sources.

4.3 Performance measures/predictive capacity (if known): Report any goodness-of-fit statistics that might be available for the assay and/or its prediction (e.g. \( r^2 \), \( r^2 \) adjusted, standard error, sensitivity, specificity, false negative and false positive rates, predictive values, etc). There are multiple appropriate ways to analyse the large number of concentration curves coming from HT screening assays, and there is no single “right way.” Therefore, in addition to specifying the fit, it is equally important to explain the curve fitting process, the assumptions used to determine the goodness-of-fit and any limitations related to the data analysis.

4.4 Scope and limitations of the assay, if known: Describe the types of substances in terms of their physical properties or similar (e.g. specific types of substances) for which the assay is appropriate. If possible characterise the applicability domain systematically using physicochemical descriptors using the training sets of reference chemicals that were used during the development or validation of the assay itself. Descriptions should include physical-chemical limitations and discussion of the chemical space that has been covered in the development of the assay(s) and use of assay results in the prediction model. Is the test

Comment [JRA10]: Refers to comment 1 (DK) on RCOM doc
amenable to a variety of chemicals such as mixtures, UVCBs, organometallics, inorganic chemicals, discrete organic chemicals?

Describe for example the inclusion and/or exclusion rules that may also define the scope of the assay, e.g. substances with high volatility or lipophilicity. Limits of applicability could also include established limitations of the assay, e.g. for the Bovine Corneal Opacity and Permeability assay, high false positive rates are known for alcohols and ketones and high negative rates have been established for solids; e.g. UV absorbance of some compounds or highly coloured substances or autofluorescence could interfere with spectrophotometric analysis.

5. Potential Regulatory applications:

This section provides information to help build a contextual weight of evidence analysis on the (quantitative) use of the prediction model for different regulatory purposes indicating all its potential applications.

5.1 Context of use: Propose possible conditions of use referring to effects investigated and the AOP if possible. This section is aimed at describing the level of scientific confidence for different end use scenarios i.e. where there is adequate scientific confidence for the use of a given prediction model and the rationale for this. Examples of end use scenarios could include:

- Initial chemical grouping: The outcomes from the assay could be used to substantiate a hypothesis for grouping substances together for the purposes of read-across;
- Priority setting: The assay might help prioritise substances within an inventory for more detailed evaluation;
- Screening level assessment of a biomarker or mechanistic endpoint: The assay might provide a reasonable indicator of the likely outcome in an in vivo endpoint assay. The screening level assessment may even be sufficient to identify a hazard, inform classification & labelling or even potency;
- Integrated approaches to testing and assessment (IATA): The assay may form one component of an IATA;

6. Bibliography: Report useful references other than those directly associated with the assay or prediction development (references describing the model development are reported in field 1.8).

7. Supporting information: Indicate whether supporting information is attached (e.g. external documents) to this MD and specify its content and possibly its utility unless previously annotated in the preceding sections.
GLOSSARY

MD: Method description

DB-ALM: Joint Research Centre's (JRC) DataBase service on ALternative Methods to animal experimentation Database for alternative methods (Last access: 21.02.2014 http://ecvam-dbalm.jrc.ec.europa.eu/)

ER: estrogen receptor

HTS: High Throughput Screening

HCS: High Content Screening

IATA: Integrated Approaches to Testing and Assessment

IOM: Institute of Medicine

OHT201: OECD Harmonised Template 201 (harmonized template for key events)

MIE: Molecular Initiating Event

Field Code Changed