GUIDANCE

ADOPTED (ECHA): 5 June 2018
ADOPTED (EFSA): 5 June 2018

doi: 10.2903/j.efsa.2018.5311

Guidance for the identification of endocrine disruptors in the context of Regulations (EU) No 528/2012 and (EC) No 1107/2009

European Chemical Agency (ECHA) and European Food Safety Authority (EFSA) with the technical support of the Joint Research Centre (JRC)

Niklas Andersson, Maria Arena, Domenica Auteri, Stefania Barmaz, Elise Grignard, Aude Kienzler, Peter Lepper, Alfonso Maria Lostia, Sharon Munn, Juan Manuel Parra Morte, Francesca Pellizzato, Jose Tarazona, Andrea Terron and Sander Van der Linden

Abstract

This Guidance describes how to perform hazard identification for endocrine-disrupting properties by following the scientific criteria which are outlined in Commission Delegated Regulation (EU) 2017/2100 and Commission Regulation (EU) 2018/605 for biocidal products and plant protection products, respectively.

Keywords: biocidal product, plant protection product, endocrine disruptor, guidance, hazard identification

Requestor: European Commission

Question numbers: EFSA-Q-2016-00825, ECHA-18-G-01-EN

Correspondence: For biological products: biocides@echa.europa.eu
For plant protection products: pesticides.peerreview@efsa.europa.eu

Acknowledgements: The ECHA and EFSA wish to thank the ECHA staff members Evelin Fabian, George Fotakis and Pascal Phrakonkham; the EFSA staff members Daniele Court Marques, George Kass, Reinhilde Schoonjans, Rositsa Serafimova and Franz Streissl; the JRC staff members David Asturiol Bofill and Andrea Richarz; the Directorate-General for Health and Food Safety (DG SANTE) staff members Laura Fabrizi, Martinus Nagtzaam, Karin Nienstedt; the German Federal Institute for risk assessment (BfR) staff members Bettina Dabrowski De Flores, Dimitra Eleftheriadou, Olena Kucheryavenko, Anja Lehmann, Glenn Lurman, Martina Rauch and Alessia Verani for the support provided to the drafting of this Guidance.

Suggested citation: ECHA (European Chemicals Agency) and EFSA (European Food Safety Authority) with the technical support of the Joint Research Centre (JRC), Andersson N, Arena M, Auteri D, Barmaz S, Grignard E, Kienzler A, Lepper P, Lostia AM, Munn S, Parra Morte JM, Pellizzato F, Tarazona J, Terron A and Van der Linden S, 2018. Guidance for the identification of endocrine disruptors in the context of Regulations (EU) No 528/2012 and (EC) No 1107/2009. EFSA Journal 2018;16(6):5311, 135 pp. https://doi.org/10.2903/j.efsa.2018.5311. ECHA-18-G-01-EN.

ISSN: 1831-4732

efsa

The EFSA Journal is a publication of the European Food Safety Authority, an agency of the European Union.

Table of contents

1. Introduction

The European Commission asked the European Food Safety Authority (EFSA) and the European Chemicals Agency (ECHA) to develop a guidance document for the implementation of the scientific criteria for the determination of endocrine-disrupting properties pursuant to the Biocidal Products Regulation (EU) No 528/2012[1] and the Plant Protection Products Regulation (EC) No 1107/2009[2].

This guidance document is written to provide guidance to applicants and assessors of competent regulatory authorities on how to identify endocrine disruptors in accordance with the ED criteria laid down in Commission Delegated Regulation (EU) No 2017/2100[3] and Commission Regulation (EU) No 2018/605[4] for biocidal products (BP) and plant protection products (PPP), respectively. The guidance document describes how to gather, evaluate and consider all relevant information for the assessment, conduct a mode of action (MoA) analysis, and apply a weight of evidence (WoE) approach, in order to establish whether the ED criteria are fulfilled.

The guidance document has been subject to several consultations which are summarised in a technical report (ECHA and EFSA, 2018). It was consulted twice with an ad-hoc Consultation Group of Member States experts and stakeholders (April–May 2017 and July–August 2017). A general public consultation was held (December 2017–January 2018) to which any interested party could respond. A targeted consultation of risk assessors from competent authorities in the plant protection and biocidal product sectors were consulted in parallel with the EFSA Scientific Committee and the EFSA Panel on plant protection products and their residues (April 2018). Finally, risk managers of the competent authorities for biocidal products and of those for plant protection products were consulted, before adoption of the guidance by ECHA and EFSA according to their procedure (May 2018). When revising the guidance following the above consultations, the overall feedback received and the status of scientific knowledge was considered and it was acknowledged that the document, in future, may need to be revised, when relevant further scientific knowledge becomes available and on the basis of the experience acquired with the application of the present guidance document.

Section 3 presents the assessment strategy for determining whether a substance meets the ED criteria. The strategy is based on the requirements outlined in the ED criteria.[3,4] An approach is proposed for analysing the information provided in a dossier submitted for approval of a substance in the context of the PPP[2] or BP[1] Regulations.

Section 4 gives an overview on the information sources that may provide suitable information for ED identification and therefore should be considered for the assessment. In addition, Section 4 provides guidance on how to consider the scientific data generated in accordance with internationally agreed study protocols in order to facilitate the evaluation of both adverse effects and endocrine activity (by following the process explained in Section 3). The rationale for grouping effects is based on the 'Guidance Document on standardised test guidelines for evaluating chemicals for endocrine disruption. Series on Testing and Assessment No. 150' provided by the Organisation for Economic Co-operation and Development (OECD, 2018b) for their interpretation with regard to estrogenic, androgenic, thyroidal and steroidogenic (EATS) modalities and adapting the Joint Research Centre's (JRC) screening methodology to identify potential endocrine disruptors (JRC, 2016).

Section 5 gives recommendations for applicants and assessors from evaluating authorities and for future research. The guidance is complemented with a list of references, abbreviations and a glossary of terms and definitions used in the text, and several appendices providing information on some specific scientific or technical issues (Appendix A – Additional considerations on how to assess the potential for thyroid disruption; Appendix B – Recommendations for design, conduct and technical evaluation of hormonal studies; Appendix C – Information requirements under the BP and PPP Regulations; Appendix D – Databases, software tools and literature-derived (Q)SARs; Appendix E – Excel template for

[1] Regulation (EU) No 528/2012 of the European Parliament and of the Council of 22 May 2012 concerning the making available on the market and use of biocidal products Text with EEA relevance. OJ L 167, 27.6.2012, p. 1–123. Available online: http://data.europa.eu/eli/reg/2012/528/oj

[2] Regulation (EC) No 1107/2009 of the European Parliament and of the Council of 21 October 2009 concerning the placing of plant protection products on the market and repealing Council Directives 79/117/EEC and 91/414/EEC. OJ L 309, 24.11.2009, p. 1–50. Available online: http://data.europa.eu/eli/reg/2009/1107/oj

[3] Commission Delegated Regulation (EU) 2017/2100 of 4 September 2017 setting out scientific criteria for the determination of endocrine-disrupting properties pursuant to Regulation (EU) No 528/2012 of the European Parliament and Council. OJ L 301, 17.11.2017, p. 1–5. Available online: http://data.europa.eu/eli/reg_del/2017/2100/oj

[4] Commission Regulation (EU) 2018/605 of 19 April 2018 setting out scientific criteria for the determination of endocrine-disrupting and amending Annex II to Regulation (EC) 1107/2009. OJ L 101, 20.4.2018, p. 33–36. Available online: http://data.europa.eu/eli/reg/2018/605/oj

reporting the available information relevant for ED assessment; Appendix F – Example on how to develop the search strategy protocol; and Appendix G – Example of MoA for non-target organisms (fish)).

2. Scope of the guidance document

This document is intended to provide guidance to applicants and assessors of the competent regulatory authorities on the implementation of the scientific criteria for the determination of endocrine-disrupting properties pursuant to Regulations (EU) No 528/2012[1] and (EC) No 1107/2009[2], as defined in Commission Delegated Regulation (EU) No 2017/2100[3] and Commission Regulation (EU) No 2018/605[4], respectively.

Like the criteria to identify endocrine disruptors, this guidance document is based on the WHO/IPCS definition of an endocrine disruptor (WHO/IPCS, 2002). It should be noted that the guidance given in this document is limited to the steps necessary to identify a substance as an endocrine disruptor. The document does not provide guidance on how to further characterise the hazard potential of a substance or the risk to humans or non-target organisms. The latter information may be needed to follow-up the regulatory consequences laid out in Regulations (EU) No 528/2012[1] and (EC) No 1107/2009[2].

The term 'substance' as used in this guidance refers in scientific terms to any 'chemical substance'. However, to which groups of 'substances' the ED assessment and the ED criteria are regulatory applicable is given in Regulations (EU) No 528/2012[1] and (EC) No 1107/2009[2], besides Commission Delegated Regulation (EU) No 2017/2100[3] and Commission Regulation (EU) No 2018/605.[4]

Although the ED criteria cover all endocrine-disrupting MoAs, i.e. adverse effects which may be caused by any endocrine modality, this guidance document mainly addresses the effects caused by EATS modalities. This is because the EATS modalities are currently the pathways for which there is a relatively good mechanistic understanding of how substance-induced perturbations may lead to adverse effects via an endocrine-disrupting MoA. In addition, only for the EATS modalities there are at present standardised test guidelines for *in vivo* and *in vitro* testing available where there is broad scientific agreement on the interpretation of the effects observed on the investigated parameters. These test guidelines are compiled in the OECD Guidance Document on Standardised Test Guidelines for Evaluating Chemicals for Endocrine Disruption (OECD GD 150) (OECD, 2018b), which includes the 'OECD Conceptual Framework (OECD CF) for Testing and Assessment of Endocrine Disrupters' providing a grouping of the studies into five levels according to the kind of information provided. OECD GD 150 including the OECD CF was updated in parallel to the preparation of this guidance, the references made in this document to the OECD GD 150 are based on the document which was adopted by OECD in April 2018 (OECD, 2018b). This guidance is focused on EATS modalities for which there is currently the most knowledge available. However, the general principles outlined in the assessment strategy (Section 3) are also applicable to other endocrine (non-EATS) modalities. Although the existing knowledge for those modalities is not as advanced as for the EATS modalities, it may, in some cases, be already possible to reach a conclusion on a non-EATS endocrine modality, e.g. where literature data provide mechanistic information, which can be linked to adverse effects measured in standard tests, e.g. histopathological findings in the pancreas.

With respect to species addressed, the focus of this guidance is on vertebrate organisms, for which the current understanding of the endocrine system and availability of test methods is most advanced, i.e. mammals, fish, and amphibians.

Due to the scarce knowledge on the endocrinology for non-target invertebrates, this guidance does not specifically cover those organisms and therefore the generation of specific data will not be triggered by applying the strategy developed in this guidance. However, if available, information on invertebrate non-target organisms (e.g. endocrine mechanistic and/or adverse effect data) should be considered in the assessment applying the general principles of this guidance.

3. Strategy to assess whether a substance meets the endocrine disruptor criteria

This chapter outlines the strategy for determining whether a substance has ED properties in accordance with the ED criteria applicable for the PPP[2] and BP[1] Regulations. Before providing an overview of the ED assessment strategy, the definition of an endocrine disruptor and the requirements for determining whether a substance meets this definition specified in the ED criteria are discussed.

The criteria for the determination of the ED properties for humans are presented separately from those applicable to non-target organisms; both sets of criteria are further sub-divided into two

sections; one section on the definition of an ED and one section on the information to be considered for the determination of the ED properties.

The first section defines when a substance shall be considered as having ED properties. This section is identical for both sets of criteria.

According to the ED criteria,[3,4] a substance shall be considered as having ED properties if it meets all of the following criteria:

a) *it shows an adverse effect in [an intact organism or its progeny]/[non-target organisms], which is a change in the morphology, physiology, growth, development, reproduction or life span of an organism, system or (sub)population*[5] *that results in an impairment of functional capacity, an impairment of the capacity to compensate for additional stress or an increase in susceptibility to other influences*;
b) *it has an endocrine mode of action, i.e. it alters the function(s) of the endocrine system*;
c) *the adverse effect is a consequence of the endocrine mode of action*.

It should be highlighted that the 'endocrine mode of action' as stated in point (b) should be interpreted as 'endocrine activity' while the term 'endocrine mode of action' in point (c) covers the link between the adverse effect and the endocrine activity identified in points a) and b), respectively.

Keeping this in mind point (b) above should be understood as (differences from above in *italics*):

it shows *endocrine activity*, i.e. it *has* the potential to alter the function(s) of the endocrine system;

Consequently point (c) above should be understood as (differences from above in *italics*):

the substance has an endocrine disrupting mode of action, i.e. there is a biologically plausible link between the adverse effect and the endocrine activity.

Since conclusions as to whether the ED criteria are met need to be drawn separately for humans and non-target organisms, the hazard identification strategy starts with two *a priori* problem formulations:

- Is there a biologically plausible link between endocrine activity and observed adverse effect(s) that are relevant for humans?
- Is there a biologically plausible link between endocrine activity and observed adverse effect(s) that are relevant for non-target organisms at population level?

Both problem formulations above must be answered and, as required by Regulation (EC) No 1107/2009[2] and Regulation (EU) No 528/2012[1], conclusions be drawn with respect to both humans and non-target organisms (see Section 3.1).

A conclusion on whether the ED criteria are met should always be drawn with respect to both humans and non-target organisms.

The information needed to assess ED properties for humans and non-target organisms may overlap. Mammalian data are always relevant for ED assessment on non-target organisms. Furthermore, there may be information on non-target organisms that could be relevant also for the ED assessment for humans.

The second section in the criteria specifies for both humans and non-target organisms what information shall be considered when determining ED properties, and how this information is to be assessed.

- According to the ED criteria, *all available relevant scientific data* must be considered in the assessment (for further details on how to gather this information see Section 3.2); and
- The ED criteria state that a weight of evidence approach shall be applied for the assessment of the available scientific data.

[5] The term (sub)population is of predominant relevance with respect to humans, therefore for non-target organisms the term population is used throughout the document.

With regard to WoE, a reference is given to the approach provided in Regulation (EC) No 1272/2008[6] on classification, labelling and packaging of substances and mixtures (CLP Regulation). According to Annex I, Section 1.1.1. of the CLP Regulation *'weight of evidence determination means that all available information bearing on the determination of hazard is considered together, such as the results of suitable in vitro tests, relevant animal data, information from the application of the category approach (grouping, read-across), (Q)SAR results, human experience such as occupational data and data from accident databases, epidemiological and clinical studies and well-documented case reports and observations. The quality and consistency of the data shall be given appropriate weight. Information on substances or mixtures related to the substance or mixture being classified shall be considered as appropriate, as well as site of action and mechanism or mode of action study results. Both positive and negative results shall be assembled together in a single weight of evidence determination'*.

The ED criteria list a number of factors (see Table 1) which must be considered in the WoE assessment. In addition, the recommendations given in the EFSA Guidance on WoE should be considered (EFSA Scientific Committee, 2017).

It should be noted that in this guidance, the WoE methodology as indicated in the criteria is used in two different contexts:

- First, WoE is applied for the evaluation of the line(s) of evidence for adversity and/or endocrine activity. Here, an assessment of the available relevant scientific data based on a WoE approach is carried out to determine whether there is sufficient empirical support for the assembled lines of evidence (see Section 3.3.1 and 3.3.2); and
- Second, WoE is used for the mode of action analysis. The result of this analysis is used to establish if there is a link between the adverse effect(s) and the endocrine activity (see Section 3.5).

Expert judgement will be necessary when considering the available lines of evidence, including the overall evaluation of the consistency of the data set as a whole.

Table 1: Factors listed in the ED criteria (Commission Delegated Regulation (EU) No 2017/2100[3] and Commission Regulation (EU) No 2018/605[4]) which must be considered in the weight of evidence assessment

Factors with respect to humans	Factors with respect to non-target organisms
'both positive and negative results'	*'both positive and negative results, discriminating between taxonomic groups (e.g. mammals, birds, fish, amphibians) where relevant'*
'the relevance of the study designs, for the assessment of adverse effects and of the endocrine mode of action' (c)	*'the relevance of the study design for the assessment of the adverse effects and its relevance at the (sub) population level, and for the assessment of the endocrine mode of action'* (c)
	'the adverse effects on reproduction, growth/development, and other relevant adverse effects which are likely to impact on (sub)populations. Adequate, reliable and representative field or monitoring data and/or results from population models shall as well be considered where available'
'the quality and consistency of the data, considering the pattern and coherence of the results within and between studies of a similar design and across different species'	*'the quality and consistency of the data, considering the pattern and coherence of the results within and between studies of a similar design and across different taxonomic groups'*
'the route of exposure, toxicokinetic and metabolism studies'	
'the concept of the limit dose, and international guidelines on maximum recommended doses and for assessing confounding effects of excessive toxicity'	*'the concept of the limit dose, and international guidelines on maximum recommended doses and for assessing confounding effects of excessive toxicity'*

[6] Regulation (EC) No 1272/2008 of the European Parliament and of the Council of 16 December 2008 on classification, labelling and packaging of substances and mixtures, amending and repealing Directives 67/548/EEC and 1999/45/EC, and amending Regulation (EC) No 1907/2006.OJ L 353, 31.12.2008, p. 1–1355. http://data.europa.eu/eli/reg/2008/1272/oj

Factors with respect to humans	Factors with respect to non-target organisms
'the biological plausibility of the link between the adverse effects and the endocrine mode of action'[(c)]	*'the biological plausibility of the link between the adverse effects and the endocrine mode of action'*[(c)]

The criteria for determining endocrine-disrupting properties[(a),(b)] state that '*in applying the weight of evidence determination the assessment of quality, reliability, reproducibility and consistency of the scientific evidence shall, in particular, consider all of the following factors*'. The factors to be considered differ depending on whether the assessment of endocrine-disrupting properties is with respect to humans or non-target organisms.

(a): Commission Delegated Regulation (EU) 2017/2100 of 4 September 2017 setting out scientific criteria for the determination of endocrine-disrupting properties pursuant to Regulation (EU) No 528/2012 of the European Parliament and Council. OJ L 301, 17.11.2017, p. 1–5. Available online: http://data.europa.eu/eli/reg_del/2017/2100/oj

(b): Commission Regulation (EU) 2018/605 of 19 April 2018 setting out scientific criteria for the determination of endocrine-disrupting and amending Annex II to Regulation (EC) 1107/2009. OJ L 101, 20.4.2018, p. 33–36. Available online: http://data.europa.eu/eli/reg/2018/605/oj

(c): 'Endocrine mode of action' should be read as 'endocrine activity' (see Section 3 for details).

3.1. General overview of the assessment strategy

In order to determine whether a substance causes adverse effect(s) that can be plausibly linked to endocrine activity, all ED relevant information and supporting toxicity information on the substance needs to be collected and assessed in accordance with this guidance.

According to the Commission Delegated Regulation (EU) No 2017/2100[3] and Commission Regulation (EU) No 2018/605[4], the conclusions as to whether the ED criteria are met need to be drawn separately with respect to humans and non-target organisms. However, it should be highlighted that there may be data available on non-target organisms relevant for the assessment of the ED properties with regard to humans. Furthermore, because of the high level of conservation of the endocrine system across taxonomic groups, the mammalian data may also be relevant for other vertebrates (OECD, 2018b). Therefore, data on mammals and other taxa are considered together in a holistic approach as part of the available evidence, but also for identifying potential data gaps when assembling lines of evidence for endocrine activity and/or endocrine-related adversity. This means, for example, that information on endocrine effects in fish/amphibians, could be used to investigate the mammalian data set with heightened scrutiny for similar effects and to target potential requests for the generation of further mammalian information, or vice versa.

It is recognised that the standard information requirements for BPs and PPPs currently require more studies which may be informative on ED properties with regard to human health and mammals than for other taxonomic groups. Thus, in line with the general principle of desired reduction of unnecessary animal testing, the assessment strategy aims at making the most efficient use of the available data set to reach a conclusion. Therefore, it is recommended to strive for a conclusion on the ED properties with regard to humans and in parallel, using the same database, to strive for a conclusion on mammals as non-target organisms. Only where, based on this assessment, the criteria are not met for mammals as non-target organisms, would the assessment need to proceed to the other taxonomic groups, which may require the generation of additional data. It is sufficient that the substance meets the ED criteria in one taxonomic group in order to conclude that a substance meets the ED criteria for non-target organisms.

Where the evidence available indicates that the criteria are not met for mammals, the assessment for non-target organisms should proceed by considering fish and amphibians, because these are the taxa where standardised test methods and knowledge on how to interpret the results are available. Information on other taxa (e.g. birds and reptiles) should be considered if available. It should be recognised that currently investigation of ED properties in these taxa is hampered by a lack of test methods investigating endocrine specific endpoints. Once such methods become available, they should be considered in the ED assessment strategy with regard to non-target organisms.

The suggested sequence for reaching the conclusions is only a general recommendation that suits most of the cases. However, it does not preclude that, depending on the available information, another sequence to reach the conclusions on the ED criteria could be more efficient. For example, in cases where a substantial amount of data is available for non-target organisms (e.g. on fish) that would allow to start the assessment from non-target organisms other than mammals.

There may be cases in which due to the knowledge on the physico-chemical and (eco)toxicological properties of the substance an ED assessment does not appear scientifically necessary or testing for this

purpose not technically possible (BP Regulation[1], Annex IV or PPP Regulation,[2] Annex, Point 1.5). In such cases, it should be justified for PPPs (Commission Regulation (EU) No 283/2013[7]) or the general rules for adaptation of the data requirements set out in Annex IV of the BP Regulation[1] shall be followed or, for PPPs, used as examples. However, it needs to be considered if possible adaptations would apply to the ED assessment in its entirety or only with respect to humans or non-target organisms.

In some cases, the ED assessment may not change the applicable regulatory consequences if the substance already fulfils any of the other exclusion criteria set out in Article 5(2) of the BP Regulation[1] or Article 4 of the PPP Regulation.[2] However, the assessment of the ED properties is still to be considered in case the active substance may be approved under restricted conditions or may be subject to mitigation measures as set out in Article 5(2) of the BP Regulation[1], points 3.6.3–3.6.5 of Annex II of the PPP Regulation,[2] or Article 4(7) of the PPP Regulation.[2]

The next sections explain the core concept of the assessment approach, i.e. the grouping of parameters relevant for identification of ED properties measured in experimental studies with respect to their capacity to inform on endocrine activity and related adversity and the steps of the assessment strategy, including specific considerations for non-EATS modalities (see Section 3.1.2).

3.1.1. Grouping of parameters relevant for identification of endocrine-disrupting properties

The OECD GD 150 lists tests (test guidelines) and parameters that are considered relevant when investigating the ED properties of substances. In addition, the OECD GD 150 provides guidance on how to interpret parameters relevant for identification of ED properties measured in the standardised test guidelines with respect to EATS modalities (considerations for dealing with non-EATS modalities are reported in Section 3.1.2).

In the context of this guidance, all the parameters listed by the OECD GD 150 (and measured in assays listed in the OECD CF) are grouped into four groups, which have been adapted for the purpose of this guidance from the JRC screening methodology to identify potential endocrine disruptors (JRC, 2016). The grouping reflects the fact that, based on OECD GD 150, some effects are considered to be strong indicators of effects being mediated by an EATS modality, while some others are considered to be potentially sensitive to, but not diagnostic of, EATS modalities. Furthermore, some parameters are measured by *in vitro* test methods and others by *in vivo* test methods. In general, *in vitro* effects provide information on the mechanism through which a substance may exert endocrine activity (e.g. by binding to and activating a receptor), whereas, *in vivo* test methods may inform on endocrine activity, adverse effects or both. The grouping of *in vivo* parameters mainly reflects the OECD CF levels, as described in OECD GD 150, OECD CF level 3 referring to '*In vivo assays providing data about selected endocrine mechanism(s)/pathway(s)*' while OECD CF levels 4 and 5 refer to '*in vivo assays providing data on adverse effects on endocrine relevant endpoints*'. Parameters measured in OECD CF levels 4 and 5 are not by default considered adverse and should be assessed according to a WoE approach as explained in the following sections of this guidance. The same attribution to the different groups can be applied when these parameters are investigated in non-guideline studies. Similarly, level 1 information ('*existing data and existing or new non-test information*') from OECD CF should be assigned to the corresponding group.

In the context of this guidance, this grouping is considered very helpful for guiding the assessors in the evaluation of the scientific evidence when identifying substances with ED properties. The four groups are:

- ***In vitro* mechanistic** – parameters measured *in vitro*, that provide information on the mechanism through which a substance could be considered endocrine active (e.g. by binding to and activating a receptor or interfering with hormone production). These parameters are measured in assays currently placed under OECD CF level 2.
- ***In vivo* mechanistic** – parameters measured *in vivo* that provide information on endocrine activity that are usually not considered adverse. This group applies mainly to parameters measured within assays placed at OECD CF level 3. In addition, changes in hormone levels are considered *in vivo* mechanistic even when they are measured in OECD CF level 4 and 5

[7] Commission Regulation (EU) No 283/2013 of 1 March 2013 setting out the data requirements for active substances, in accordance with Regulation (EC) No 1107/2009 of the European Parliament and of the Council concerning the placing of plant protection products on the market. OJ L 93, 3.4.2013, p. 1–84. Available online: http://data.europa.eu/eli/reg/2013/283/oj

assays.[8] It should be noted that certain parameters within OECD CF level 3 *in vivo* assays when measured in an intact animal model (e.g. Hershberger assay OECD TG 441 (OECD, 2009d) or fish short-term reproduction assays OECD TG 229 (OECD, 2012a)) may also provide additional information on adversity in certain circumstances (Tables 13, 15 and 16) and therefore should be treated as those parameters grouped as 'EATS-mediated' or 'sensitive to, but not diagnostic of EATS' (see below).

- **EATS-mediated** – parameters measured *in vivo* that may contribute to the evaluation of adversity, while at the same time (due to the nature of the effect and the existing knowledge as described in OECD GD 150) they are also considered indicative of an EATS MoA and thus (in the absence of other explanations) also imply underlying *in vivo* mechanistic information. This group includes the parameters mainly from OECD CF level 4 and 5 tests labelled in OECD GD 150 as 'endpoints for estrogen-mediated activity', 'endpoints for androgen-mediated activity', 'endpoints for thyroid-related activity' and/or 'endpoints for steroidogenesis-related activity'. In specific cases, as already explained in the previous group, when measured in an intact animal, also parameters measured in OECD CF level 3 assays can provide EATS-mediated information.
- **Sensitive to, but not diagnostic of, EATS** – parameters measured *in vivo* that may contribute to the evaluation of adversity, however, due to the nature of the effect and the existing knowledge as described in OECD GD 150, these effects cannot be considered diagnostic on their own of any one of the EATS modalities. Nevertheless, in the absence of more diagnostic parameters, these effects might provide indications of an endocrine MoA that might warrant further investigation. This includes parameters from OECD CF level 3, 4 and 5 *in vivo* assays and labelled in OECD GD 150 as endpoints potentially 'sensitive to, but not diagnostic of, EATS modalities'.

Tables 12, 13, 14, 15, 16 and 17 in Section 4 report the main parameters investigated in the test guidelines and their attribution to the different groups outlined above.

3.1.2. Considerations on non-EATS modalities

Adversity associated with some 'sensitive to, but not diagnostic of, EATS' parameters may be also a consequence of disruption in other endocrine modalities, i.e. non-EATS. For example adversity in the adrenal and/or pituitary can be consequent to disruption of the hypothalamic-pituitary-adrenal axis resulting in altered stress response. While, the currently available OECD test guidelines can detect apical effects potentially relevant for ED identification through other modalities than EATS (European Commission, 2018; Manibusan and Touart, 2017), there are currently no OECD test methods to elucidate the potential non-EATS mechanism eliciting those apical effects. However, there are methods described in scientific literature which could provide mechanistic information for non-EATS modalities.

In addition, other parameters measured in standard tests not labelled as 'EATS-mediated' or 'sensitive to, but not diagnostic of, EATS' may also be indicative of non-EATS endocrine effects (e.g. histopathological findings in the pancreas or serum levels of corticosterone, insulin, glucose, etc.).

In the absence of internationally validated test methodologies, no specific guidance can be given here on how to identify potential links between such effects to non-EATS endocrine modalities. However, concerns for non-EATS endocrine effects warrant additional investigation to the extent possible. If information on non-EATS modalities is available, e.g. where literature data provide mechanistic information, this shall be taken forward in the assessment and considered for assembling lines of evidence and in the MoA analysis, as explained in Sections 3.3–3.5.

3.1.3. The assessment strategy

The assessment strategy is based on the three conditions stipulated in the ED criteria (adversity, endocrine activity and a biologically plausible link between the two) and on the grouping of the parameters as described above. The 'EATS-mediated' parameters listed in the OECD GD 150 drive the assessment strategy because, by providing evidence for both endocrine activity and the resulting potentially adverse effects, they are considered indicative of an endocrine MoA.

It should be noted that generally parameters which are considered as 'sensitive to, but not diagnostic of, EATS' and 'EATS-mediated' parameters are normally investigated in the same level 4 or level 5 tests. Thus, if there is no adversity seen in the 'EATS-mediated' parameters, but adversity is

[8] Further guidance on measurement of hormone levels is given in Appendix B.

observed in the same study in parameters considered 'sensitive to, but not diagnostic of, EATS', then this adversity is not likely to be caused by alterations of the EATS modalities. However, there may be situations where the 'EATS-mediated' parameters are not sufficiently investigated (i.e. in level 4 or 5 tests carried out according to outdated guidelines or in the case of non-target organisms where it is likely that only results from OECD TG 210 (OECD, 2013b) tests are available). In such cases, any adversity observed in parameters considered 'sensitive to, but not diagnostic of, EATS', cannot be dismissed. Further guidance on how to proceed with the assessment in case only 'sensitive to, but not diagnostic of, EATS' parameters can be found in Sections 3.4.4.2 and 3.5.2. The general principles of the assessment strategy are applicable also to non-EATS modalities (see further guidance in Sections 3.3 and 3.5.2).

The assessment strategy is applicable both for humans and non-target organisms and in both cases the guidance specified in Sections 3.2–3.5 of this document needs to be followed. Figure 1 illustrates the steps of the assessment. Each of the steps outlined in the figure are described below. The general assessment strategy includes:

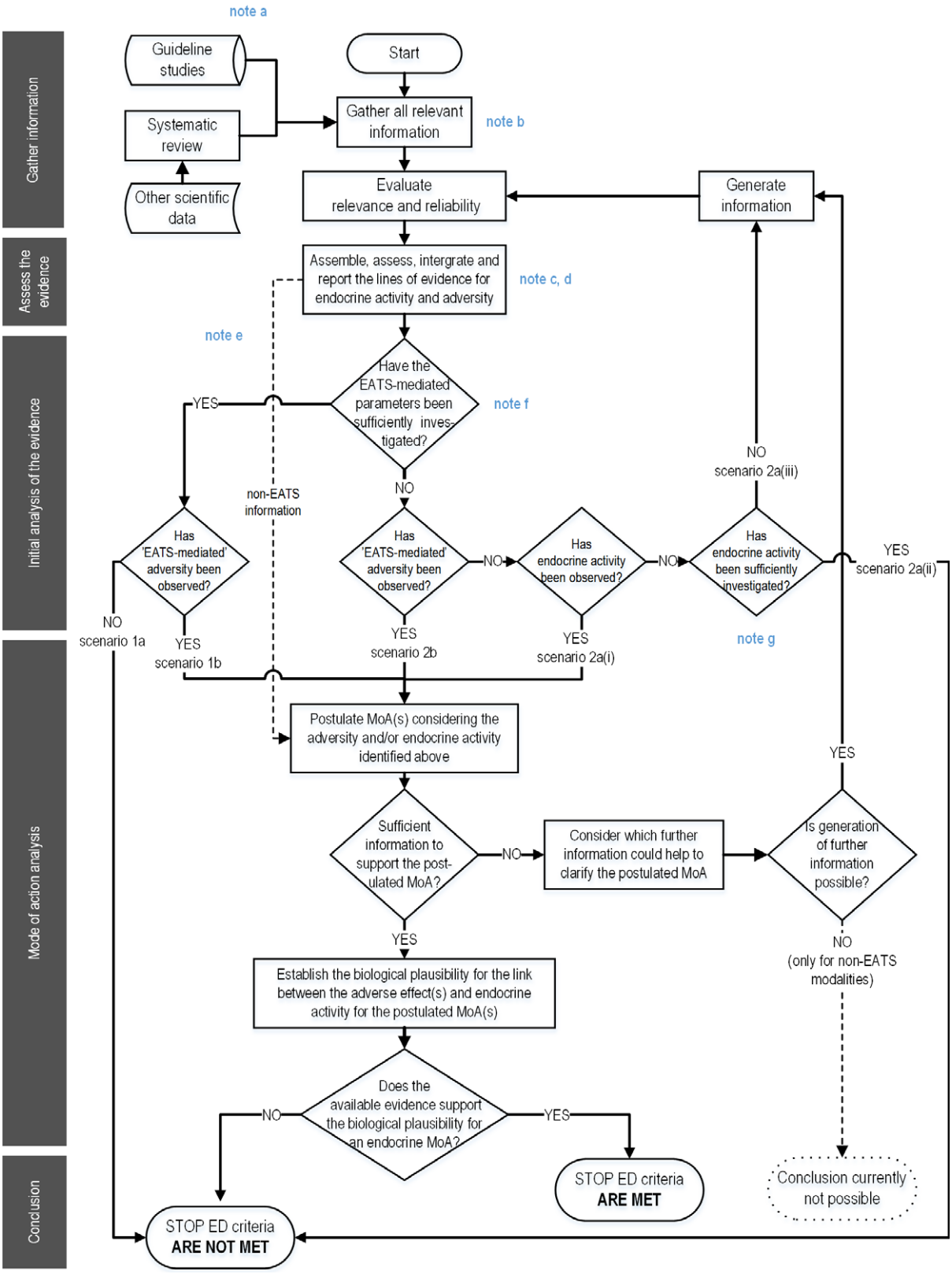
Gather information. In this step, all available relevant information is gathered both in terms of scientific data generated in accordance with internationally agreed study protocols, and different types of scientific data retrieved with systematic review methodology. All types of data described in Section 4 could be considered, and where relevant, included in the dossier for enabling the assessment of the ED properties. This information shall be evaluated for its relevance and reliability, and extracted and reported in the draft assessment-, renewal assessment- and competent authority reports (DAR/RAR/CAR). Guidance on how to perform this step is given in Section 3.2.

Assess the evidence. In this step, the information is assembled into lines of evidence, integrating information for both adversity and endocrine activity. The lines of evidence are assessed and reported in the dossier/DAR/RAR/CAR. Guidance on how to perform this step is given in Section 3.3. If there is indication of non-EATS-related endocrine activity and/or effects, this should be taken forward to the MoA analysis step because the questions asked in the next step are tailored to the EATS modalities.

Initial analysis of the evidence. This step includes a decision tree with different possible scenarios. The scenarios are driven by the availability of 'EATS-mediated' parameters and/or evidence of endocrine activity and provide indication to the assessor and the applicant of the situations where the available evidence either allows to conclude that a substance does not meet the ED criteria, or where additional information is needed, or where a MoA analysis is required to conclude on the ED properties. Guidance on how to perform this step is given in Section 3.4.

MoA analysis. This step aims to establish if there is a biologically plausible link between observed adverse effects and endocrine activity. Different situations are outlined. Depending on the available evidence, the applicant and the assessor need to identify the information that may need to be generated and included in the dossier in order to further investigate the adversity or the endocrine activity, or any potential alternative MoA(s). Guidance on how to conduct and document a MoA analysis and how to establish if there is a biologically plausible link between observed adverse effects and endocrine activity is given in Section 3.5. In this step, it should be further investigated whether it is possible to establish if there is a plausible link between non-EATS endocrine activity and observed adversity or whether further information could be generated to clarify whether there is a non-EATS endocrine-disrupting MoA.

Conclusion on the ED criteria. In this step, the conclusion as to whether the ED criteria are met with respect to humans and non-target organisms is drawn and transparently documented, including the remaining uncertainties (see Section 3.6). Different situations are outlined, depending on the outcome of the MoA analysis.


Gather information
Assess the evidence
Initial analysis of the evidence
Mode of action analysis
Conclusion
note a
Guideline studies
Systematic review
Other scientific data
Start
Gather all relevant information
note b
Evaluate relevance and reliability
Generate information
Assemble, assess, intergrate and report the lines of evidence for endocrine activity and adversity
note c, d
note e
Have the EATS-mediated parameters been sufficiently inves-tigated?
note f
YES
NO
non-EATS information
NO scenario 2a(iii)
Has 'EATS-mediated' adversity been observed?
Has 'EATS-mediated' adversity been observed?
NO
Has endocrine activity been observed?
NO
Has endocrine activity been sufficiently investigated?
YES scenario 2a(ii)
NO scenario 1a
YES scenario 1b
YES scenario 2b
YES scenario 2a(i)
note g
Postulate MoA(s) considering the adversity and/or endocrine activity identified above
Sufficient information to support the post-ulated MoA?
NO
Consider which further information could help to clarify the postulated MoA
Is generation of further information possible?
YES
NO (only for non-EATS modalities)
YES
Establish the biological plausibility for the link between the adverse effect(s) and endocrine activity for the postulated MoA(s)
Does the available evidence support the biological plausibility for an endocrine MoA?
NO
YES
STOP ED criteria ARE MET
Conclusion currently not possible
STOP ED criteria ARE NOT MET

Notes to Figure 1: The assessment strategy illustrated in the flow chart is applicable both for humans and non-target organisms and in both cases the guidance specified in Sections 3.2–3.5 of this document needs to be followed. The assessment strategy is driven by the availability of 'EATS-mediated' parameters as these provide evidence for both endocrine activity and the resulting potentially adverse effects. However, there may be situations where the 'EATS-mediated' parameters are not, sufficiently, investigated (e.g. this is very likely the case for non-target organisms). In such cases, it may be possible to follow the assessment strategy using the 'sensitive to, but not diagnostic of, EATS' parameters (without the need to generate additional information on EATS-mediated parameters i.e. in case of scenarios 2a(i) or 2b). If the required data are available, it is in principle possible to establish endocrine disrupting MoA(s) on the basis of parameters indicating 'sensitive to, but not diagnostic of, EATS' potential adversity and EATS endocrine activity.
Note a: According to the ED criteria3 for BCPs '*scientific data generated in accordance with internationally agreed study protocols, in particular those referred to in Annexes II and III of Regulation (EU) No 528/2012*'; according to the ED criteria4 for PPP '*scientific data generated in accordance with internationally agreed study protocols, in particular, those listed in the Commission Communications in the framework of setting out the data requirements for active substances and plant protection products, in accordance with this Regulation'.*
Note b: As discussed in Section 3.1, some substances may not need to be assessed for ED properties.
Note c: For details on population relevance see Section 3.3.1.4.
Note d: For details on effects secondary to other toxicities see Section 3.3.1.1.
Note e: If information on non-EATS modalities becomes available, e.g. through systematic review of the literature, this needs be taken forward in the assessment. In such cases, after gathering and assessing the information, and assembling and reporting the lines of evidence (see Section 3.3), the non-EATS information can be taken forward directly to the MoA analysis (the step 'initial analysis of the evidence' is not applicable), as explained in Section 3.5.
Note f: For details on EATS-mediated adversity considered sufficiently investigated see Section 3.4.
Note g: For details on endocrine activity considered sufficiently investigated: see Section 3.4.2 and 3.4.3.

Figure 1: Flow chart illustrating the ED assessment strategy

3.2. Gather all relevant information

According to the ED criteria, *the identification of a [...] substance [...] as having endocrine-disrupting properties [...] shall be based on all of the following points:*

(1) *all available relevant scientific data (in vivo studies or adequately validated alternative test systems predictive of adverse effects in humans or animals; as well as in vivo, in vitro, or, if applicable, in silico studies informing about endocrine modes of action):*

- *scientific data generated in accordance with internationally agreed study protocols [...];*
- *other scientific data selected applying a systematic review methodology [...].'*

The applicant should provide all relevant scientific data, which can give information on (potential) ED properties in the dossier. This means that the dossier must provide all the required information, i.e. standard guidelines studies as required in the respective data requirements and any other relevant scientific data.

The standard information requirements for PPPs and BPs include a number of studies, in accordance with internationally agreed study protocols (standard studies), that are useful for the ED assessment. These are listed in Tables C.1 and C.2 in Appendix C according to the current legal frameworks.

It should be highlighted that, applicants should generate all the information needed to enable a conclusion. Further details on what types of potential data are needed are given in Sections 3.4 and 3.5 and Section 4.

The dossier should also include other scientific data selected applying a systematic review methodology. Systematic review is a method that aims to systematically identify, evaluate and synthesise evidence for a specific question. In addition, it promotes a more structured and transparent use of the body of evidence and it reduces bias in the selection of the studies by the extensiveness and reproducibility of the entire process.

EFSA guidance on application of systematic review methodology to food and feed safety assessments to support decision making (EFSA, 2010); and the EFSA guidance on submission of scientific peer-reviewed open literature for the approval of pesticide active substances (EFSA, 2011) should be followed both for PPPs and BPs ED hazard assessment. These guidance documents provide instructions on how to identify and select scientific literature according to the principles of the

systematic literature review. To ensure those fundamental features of the systematic literature search, definition of the review question and the criteria for relevance and reliability should be defined *a priori*.

When conducting a systematic literature search the search strategy can be conducted following two general search approaches as recommended by (EFSA, 2011) for an example see Appendix F:

- A single concept search strategy in order to capture all the information about the substance in one search. This is performed by using search terms related to the substance and its synonyms (e.g. CAS number, IUPAC name, etc.).
- A targeted search strategy for individual endpoints. For endocrine disruption, if this option is used, a proper search strategy should be designed allowing to avoid bias and capture as much as possible relevant scientific literature data.

In the context of this Guidance, it is suggested to perform as a starting point, the literature search by using the single concept approach since it is considered to be highly sensitive, and less time consuming than the targeted search strategy. If a large number of hits is retrieved by using the single concept approach, this can be further refined by running a search targeted on the information requirements.

It is recognised that a systematic literature review would identify most of the published information on a substance and could therefore be a mix of summaries of standard guideline studies (if published), academic investigations (generally non-guideline), (Q)SAR models, epidemiological studies, environmental field studies, monitoring data and population modelling, etc.

The systematic review should include all relevant published scientific information. There may be information contained within various databases (see Table 10 and Table D.1 in Appendix D), which are highly relevant for the identification of ED properties and, therefore, it is recommended that those should be always searched and documented in the dossier. If available this kind of information must be assessed for its relevance and reliability (see Section 3.2.1).

In addition to the above sources of data, there may also be specific non-guideline information (e.g. (Q)SAR predictions, population modelling) conducted by the applicant to support the approval. Although not conducted following internationally agreed study protocols or retrieved through the systematic literature review such data should be considered as part of the information for ED hazard identification, after an assessment of their relevance and reliability according to Section 3.2.1. This may also include level 1 data of the OECD CF (Table 9).

3.2.1. Evaluate relevance and reliability of the data

Each piece of information provided in the dossier (e.g. experimental study, (Q)SAR prediction, etc.) has to be assessed for its relevance and reliability.

Relevance – Data relevance refers to the appropriateness of the data for the intended purpose of the assessment; adapted from EFSA (2015); Klimisch et al. (1997); Vermeire et al. (2013).

Reliability – in Klimisch et al. (1997), reliability is defined as 'the inherent quality of a test report or publication relating to preferably standardised methodology and the way the experimental procedure and results are described to give evidence of the clarity and plausibility of the findings'. However, while test reporting is considered essential for assessing the reliability, it is not itself a part of reliability criteria (EFSA, 2015). Reliability of data is closely linked to the reliability of the test method used to generate the data.

For BPs, further guidance on relevance and reliability is provided in the ECHA 'Guidance on information requirements and chemical safety assessment' (ECHA, 2011), the ECHA 'Guidance on the Biocidal Products Regulation: Volume III Human Health, Assessment and Evaluation (Parts B+C) (ECHA, 2017a), and the ECHA 'Guidance on the Biocidal Products Regulation: Volume IV Environment, Assessment and Evaluation (Parts B+C)' (ECHA, 2017b). Further information on relevance and reliability for PPPs is provided in the EFSA Guidance (EFSA, 2010, 2011).

3.2.1.1. Data from standard studies

Studies generated according to EU test methods and/or internationally agreed study protocols, when they include parameters which are informative of endocrine-related adversity and/or endocrine activity, are generally considered relevant for the identification of ED properties of a substance.

In the context of the ED hazard identification, new studies should be carried out according to the latest version of the corresponding test guidelines to be considered fully relevant. This is of particular importance

since in recent years a number of test guidelines have been revised to include additional parameters which are relevant for identification of ED properties (e.g. the latest version of OECD TG 416 (OECD, 2001)).

It is recognised that the available information on a substance generated according to older versions of guidelines may be relevant for the identification of ED properties. However, parameters considered highly relevant for the identification of ED properties may be missing. Therefore, when evaluating the relevance of studies conducted according to outdated guidelines, it is very important to consider what parameters relevant for identification of ED properties were included in the study design. Missing parameters with respect to the updated versions of the test guidelines should be clearly reported, and this may lead to the need to generate additional information. Further guidance on how to proceed if the EATS-mediated parameters are not sufficiently investigated is given in Section 3.4.

Additionally, when evaluating the relevance of toxicity studies, effects are considered adequately characterised if doses up to the maximum tolerated dose or limit dose, as defined in the related OECD guidelines, are used. For ecotoxicology, the highest test concentration should be set by the maximum tolerated concentration determined from a range-finding test or from other toxicity data.

Relevant studies should undergo a reliability assessment. When evaluating the standard studies, their reliability is considered based on the validity criteria of the test guidelines. Deviations with respect to the recommendations in the standard guidelines should be reported and their influence on the study results should be evaluated on a case-by-case basis.

3.2.1.2. Other scientific data

The following section is intended to provide additional guidance on how to evaluate data quality for different types of scientific data which will be selected using systematic review. Furthermore, general indications are given on how to consider data that may be available in the dossier, but were not retrieved through systematic review (e.g. *in silico* predictions/modelling).

According to the EFSA guidance documents (EFSA, 2011), the selection of relevant studies is normally carried out in two steps. An initial rapid assessment based on the screening of titles and abstracts is conducted in order to exclude those publications which are clearly irrelevant. Those studies which are of unclear relevance and the ones which appear to be relevant go to the second step, i.e. detailed assessment of the full text. The guidance only gives general principles with regard to the assessment of relevance and reliability of the literature. Relevance criteria should not be too restrictive and the identification of relevance criteria should be considered an iterative process that starts with a clear analysis of the different components of the data requirements to set the main characteristics a relevant study should have. A preliminary search of the literature may be useful to test and refine the relevance criteria on a subset of summary records or full-text documents, to assess their applicability. The assessment of study relevance does not involve considerations of study reliability, which refers to the evaluation of the inherent quality of a study, its precision and accuracy and refers to the extent to which a study is free from bias.

When assessing reliability, some general considerations could be taken into account, such as statistical power, verification of measurement methods and data, control of experimental variables that could affect measurements, etc. Studies retrieved through the literature review may be conducted according to standardised protocols. In this case, a reasonable approach for evaluation would be to apply validity and quality criteria that are included in the most relevant test guidelines.

Non-guideline studies

Literature retrieved through the systematic review can also include non-guideline information. Non-guideline information is evaluated for quality on a case-by-case basis. In general, the same principles for relevance and reliability apply as for literature data outlined above. However, as the parameters investigated in the studies may be non-standardised, additional considerations may be needed to establish the reliability and relevance of such studies.

(Q)SAR models and read-across approaches

Publications describing the output of (Q)SAR models and/or read-across approaches may be available in the literature or performed by the applicant on a case-by-case basis. The scientific validity and reliability of a (Q)SAR model is evaluated following the five OECD principles for validation of (Q)SAR models (OECD, 2014). A model is considered valid when it models a defined endpoint; has an unambiguous algorithm; has a defined domain of application; includes appropriate measures of goodness-of-fit, robustness and predictivity and it is related to mechanistic interpretation. The relevance of the QSAR model predictions needs to be assessed on a case-by-case basis. The reliability

of an *in silico* prediction is related to the definition of the chemical space covered by the model, i.e. the applicability domain of the model. The target substance should be within the applicability domain of the model for a reliable prediction. Knowledge-based models do not have a defined training set and therefore the information on the applicability domain is missing. However, these models might provide complementary information, e.g. suggested endocrine activity, examples and references that can be used to assess the reliability of the prediction. Guidance on how to report (Q)SARs is provided by the ECHA Guidance on information requirements and chemical safety assessment, Chapter R.6: QSARs and grouping of chemicals (ECHA, 2008).

The relevance and reliability of a read-across prediction can be evaluated following the ECHA 'Read-across assessment framework' (ECHA, 2017c,d). General guidance on read-across and grouping of substances are provided by the ECHA Guidance on information requirements and chemical safety assessment, Chapter R.6: QSARs and grouping of chemicals (ECHA, 2008).

In vitro *methods*

Mechanistic *in vitro* data can potentially provide evidence for endocrine activity, which is considered a key information in the assessment. However, only few *in vitro* assays are currently available as OECD test guidelines. The assessment of available data should at least consider the relevance of the cell system used, the exposure concentrations and metabolic capacity of the test system. There are many factors to be considered when conducting or evaluating *in vitro* assays. A guidance document on Good *In Vitro* Method Practices (GIVIMP) for the development and implementation of *in vitro* methods for regulatory use in human safety assessment has recently published (OECD, 2018a). The document is intended to reduce the uncertainties in cell and tissue-based *in vitro* method derived predictions by applying all necessary good scientific, technical and quality practices from *in vitro* method development to *in vitro* method implementation for regulatory use (OECD, 2018a). This document describes the process of validation, interpretation of data and sources of interference that need to be considered as they might lead to false positive or negative results.

Databases of compiled data

No specific indication can be given for the evaluation of data extracted from existing databases (e.g. ToxCast and others listed in Table 10 and in Appendix D). Therefore, a case-by-case evaluation of these data should be performed where sufficient details (e.g. experimental design details as concentrations tested, positive and negative controls, cell medium, time of incubation, etc.) are provided to allow an independent assessment.

Epidemiological data

According to data requirements for PPPs and BPs, relevant epidemiological studies shall be submitted, and considered where available. Epidemiological outcomes (e.g. positive association observed between PPP or BPs exposures and occurrence of potentially endocrine-related effects) should be considered as part of the WoE approach and integrated with the experimental toxicological data. EFSA recently published a scientific opinion on the use of epidemiological data, including relevance and reliability, and proposed their integration with experimental data (EFSA PPR Panel, 2017).

Field studies, monitoring data and population modelling

These types of studies can be available in the literature or conducted *ad-hoc* by the applicant. In particular population models are generally published. However, substance-specific modelling can be included in the dossier. Setting general rules for the evaluation of field studies and monitoring data is complicated. In general, it is necessary to perform a case-by-case evaluation, i.e. due to the high variability it is not possible to set common criteria. These studies should be evaluated for their scientific merit by following the indications already included in available guidance documents (e.g. the EFSA Guidance on Risk Assessment for Birds and Mammals (EFSA, 2009)). Different population models are available which can be used for different purposes and for answering different questions. For instance, a key question which could be addressed by population models is the degree of reproductive impairment which is likely to trigger consequences at the population level. Models are available to address this specific question (Topping and Luttik, 2017; White et al., 2014), however, generally these models are more suitable for risk assessment purpose. In addition, although models have the advantage that different environmental situations can be simulated and extrapolation in time is possible, at present they are not routinely used for the approval of active substance at EU level due to the lack of standard and validated models. The standardisation and validation of models should ensure

that model predictions at population level are reliable and realistic (Kramer et al., 2011). Although there is currently no generally accepted models, a detailed description of how to develop models for regulatory purposes and how to evaluate them is provided in the EFSA PPR opinion on good modelling practice (EFSA PPR Panel, 2014). In conclusion, while the mentioned tools are considered promising, they currently cannot be used to dismiss the population relevance of an adverse effect without further investigating the link between the effects observed in laboratory test and the population dynamics (Marty et al., 2017; Matthiessen et al., 2018; Mintram et al., 2018).

3.2.2. Extracting and reporting the information

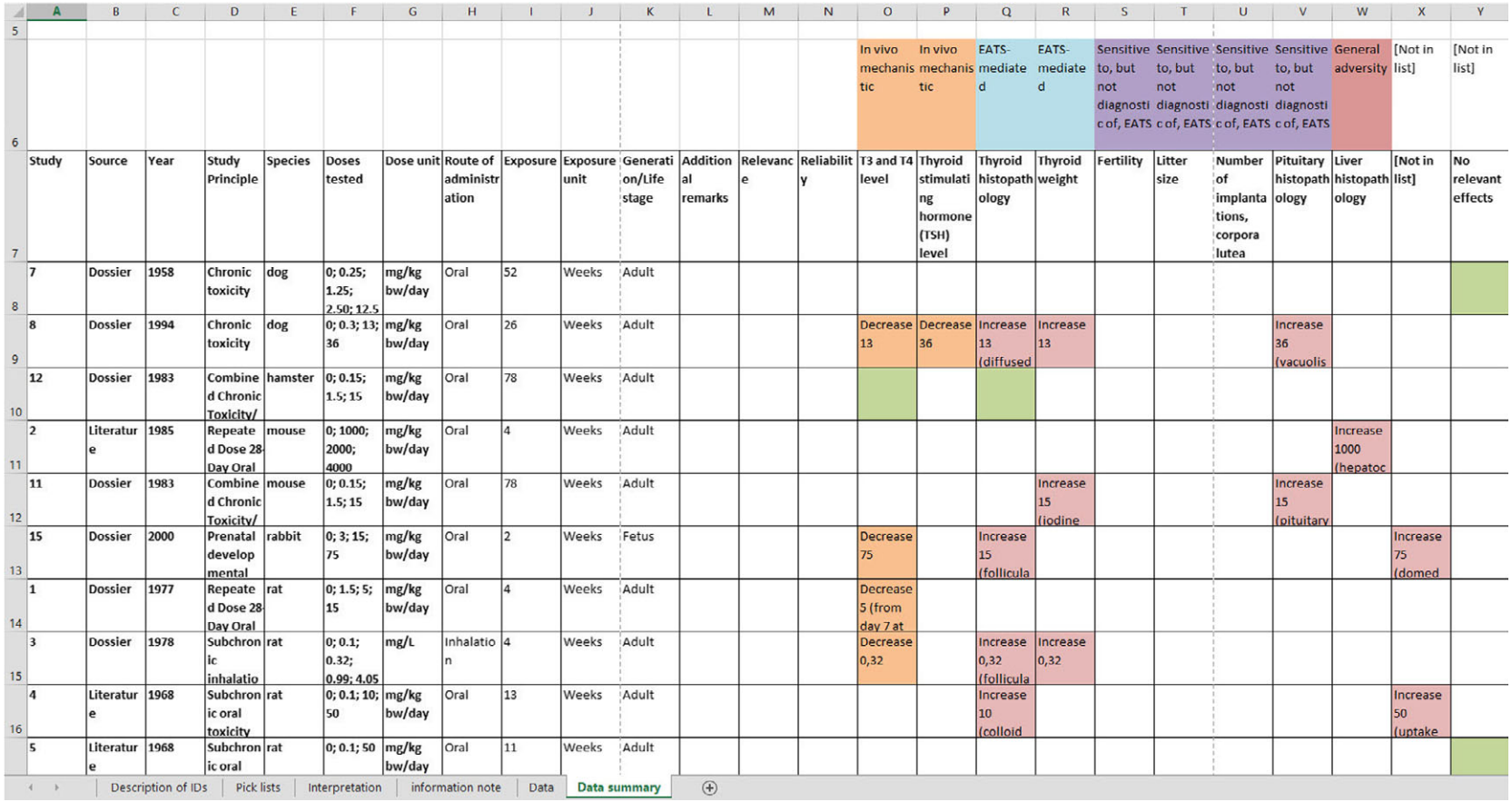
As a matter of normal practice, each study provided with the dossier by the applicants must be evaluated and summarised by the rapporteur Member State Competent Authorities with sufficient level of detail in the DAR/RAR/CAR. The literature review should also be included and transparently reported and evaluated. A summary of the relevant studies retrieved from the literature should be included with an evaluation of their relevance and reliability. The applicant should provide summaries of the studies with the dossier. Applicants are strongly recommended to use the OECD harmonised templates[9] when reporting the studies in the summary dossier.

All the parameters which are useful for the ED assessment, identified in each relevant and reliable study, should be reported in a tabular form to be provided by the applicant with the dossier in editable format.

It is suggested that available information is reported in the Excel template provided with this guidance (see Appendix E). This should also include consideration of systemic toxicity. Additional instructions on the elements (e.g. category of EATS modalities, dose–response, consistency within each study, etc.) to consider when completing the excel spreadsheet are provided in Appendix E. Both positive and negative results should be recorded. Both data from the mammalian toxicology section and the ecotoxicology section should be tabulated in a single spreadsheet (see Appendix E). The excel template consists of several columns which capture different type of information related to the study design and the effects observed in relevant parameters (e.g. OECD TG used, animal species, doses administered, exposure duration, type of effect observed, lowest-observed-adverse-effect level (LOAEL), etc.). In the template, each row reports the changes observed in a certain parameter within a specific study. Therefore, when in the same study, there are multiple effects reported there will be an equal number of rows to be filled in the template (one row per effect). In order to facilitate the evaluation of the data collected in the template, the data collected are re-organised into a data matrix which is built automatically. The advantage of the data matrix is that it shows within one row all the effects observed from one study (this allows summarising the information collected that was before spread over several rows in the template).

A screenshot of the data matrix is shown in Figure 2 as example.

[9] https://www.oecd.org/ehs/templates/harmonised-templates.htm



	Study	Source	Year	Study Principle	Species	Doses tested	Dose unit	Route of administration	Exposure	Exposure unit	Generation/Life stage	Additional remarks	Relevance	Reliability	T3 and T4 level	Thyroid stimulating hormone (TSH) level	Thyroid histopathology	Thyroid weight	Fertility	Litter size	Number of implantations, corpora lutea	Pituitary histopathology	Liver histopathology	[Not in list]	No relevant effects
															In vivo mechanistic	In vivo mechanistic	EATS-mediated	EATS-mediated	Sensitive to, but not diagnostic of, EATS	Sensitive to, but not diagnostic of, EATS	Sensitive to, but not diagnostic of, EATS	Sensitive to, but not diagnostic of, EATS	General adversity	[Not in list]	[Not in list]
8	7	Dossier	1958	Chronic toxicity	dog	0; 0.25; 1.25; 2.50; 12.5	mg/kg bw/day	Oral	52	Weeks	Adult														
9	8	Dossier	1994	Chronic toxicity	dog	0; 0.3; 13; 36	mg/kg bw/day	Oral	26	Weeks	Adult				Decrease 13	Decrease 36	Increase 13 (diffused	Increase 13				Increase 36 (vacuolis			
10	12	Dossier	1983	Combined Chronic Toxicity/	hamster	0; 0.15; 1.5; 15	mg/kg bw/day	Oral	78	Weeks	Adult														
11	2	Literature	1985	Repeated Dose 28-Day Oral	mouse	0; 1000; 2000; 4000	mg/kg bw/day	Oral	4	Weeks	Adult												Increase 1000 (hepatoc		
12	11	Dossier	1983	Combined Chronic Toxicity/	mouse	0; 0.15; 1.5; 15	mg/kg bw/day	Oral	78	Weeks	Adult							Increase 15 (iodine				Increase 15 (pituitary			
13	15	Dossier	2000	Prenatal developmental	rabbit	0; 3; 15; 75	mg/kg bw/day	Oral	2	Weeks	Fetus				Decrease 75		Increase 15 (follicula							Increase 75 (domed	
14	1	Dossier	1977	Repeated Dose 28-Day Oral	rat	0; 1.5; 5; 15	mg/kg bw/day	Oral	4	Weeks	Adult				Decrease 5 (from day 7 at										
15	3	Dossier	1978	Subchronic inhalatio	rat	0; 0.1; 0.32; 0.99; 4.05	mg/L	Inhalation	4	Weeks	Adult				Decrease 0,32		Increase 0,32 (follicula	Increase 0,32							
16	4	Literature	1968	Subchronic oral toxicity	rat	0; 0.1; 10; 50	mg/kg bw/day	Oral	13	Weeks	Adult						Increase 10 (colloid							Increase 50 (uptake	
	5	Literature	1968	Subchronic oral	rat	0; 0.1; 50	mg/kg bw/day	Oral	11	Weeks	Adult														

Description of IDs | Pick lists | Interpretation | information note | Data | Data summary

Figure 2: Screenshot of the Excel table provided in Appendix E, showing how the information collected is summarised

3.3. Assemble and assess lines of evidence for endocrine activity and adversity

A line of evidence is in broad terms a '*set of relevant information grouped to assess a hypothesis*' (EFSA Scientific Committee, 2017). In general, the lines of evidence are not fixed and different subsets of information can be identified according to the contribution they make towards answering the problem formulated.

In the context of this guidance, the assembling of lines of evidence should take into consideration all the available evidence (positive and negative) that have been evaluated as relevant and reliable as explained in Section 3.2.1. Relevant and reliable parameters should be assembled to determine whether and how they contribute to the lines of evidence for adversity and/or endocrine activity.

In particular, for the purpose of building lines of evidence, the parameters are grouped according to their potential to inform on EATS modalities into the groups described in Section 3.1.1 (based on the guidance provided by OECD GD 150), i.e. '*in vitro* mechanistic'-, '*in vivo* mechanistic'-, 'EATS-mediated'- and 'sensitive to, but not diagnostic of, EATS' parameters. However, if information on non-EATS modalities is available (see Section 3.1) this should be further considered and the same approach described in this section should be followed.

The lines of evidence for adverse effects and endocrine activity will be used to postulate (endocrine) MoA(s) and to understand if there is a biologically plausible link between the observed adverse effects and endocrine activity. If available, published MoAs and adverse outcome pathways (AOPs) could be useful for guiding the assembling of line(s) of evidence (see OECD AOP Knowledge Base (http://aopkb.org/) and as an example AOP 25 which was used to build Table 3). In particular, they may be useful to structure the information for facilitating the following steps of the assessment strategy (see also Figure 6 and Appendix G).

3.3.1. Assembling and assessing the line(s) of evidence for adverse effects

In the ED criteria, the identification of adverse effects is based on the WHO definition (WHO/IPCS, 2009) which is '*A change in the morphology, physiology, growth, development, reproduction or life span of an organism, system or (sub)population that results in an impairment of functional capacity, an impairment of the capacity to compensate for additional stress or an increase in susceptibility to other influences*'.

The definition of adversity is generic and not specific to the endocrine assessment and current practices are applicable for deciding whether the observed effects are treatment-related and should be considered adverse. A substance with ED properties will in most of the cases display a pattern of effects. Therefore, for the scope of this guidance, effects related to all parameters labelled as 'EATS-mediated' and/or 'sensitive to, but not diagnostic of, EATS' should be considered together when assembling the lines of evidence for adverse effects. Level 3 tests using intact (immature) animals might also provide additional evidence of adverse effects.

The definition of adversity is generic and not specific to the endocrine assessment and current practices are applicable for deciding whether the observed effects are treatment-related and should be considered adverse.

A line of evidence for adversity may be based on a single parameter (e.g. histopathological findings in the testis observed in one or more studies; decrease in fecundity of fish observed in one or more studies), however, it should be highlighted that some individual parameters may not be considered adverse in isolation. In such cases, the conclusion on adversity relies on a combination of parameters following the integration of the different lines of evidence as explained in Section 3.3.3. Therefore, it requires expert judgement to assemble and assess the lines of evidence for adversity.

For non-target organisms, separate lines of evidence for adversity could be assembled for the different species/taxa. In particular, data on fish could be used for assembling lines of evidence for E, A and S modalities while data on amphibians could be used for assembling lines of evidence for the T modality. In some cases, data on amphibians may also inform about the E, A and S modalities. The lines of evidence for adversity on non-target organisms could be built by considering either the reproduction (e.g. fertility, fecundity, etc.) in the case of E, A and S modalities and/or the development/growth (hindlimb length, developmental stage, time to metamorphosis, thyroid histopathology) for the T modality. Data on other taxa (e.g. birds) can, on a case-by-case basis, be considered as complementary information.

The assessment of the lines of evidence should be based on the available empirical support and expert judgement. The empirical support consists of dose-response, temporal concordance, consistency among studies and species and repeatability for the line of evidence. Expert judgement could be necessary when assessing the available lines of evidence, including the overall evaluation of the consistency of the data set as a whole.

In the case of the lines of evidence for adversity related to non-target organisms, the empirical support will be mainly based on the evaluation of the dose-response relationship due to the available data set not often allowing for the evaluation of the temporal concordance and consistency among species (often only studies on a single species are available).

For both human health and non-target organisms, lack of a proper dose response or consistency between species and studies should not imply that the empirical support is judged as insufficient as long as this can be explained, for example by the lack of a proper dose spacing and/or differences in study designs.

When assembling and assessing the line of evidence, any available epidemiological studies should be considered as supportive evidence for the evaluation of whether an ED is likely to have adverse effects for humans. However, they cannot be used to override or dismiss evidence of adversity found in laboratory studies, nor can they replace laboratory studies.

Similarly, when assembling the lines of evidence for non-target organisms any field and monitoring studies and population modelling can be considered as supportive evidence.

3.3.1.1. Effects secondary to other toxicities

According to the criteria, *adverse effects that are non-specific secondary consequences of other toxic effects shall not be considered for the identification of the substance as endocrine disruptor*.

In principle, the top dose/concentration selected for the conduction of the (eco)toxicological studies should provide information on substance toxicity at an exposure of the tested agent that should be tolerated without inducing significant chronic physiological dysfunctions, be compatible with animal survival and permits data interpretation in the context of the use of the study. The concepts of maximum tolerated dose (MTD) and maximum tolerated concentration (MTC) are then useful for top dose/concentration selection and should be considered as a starting point for the evaluation of changes which could be due to excessive systemic toxicity.

The aim of the MTD is to produce a minimum toxic effect over the course of the study. Elements to consider are alterations in physiological function, including: no more than 10% decrease in body weight gain relative to control, target organ toxicity and alterations in clinical pathological parameters. Although these parameters can only be considered indicative and expert judgement is necessary to define the MTD on a case-by-case basis. Elements which indicate that the MTD has been exceeded are reported in the OECD Guidance on the Recognition, Assessment and Use of Clinical Signs as Humane Endpoints for Experimental Animals Used in Safety Evaluation (OECD, 2000).

Equally, in ecotoxicology, the MTC is defined as the highest test concentration of the chemical which results in less than 10% mortality (Hutchinson et al., 2009; Wheeler et al., 2013; Ankley and Jensen, 2014). For tests on aquatic organisms, the maximum solubility in water, or the limit concentration as defined in the relevant OECD guidelines should be considered.

All these elements should be also evaluated when considerations are made on concluding that endocrine mediated adverse effects are consequent of excessive systemic toxicity.

These elements should, however, be judged in the context of the dose response relationship and of their severity (i.e. should not be so severe that physiological functions are compromised).

There are two situations foreseen where adverse effects may be non-specific secondary consequences of other toxicities:

1) Where potentially endocrine-related adverse effects are only observed at excessive toxic dose/concentration (i.e. only observed above the MTD or MTC) they should not be considered indicative of endocrine disruption. Justification of this excessive toxicity should be provided. However, some specific considerations should be made when dealing with effects that are indeed also observable following endocrine imbalances (e.g. changes in body weight consequent to mimetic activity for testosterone and oestradiol (Andersson et al., 2005; Hotchkiss et al., 2007)).
2) In other situations, potentially endocrine-related adverse ffects observed at, or below the MTD or MTC, can be considered as secondary to other (non-endocrine) toxicities only if substantiated by the MoA analysis.

3.3.1.2. Low-dose effects and non-monotonic dose response (NMDR)

It is acknowledged that there is a lack of consensus in the scientific community with regard to the existence and/or relevance of low-dose effects and NMDR curves in (eco)toxicology in connection with endocrine disruption (EFSA Scientific Committee, 2013). However, some evidence is available from experimental data for such NMDR (Beausoleil et al., 2013). Although NMDR should not by default be dismissed as not supporting the assessment for hazard identification, in most of the cases the design of standard *in vivo* toxicity studies (mainly because of the limited number of doses) does not allow concluding on the presence of a NMDR. Evidence of non-monotonicity in *in vitro* studies (where many concentrations can be tested) could provide additional information relevant for supporting the biological plausibility of an endocrine MoA where endocrine-related adversity is observed. Furthermore, it should be noted that standard toxicity studies are designed to identify hazard (i.e. the adverse effect), and therefore the likelihood of not detecting an adverse effect in the presence of a NMDR is considered low.

3.3.1.3. Human relevance

According to the scientific criteria for determining ED properties applicable to the BP and PPP Regulations, '*A substance shall be considered as having endocrine-disrupting properties that may cause adverse effect in humans [...] unless there is evidence demonstrating that the adverse effects identified are not relevant to humans*'.

Note that the assessment of human relevance does not refer to adversity as such, but rather to the question as to whether an effect elicited by a substance in a test animal could also be elicited in a human being. Therefore, to disprove human relevance it is necessary to demonstrate differences in the mechanisms of action of the substance in human and in test animals by having recourse to the MoAs. Therefore, human relevance is addressed in the context of MoA analysis (Section 3.5.4.4).

3.3.1.4. Population relevance

According to the scientific criteria set in Commission Delegated Regulation (EU) No 2017/2100[3] and in Commission Regulation (EU) No 2018/605[4], for determining ED properties applicable to the BP and PPP Regulations, '*A substance shall be considered as having endocrine-disrupting properties that may cause adverse effects on non-target organisms [...] unless there is evidence demonstrating that the adverse effects identified are not relevant at the (sub)population level for non-target organisms*'. The criteria also stipulate that, in applying the WoE approach, the assessment of the scientific evidence shall consider *the adverse effects on reproduction, growth/development, and other relevant adverse effects which are likely to impact on (sub)populations*.

Effects on growth, development, reproduction in single species are generally regarded relevant for the maintenance of the wild population (European Commission, 2011). Therefore, the relevance of such effects at the population level should be assumed when determining the adversity in the absence of appropriate scientific data demonstrating non-relevance. Behavioural changes and impaired ability to cope with additional stress are factors implicitly covered by the WHO definition of adversity, since they would affect the reproductive performance and the development. Therefore, behavioural changes or impaired ability to cope with additional stressors which have the potential to impact the population stability of non-target organisms would be considered in the definition of adversity. It is acknowledged however, that current standard tests are not specifically designed to specifically capture all behavioural effects (European Commission, 2018).

According to the assessment strategy proposed by this guidance, further consideration is needed to evaluate whether some effects observed in mammals can be considered adverse for mammals as non-target organisms. For example, thyroid histopathological findings observed in the rat are likely not relevant at population level if observed in isolation without impairment of growth/development and/or reproduction. Therefore, in order to reach a conclusion, it may be needed to reconsider the mammalian data package. Similarly, in the case of amphibians, changes in thyroid histopathology should be considered adverse at the population level only when observed together with effects on development (i.e. delay or acceleration). This is due to the fact that thyroid histopathology often represents compensation to thyroid insufficiency (Marty et al., 2017). Nevertheless, changes in development in amphibians even if observed in the absence of investigation of thyroid histopathology are considered population relevant effects. However, the degree of delay or acceleration in the development that can be considered adverse at population level is uncertain (Marty et al., 2017). Therefore, such effects should be considered relevant at the population level unless available

information demonstrates the contrary. According to the strategy described in this guidance, the population relevance of the observed adverse effect should be assessed taxon by taxon.

3.3.2. Assembling and assessing the line(s) of evidence for endocrine activity

Parameters labelled as '*in vitro* mechanistic' or '*in vivo* mechanistic', should be considered when assembling lines of evidence for endocrine activity. As indicated above, 'EATS-mediated' parameters due to the nature of the effect and the existing knowledge also provide *in vivo* mechanistic information for at least one EATS modality. The lines of evidence for endocrine activity should be organised by modality.

Similarly to the evidence for adversity, the evidence for endocrine activity is evaluated on the basis of the empirical support and expert judgement. The empirical support consists of dose/concentration–response, consistency among studies and repeatability for the line of evidence.

3.3.3. Integration of the lines of evidence for adverse effects and endocrine activity

Once assembled, the available lines of evidence should be integrated for the assessment of adversity and endocrine activity for each modality. The lines of evidence should be reported in a tabular format. Additional information, e.g. on systemic general toxicity or other target organ effects, may be used at this point, on a case-by-case basis, in order to contextualise the presence or absence of an adverse effect potentially linked to an endocrine activity. The assessment of the integrated lines of evidence should allow an evaluation of whether the data set is sufficient to support robust conclusion on adversity and/or endocrine activity (see Section 3.4).

Two examples, including assembling, assessing and integrating lines of evidence, are reported in Tables 2 and 3.

In the example in Table 2, for endocrine activity the evidence comes from three different sources: an *in silico* prediction, hormonal measurements in repeated dose toxicity studies and a mechanistic *in vivo* study with amphibians. The lines of evidence related to decrease in thyroid hormonal levels and inhibition of iodine transport are suggested to be integrated as they are both informative of T modality. For EATS-related adversity, the evidence comes from histopathological findings in repeated dose toxicity studies and a field study with reptiles. The repeated dose toxicity studies are also used to establish lines of evidence for general systemic toxicity. The lines of evidence related to increase in follicular cell hyperplasia and thyroid weight in different species can be integrated as they both inform on adversity through T modality.

In the example in Table 3, for endocrine activity the evidence comes from: mechanistic *in vitro* studies for EAS modalities, hormonal and biomarker measurements from *in vivo* mechanistic data. In addition, effects on gonad histopathology (EATS mediated) as well as effects on fecundity (sensitive to but not diagnostic of EATS parameters) are considered for the definition of adversity. The *in vivo* evidence is derived from level 3 and 5 studies (i.e. fish short-term reproduction assay and fish life cycle toxicity test (FLCTT)). In the FLCTT, evidence of general toxicity (liver histopathology) was also reported. Lines of evidence related to changes in specific female gonad histopathology and decrease in fecundity in fish are integrated because they inform on adversity through S modality. Aromatase inhibition, decrease in oestradiol and VTG levels in female fish are integrated since they are all lines of evidence for S modality.

3.3.4. Reporting the lines of evidence

The lines of evidence should be reported in a tabular format as exemplified in Tables 2 and 3. More specifically, the lines of evidence should be reported and organised according to their contribution to the assessment. Indication of general systemic toxicity should also be reported to allow the assessment of potential secondary effects as described in Section 3.3.1. The reliability of each study is useful information which should be reported. In the examples, the available information was assembled into lines of evidence depending on whether the parameters contribute with information on endocrine activity and/or EATS-related adversity. As shown in the examples, details such as the species tested, exposure duration and route of exposure, and doses/concentration should be provided for each piece of evidence together with the observed effects and the likely endocrine modality.

Table 2: Example showing how to assemble, integrate and assess the lines of evidence for thyroid disruption in mammals

	Grouping	Line(s) of evidence	Species	Exposure weeks	Route of exposure	Effect dose mg/kg/day	Observed effects (positive and negative)	Assessment of each line of evidence	Assessment of the integrated line of evidence	Modality
Integrated line of evidence for endocrine activity	*In silico* prediction	(Q)SAR prediction					Predicted to Inhibit iodine transport	Supporting evidence	Overall positive evidence for endocrine activity	Thyroid
	In vivo mechanistic	Hormonal changes T3, T4	Dog	26	Oral	13	Dose-dependent decrease	Hormone changes observed in three species in a dose-related manner		Thyroid
			Hamster	78	Oral	15	No effect; highest dose tested 15			
			Rat	4	Oral	5	Dose-dependent decrease			
			Rat	4	Inhalation	0.32	Dose-dependent decrease			
			Rabbit	2	Oral	75	Dose-dependent decrease			
Integrated line of evidence for adversity	EATS-mediated parameter	Hindlimb length	Frog	3	Water	1.75 mg/L	Dose-dependent decrease	Sufficient: changes observed in a dose-dependent manner	Overall positive evidence for adversity	Thyroid
		Thyroid (histopathology)	Frog		Water	1.75 mg/L	Dose-dependent increase	Sufficient: changes observed in a dose-dependent manner		
		Thyroid (histopathology)	Lizard	4	Intraperitoneal injections	5	Changes in epithelium height of the follicular cells at all the tested doses	Supportive (non-standard species and study design) evidence of changes in histopathology in a dose-dependent manner		

	Grouping	Line(s) of evidence	Species	Exposure weeks	Route of exposure	Effect dose mg/kg/day	Observed effects (positive and negative)	Assessment of each line of evidence	Assessment of the integrated line of evidence	Modality
		Thyroid (histopathology)	Dog	26	Oral	13	Follicular cell hyperplasia; dose-dependent increase	Sufficient: observed in 2 species in a dose related manner		
			Hamster	78	Oral	15	No effect; highest dose tested 15			
			Rat	4	Inhalation	0.32	Follicular cell hyperplasia; dose-dependent increase			
			Rat	13	Oral	10	Colloid and capillary density; dose-dependent increase			
			Rat	104	Oral	5	Follicular cyst/ follicular cell adenoma and adenocarcinoma; dose-dependent increase			
			Rat	2 generation	Oral	1.64	Follicular cell hyperplasia; dose-dependent increase; at the top dose follicular cells hyperplasia/ adenoma			
		Thyroid(organ weight)	Dog	26	Oral	13	Dose-dependent increase	Sufficient: observed in 2 species in a dose-dependent manner		
			Mouse	78	Oral	15	Dose-dependent increase			
			Rat	4	Inhalation	0.32	Dose-dependent increase			
			Rat	104	Oral	5	Dose-dependent increase			

	Grouping	Line(s) of evidence	Species	Exposure weeks	Route of exposure	Effect dose mg/kg/day	Observed effects (positive and negative)	Assessment of each line of evidence	Assessment of the integrated line of evidence	Modality
	Parameter sensitive to, but not diagnostic of, EATS	Pituitary (histopathology)	Dog	26	Oral	36	Vacuolisation of pale cells	Sufficient: observed in 3 species in a dose-related manner		
			Mouse	78	Oral	15	Hyperaemia; dose-dependent increase			
			Rat	104	Oral	5	Adenoma			
			Rat	2 generation	Oral	15.64	Vacuolated cells			
Evidence of general toxicity		Body weight	Dog	26	Oral	36	Decrease (5%)	Minor effects in body weight in the high-dose groups		
			Hamster	78	Oral	15	No effect; highest dose tested 15			
			Rat	4	Inhalation	0.66	No effect; highest dose tested 0.66			
			Rat	13	Oral	13	Dose-dependent decrease 10% at highest does 30			
			Rat	104	Oral	5	No effect			
			Rat	2 generation	Oral	3	No effect			
			Mouse	78	Oral	15	Dose-dependent decrease 10% at highest does 45			
		Liver weight (relative)	Dog	26	Oral	36	Increase 5%	Minor effects in relative liver weight in the high-dose groups		
			Hamster	78	Oral	15	No effect; highest dose tested 15			
			Rat	4	Inhalation	0.66	No effect			
			Rat	13	Oral	30	Increase 7%			
			Rat	104	Oral	5	No effect			
			Rat	2 generation	Oral	3	No effect			
			Mouse	78	Oral	45	Increase 10%			

	Grouping	Line(s) of evidence	Species	Exposure weeks	Route of exposure	Effect dose mg/kg/day	Observed effects (positive and negative)	Assessment of each line of evidence	Assessment of the integrated line of evidence	Modality
		Kidney weight (relative)	Dog	26	Oral	36	No effect	No indication of kidney toxicity		
			Hamster	78	Oral	15	No effect; highest dose tested 15			
			Rat	4	Inhalation	0.66	No effect			
			Rat	13	Oral	30	No effect			
			Rat	104	Oral	5	No effect			
			Rat	2 generation	Oral	3	No effect			
			Mouse	78	Oral	45	No effect			

Table 3: Example showing how to assemble, integrate and assess the lines of evidence for aromatase inhibition leading to reproductive dysfunction in fish

	Grouping	Line(s) of evidence	Species/cell line(s)	Exposure (weeks)	Route of exposure	Effect dose (mg/L)	Observed effects (positive and negative)	Assessment of each line of evidence	Assessment of the integrated line of evidence	Modality
Integrated line of evidence for endocrine activity	*In vitro* mechanistic data	Aromatase activity	Human placental microsomes CYP19				Inhibition of CYP19 activity	Sufficient	Overall positive evidence for endocrine activity	S
			H295R				Inhibition of CYP19			
			Recombinant human microsomes (2)				Inhibition			
			Human placental microsomes				Inhibition			
			JEG-3 (2)				Positive after 2 h incubation. No effect after 24 h incubation. No effect on aromatase expression. Weak activation at lower concentration. Apparent inhibition at higher concentration			
			Yeast and human CYP51				Inhibition			
			Recombinant zebrafish CYP51				CYP51 binding			

	Grouping	Line(s) of evidence	Species/cell line(s)	Exposure (weeks)	Route of exposure	Effect dose (mg/L)	Observed effects (positive and negative)	Assessment of each line of evidence	Assessment of the integrated line of evidence	Modality
	In vivo mechanistic	Hormonal changes: oestradiol	*Pimephales promelas*	3	Water	0.5	Dose-dependent decrease	Sufficient: Oestradiol decrease observed in a dose-related manner but measured in one study only		S
		Vitellogenin (VTG) in females	*Pimephales promelas*	3	Water	1	Decrease only at the highest dose (large dose spacing; the previous dose is 0.12)	Sufficient: Dose-related changes in VTG. When the dose dependence could not be demonstrated, this is considered to be due to the test design (dose spacing and tested doses)		
			Pimephales promelas	3	Water	0.5	Dose-dependent decrease			
			Pimephales promelas	36	Water	0.558	Decrease only at the highest dose			
Integrated line of Evidence for adversity	EATS-mediated parameters	Histology: Specific female gonad histopathology	*Pimephales promelas*	36	Water	0.558	Only at the highest dose (decreased yolk formation; decreased post ovulatory follicles; decreased mean ovarian stages scores)	Supportive evidence. The parameter was only measured in one study	Overall positive evidence for adversity	S
	Sensitive to, but not diagnostic of EATS	Fecundity	*Pimephales promelas*	3	Water	1	Decrease only at the highest dose	Sufficient. Dose-related decrease in fecundity. When the dose dependence could not be demonstrated this is considered to be due to the test design (dose spacing and tested doses)		S
			Pimephales promelas	3	Water	0.5	Dose-dependent decrease			
			Pimephales promelas	36	Water	0.558	Decrease only at the highest dose			

	Grouping	Line(s) of evidence	Species/cell line(s)	Exposure (weeks)	Route of exposure	Effect dose (mg/L)	Observed effects (positive and negative)	Assessment of each line of evidence	Assessment of the integrated line of evidence	Modality
Evidence of general toxicity		Liver histopathology	*Pimephales promelas*	36	Water	0.558	Increase nuclear pleomorphism, multi-nucleation, cystic degeneration, necrosis, pigmented macrophages, aggregates and anisocytosis in hepatocytesof males and females	Effects on liver were only investigated in one study and only observed at the highest tested dose		

3.4. Initial analysis of the evidence

Once all relevant information has been gathered, evaluated and assembled into lines of evidence using a WoE approach as explained in Section 3.2, an analysis of the data set with respect to indication of EATS-mediated adversity or EATS-mediated endocrine activity has to be carried out.

The initial analysis of the evidence comprises an assessment whether either EATS-mediated adversity or EATS endocrine activity has been 'sufficiently' investigated (see Sections 3.4.1 and 3.4.2). This will allow to stop the ED assessment in case no EATS-mediated adversity or endocrine activity have been observed or to decide whether further data need to be generated (see Section 3.4.3). As explained in the assessment strategy (see Section 3), the 'EATS-mediated' parameters listed in the OECD GD 150 drive the assessment because by providing evidence for both endocrine activity and the resulting adverse effects, they are considered indicative of an endocrine MoA.

This initial analysis is not relevant for non-EATS modalities. Instead, if there is indication of non-EATS mediated endocrine activity or adversity, this should be directly taken forward to the MoA analysis (see Section 3.5).

In the following two sections, there is a description of what is considered a sufficient data set to support the conclusion that the ED criteria are not met on the basis of absence of EATS-mediated adversity and EATS endocrine activity.

3.4.1. Sufficient data set for EATS-mediated adversity to support a conclusion on absence of EATS-mediated adversity

Based on the current knowledge and available test guidelines, to consider the EATS-mediated adversity sufficiently investigated with respect to humans and mammals (as non-target organisms), the information described below needs to be available in order to support a conclusion on absence of EATS-mediated adversity.

To have the EAS-mediated adversity with regard to humans and mammals (as non-target organisms) sufficiently investigated, all the data requirements of the specific Regulations, must be fulfilled. This should include all the 'EAS-mediated' parameters foreseen to be investigated in an extended one-generation reproductive toxicity study (EOGRTS; OECD TG 443; with cohort 1a/1b including the mating of cohort 1b to produce the F2 generation (OECD, 2012b)) or a two-generation reproductive toxicity study (OECD TG 416; test protocol according to latest version of January 2001 (OECD, 2001)) (see also Table 14 in Section 4). To have the EAS-mediated adversity for other non-target organisms sufficiently investigated, the 'EAS-mediated' parameters foreseen to be measured in the Medaka extended one-generation test (MEOGRT, OECD TG 240 (OECD, 2015c)) should have been investigated and the results included in the dossier. Alternatively, a FLCTT covering all the 'EAS-mediated' parameters foreseen to be measured in the MEOGRT is acceptable (see also Table 15 in Section 4).

To have the T-mediated adversity with regard to humans and mammals (as non-target organisms) sufficiently investigated, the thyroid parameters foreseen to be investigated in the following studies OECD test guidelines 407, 408, 409 (and/or the one-year dog study, if available), 416 (or 443 if available) and 451-3 should have been measured and the results included in the dossier. If there is no indication of effects in these studies, the T modality is considered to be sufficiently covered. However, if any thyroid effect is observed additional guidance on how to proceed is provided in Appendix A.

In principle, to have the T-mediated adversity with regard to other non-target organisms sufficiently investigated the results from all the 'T-mediated' parameters foreseen to be investigated in the Larval amphibian growth and development assay (LAGDA; OECD TG 241 (OECD, 2015d)) would be needed. However, if the T-mediated parameters foreseen to be investigated in an amphibian metamorphosis assay (AMA, OECD TG 231 (OECD, 2009c)) are negative, this would be sufficient to support that T-mediated adversity is unlikely because no T-related endocrine activity has been observed (see also Table 16 in Section 4).

It has to be noted, that the determination of adversity shall be based on a WoE approach taking into account all the available information in the dossier. This means that the studies abovementioned should not be considered in isolation. The approach described in Section 3.2 on assembling and assessing the lines of evidence should be followed.

3.4.2. Sufficient data set for EATS-related endocrine activity to support a conclusion on absence of EATS-related endocrine activity

According to the assessment strategy (see Figure 1), following the initial analysis of the evidence, if 'EATS-mediated' adversity has not been sufficiently investigated and no 'EATS-mediated' adversity has been observed – then EATS-related endocrine activity should be further considered. Based on the current knowledge and available test guidelines, to consider the EATS-related endocrine activity sufficiently investigated with respect to humans and mammals (as non-target organisms), the information described below needs to be available in order to support a conclusion on absence of EATS-related endocrine activity.

E-modality – The output data from the ToxCast ER Bioactivity Model or 'Uterotrophic bioassay in rodents' (OECD TG 440) (OECD, 2007d).

A-modality – 'Hershberger bioassay in rats' (OECD TG 441) (OECD, 2009d).

T-modality – *In vitro* mechanistic test guidelines for the T modality are currently not available as well as specific *in vivo* mechanistic tests on mammals. Hence, to consider the T modality as 'sufficiently investigated' for mammals the thyroid parameters foreseen to be investigated in the following studies OECD test guidelines 407, 408, 409 (and/or the one-year dog study, if available), 416 (or 443 if available) and 451-3 should have been measured and the results included in the dossier (see Section 3.4.1).

S-modality – The level 2 *in vitro* assays 'H295R steroidogenesis assay' OECD TG 456 (OECD, 2011c) and the 'aromatase assay (human recombinant)' OPPTS 890.1200 (US EPA, 2009b). There are currently no level 3 tests that fully cover this modality, however it is partially covered by OECD TG 441. Therefore, the results of the above *in vitro* assays should be considered together with the results of the E and A modalities in order to conclude on the absence of endocrine activity for the S modality.

To consider the **E, A, S modalities** for **non-target organisms other than mammals** sufficiently investigated, preferably the 'Fish short term reproduction assay' (FSTRA; OECD TG 229) should have been conducted; however the 21-day fish assay OECD TG 230 (OECD, 2009b) is acceptable as well. If data are already available covering the mechanistic parameters investigated in OECD TG 229 or OECD TG 230 (e.g. OECD TG 234), then those data could be used instead.

To consider the **T-modality** sufficiently investigated, an 'Amphibian metamorphosis assay' (AMA; OECD TG 231 (OECD, 2009c)) should have been conducted.

For further considerations on endocrine activity, see Section 3.4.3 and Figure 3.

3.4.3. Considerations on the generation of further data

When further information needs to be generated to enable a conclusion on the ED criteria, the applicant should determine the information needed to clarify the concern and thereafter agree this with the risk assessors. This is particularly important if additional vertebrate testing is considered necessary, because in accordance with Directive 2010/63/EU[10], Regulation (EU) No 528/20121 (Recitals 57 and 59, Articles 2.3(p), 37.4 and 62.1, Annex II, para 6, and Annex III, para 6), Regulation (EC) No 1107/20092 (Recitals 11 and 40, Articles 8.1(d), 18(b), 33.3(c) and 62.1) and Regulation (EU) No 2013/2838, unnecessary animal testing should be avoided. In addition, agreeing with the risk assessors on what type of information is needed to clarify the concern may avoid additional data requests later in the process, and thus facilitate decision making.

It should be noted that the generation of a sufficient data set including the studies mentioned in Sections 3.4.1 and 3.4.2 might be considered by the applicants when a new dossier has to be prepared. When a dossier is already available but the data set is not sufficient, it is recommended to first consider the existing data (see scenarios 2a (i) and 2b under Section 3.4.4.2) before performing additional vertebrate tests. This is because the available information may still be enough to support a conclusion that the ED criteria are met, and therefore additional vertebrate testing might not be needed.

It is noted that further investigation of the endocrine activity is always required when no adversity based on EATS-mediated parameters is observed on the basis of an insufficient data set. Furthermore, by following the assessment strategy, generation of further data (e.g. additional tests to investigate adversity) can be triggered by the MoA analysis on a case-by-case basis and depending on the information already available.

[10] Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the protection of animals used for scientific purposes. OJ L 276, 20.10.2010, p. 33–79. Available online: http://data.europa.eu/eli/dir/2010/63/oj

To investigate the endocrine activity, a stepwise approach should be followed by using the tests listed in Table 4 and by following the strategy outlined in Figure 3 considering the existing information. The strategy suggests conducting (or using equivalent information from the literature) the level 2 tests listed in Table 4 in order to reduce animal testing and facilitate the MoA analysis under Section 3.5.[11] When, following the strategy in Figure 3, the *in vitro* tests are positive, this could be sufficient to go to the MoA analysis (see Section 3.5). However, if *in vitro* tests are negative this is not usually sufficient to demonstrate lack of endocrine activity *in vivo* due to the complexity of the endocrine system and the current limitations of the *in vitro* assays.

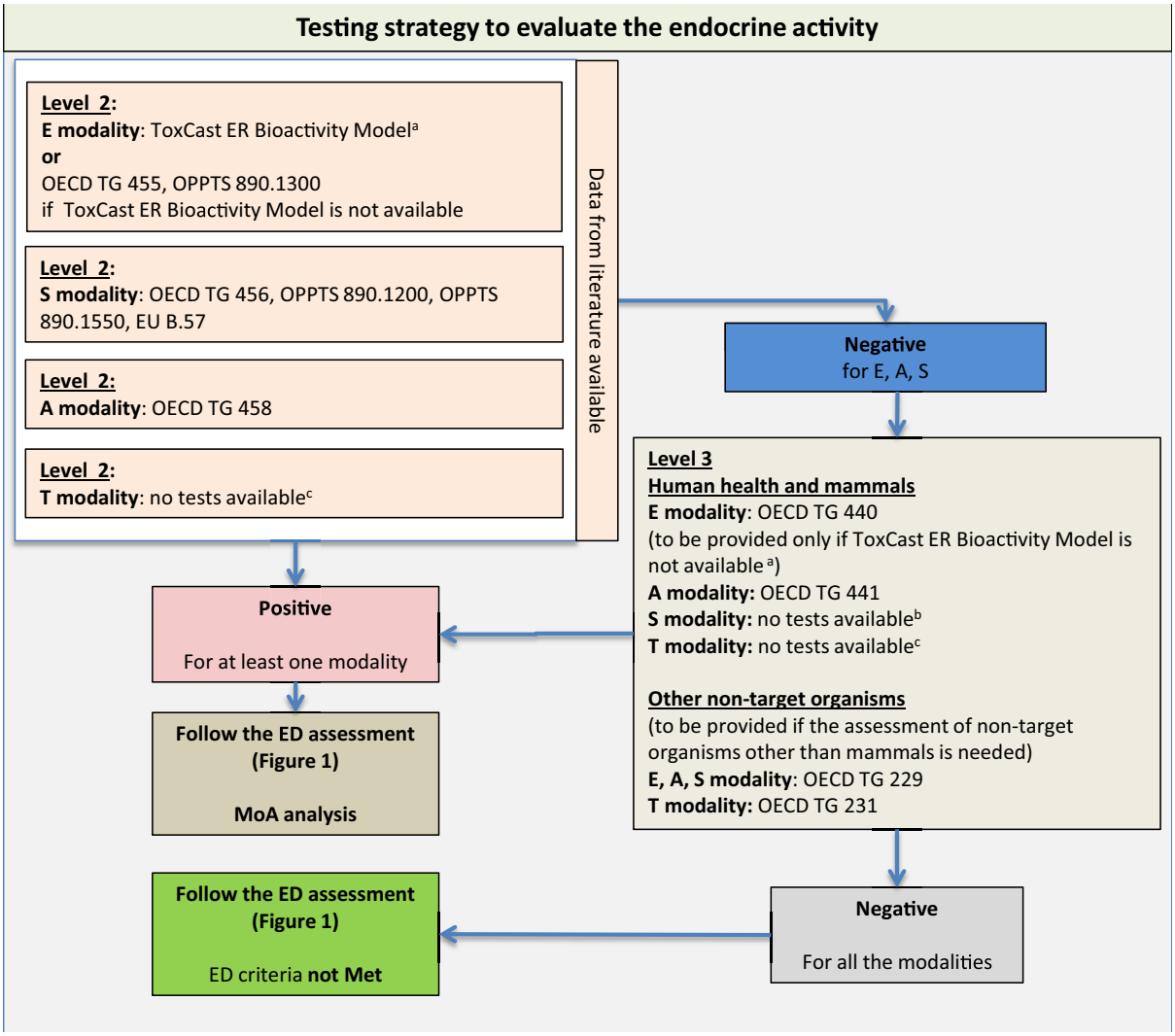
When interpreting the results of *in vitro* tests, the lack of a metabolic system, as well as the lack of consideration of other adsorption, distribution, metabolism, excretion (ADME) properties, should be considered. In part, this is because *in vitro* systems currently consist of (a monolayer of) one cell type that focuses on a specific pathway. In general, the *in vitro* tests lack the complexity of an intact organism. In particular, considerations of ADME properties are not covered by current test guidelines. To partly overcome this limitation, several *in vitro* tests can be run by adding (part of the) metabolising systems, potentially metabolising the parent compound into an active, less active or inactive substance/metabolite. Activities on including a metabolising step are currently on the OECD test guideline work programme (OECD, 2017e).

As described in Section 4, most current *in vitro* assays focus on nuclear hormone receptors, not all ED effects are mediated through a direct action on these receptors. However, as compounds might be able to act via more than one mechanism, no single *in vitro* test can be expected to detect all types of endocrine disruption: the eventual ED effect *in vivo* might be a consequence of disturbance of several pathways simultaneously, some of which might not be covered by our current *in vitro* testing strategy. Because of this, and because of the inherent limitations of *in vitro* systems, conclusions can only be drawn in the context of what the *in vitro* assay evaluates and a negative *in vitro* result alone cannot be used to exclude possible endocrine-disrupting activity on the endocrine modality under investigation.

Table 4: Recommended tests methods to investigated EATS-related endocrine activity

Pathway	OECD CF Level	Assay family	OECD guideline	Other guidelines	Other sources of data
Estrogen	Level 2	Transactivation assay	OECD TG 455	OPPTS 890.1300	ToxCast ER Bioactivity Model
	Level 3		OECD TG 440		
			OECD TG 229		
Androgen	Level 2	Transactivation assay	OECD TG 458		
	Level 3		OECD TG 441		
	Level 3		OECD TG 229		
Steroidogenesis	Level 2	Steroidogenesis	OECD TG 456	OPPTS 890.1550; EU B.57	
	Level 3	CYP19	OECD TG 229	OPPTS 890.1200	
Thyroid	Level 3		OECD TG 231		

[11] For the E and A modalities, it may not be necessary to conduct the level 2 studies if the corresponding level 3 test(s) are available (see Section 3.4.2).



Note a: The ToxCast Estrogen Receptor (ER) model integrates multiple *in vitro* assays, high-throughput ToxCast screenings assays measuring receptor (ER) binding, dimerisation, chromatin binding, transcriptional activation and ER-dependent cell proliferation (Judson et al., 2015). The multiple *in vitro* assays provide comprehensive pathway coverage for the biology of the ER signalling pathway (Browne et al., 2015). US EPA is accepting ToxCast ER model for 1812 chemicals as alternatives for EDSP tier 1 ER binding, ER transactivation, and uterotrophic assays.
Note b: Partially covered by the OECD TG 441. However, it is recommended to always investigate the S modality with the E and A modalities and not to interpret negative results in isolation.
Note c: Level 2 tests and, for mammals, specific level 3 tests are not yet available. The T modality is included in this figure for completeness. To consider the T modality as 'sufficiently investigated' for human health and mammals the thyroid parameters from OECD test guidelines 407, 408, 409 (and/or the one-year dog study, if available), 416 (or 443 if available) and 451-3 should have been measured (see Section 3.4.1).

Figure 3: Strategy to investigate EATS-related endocrine activity in the context of the ED assessment (see also Figure 1)

3.4.4. Scenarios

In this section, different scenarios are described providing guidance on how to proceed with the assessment, depending on the information available on EATS-mediated adversity and endocrine activity (see Table 5 for a summary of the scenarios). The decision tree included in the initial analysis of the evidence, as illustrated in the flow chart presented in Section 3.1 and also reported in the zoom-in of the flow chart below (Figure 4), is detailed in the following sections.

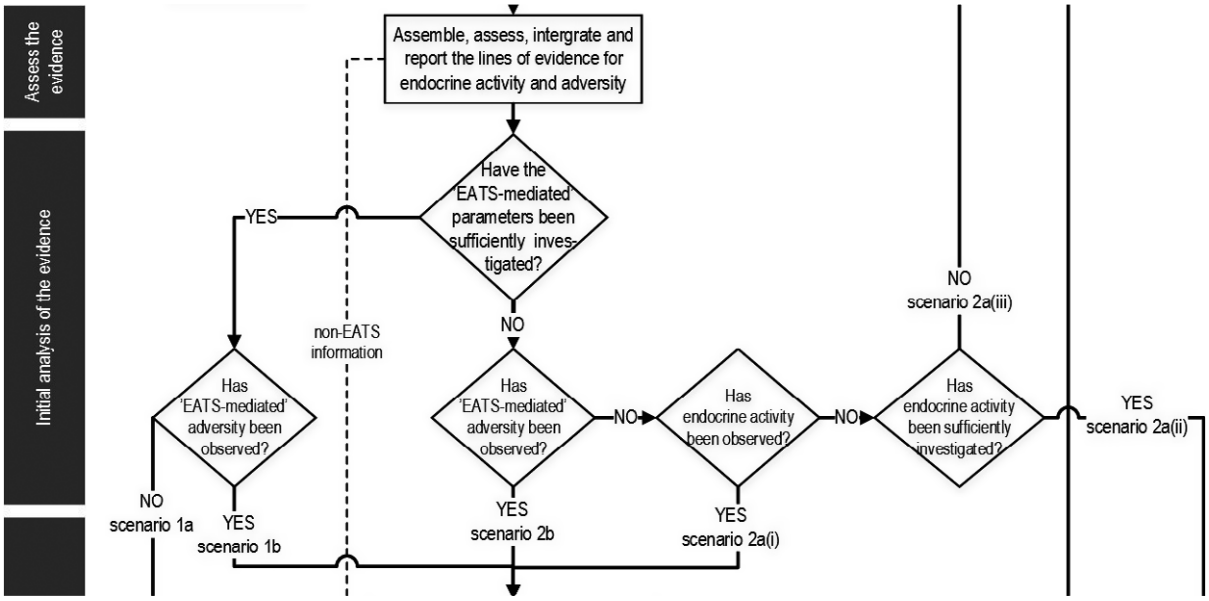


Figure 4: Zoom in on the initial analysis of the evidence from the flow chart in Figure 1

3.4.4.1. Scenarios based on 'EATS-mediated' parameters sufficiently investigated

This section is meant to cover the situations where the answer to the question in Figure 1 and its zoom-in shown in Figure 4 'Have all "EATS-mediated" parameters been investigated?' is YES.

These scenarios cover the cases where the 'EATS-mediated' parameters have been sufficiently investigated as explained in Section 3.4.1 with regard to humans and non-target organisms.

Two scenarios can be foreseen as explained below.

Scenario 1a – No adversity indicated by 'EATS-mediated' parameters

When there is an overall indication of no adversity based on EATS-mediated parameters, the first condition of the ED criteria is not met; therefore, it is possible to conclude that **the substance does not meet the ED criteria**.

The approach taken to reach this conclusion must be transparently documented in the dossier (see Section 3.6).

Scenario 1b – Adversity indicated by 'EATS-mediated' parameters

When adversity is observed based on 'EATS-mediated' parameters (see Table 2 for an example), the biological plausibility of the link between the 'EATS-mediated' adversity and endocrine activity should be documented through a MoA analysis (see Section 3.5.1 for further details).

3.4.4.2. Scenarios based on 'EATS-mediated' parameters not sufficiently investigated

This section is meant to cover the situations where the answer to the question in Figure 1 and its zoom-in shown in Figure 4 'Have the "EATS-mediated" parameters been sufficiently investigated?' is NO.

These scenarios cover the cases where the data set does not match the descriptions of a sufficient data set in Section 3.4.1. This is the most common situation for existing dossiers where for example, studies according to the OECD TG 416 (outdated version) or only a study according to OECD TG 210 (ELS study on fish) are available.

Two scenarios can be foreseen, depending on whether adversity is indicated by the 'EATS-mediated' parameters that have been investigated.

Scenario 2a – No adversity indicated by the 'EATS-mediated' parameters

When no EATS-mediated adversity is observed or only 'sensitive to, but not diagnostic of, EATS' parameters are available (either indicating or not indicating adversity), the analysis should proceed with the endocrine activity, and, as a starting point, it should be considered if information is already available and if it is sufficient as explained in Section 3.4.2. Generation of further data may be needed as described in Table 4 and in Figure 3.

Three subscenarios can be distinguished in this case, depending on whether endocrine activity has been observed, or not observed, or not sufficiently investigated.

i) Endocrine activity observed

If the available mechanistic information gives indication of endocrine activity, for at least one of the modalities, this would be sufficient as a starting point for a MoA analysis which is required to establish the biological plausibility of the link between the observed endocrine activity and potential adverse effect for the postulated MoA(s) (see Section 3.5). As all 'EATS-mediated' parameters have not been investigated, additional information e.g. from level 3, 4 or 5 studies may need to be generated.

ii) No endocrine activity observed, but sufficiently investigated

If the available mechanistic information (see Table 4 and Figure 3) does not give indication of endocrine activity (i.e. all the E, A, S, T modalities are negative), and the data set is sufficient as explained in Section 3.4.2 then **the substance does not meet the ED criteria**.

The approach taken to reach this conclusion must be transparently documented in the dossier.

iii) No endocrine activity, but not sufficiently investigated

If the endocrine activity has not been sufficiently investigated (see Section 3.4.2), it is necessary to generate further information (see Table 4 and Figure 3). Alternatively, applicants may consider complementing the information available on 'EATS-mediated' adversity, e.g. by carrying out level 5 studies, as described in Section 3.4.1. Depending on the outcome of these further investigations, the assessment needs to be continued following the corresponding scenario.

Scenario 2b – Adversity indicated by 'EATS-mediated' parameters

When adversity is observed based on 'EATS-mediated' parameters, the biological plausibility of the link between the 'EATS-mediated' adversity and endocrine activity should be documented through a MoA analysis (see Section 3.5.1 for further details).

Table 5: High level summary of the scenarios, including the next steps in the assessment; for a full description of the scenario, refer to Section 3.4.4

Adversity based on 'EATS-mediated' parameters	Positive mechanistic OECD CF level 2/3 test	Scenario	Next step of the assessment
No (sufficiently investigated)	Yes/No	1a	Conclude: ED criteria not met because there is no 'EATS-mediated' adversity
Yes (sufficiently investigated)	Yes/No	1b	Perform MoA analysis (postulate and document the MoA, see Section 3.5.1)
No (not sufficiently investigated)	Yes	2a (i)	Perform MoA analysis; additional information may be needed for the analysis
No (not sufficiently investigated)	No (sufficiently investigated)	2a (ii)	Conclude: ED criteria not met because no endocrine activity has been observed for the EATS modalities
No (not sufficiently investigated)	No (not sufficiently investigated)	2a (iii)	Generate missing level 2 and 3 information. Alternatively, generate missing 'EATS-mediated' parameters. Depending on the outcome of these tests move to the corresponding scenario
Yes (not sufficiently investigated)	Yes/No	2b	Perform MoA analysis (postulate and document the MoA, see Section 3.5.1)

EATS: Estrogen, androgen, thyroid, steroidogenic; MoA: Mode of action.

3.5. Mode of Action analysis

When potentially endocrine-related adverse effects and endocrine activity are identified, the link between the two, according to the ED criteria, shall be established based on biological plausibility which shall be determined in the light of current scientific knowledge and under consideration of internationally agreed test guidelines, using a WoE approach.

There are different frameworks which could be helpful in establishing the biological plausibility of the link between an adverse effect and endocrine activity. The International Programme on Chemical Safety (IPCS) Mode of Action (MoA) and human relevancy framework (Boobis et al., 2006, 2008; Meek et al., 2014b) provides a methodology for analysing and transparently laying out the evidence for the MoA of a substance. The WoE methodology, i.e. modified Bradford Hill considerations, is applicable to the assessment of any MoA including endocrine-disrupting MoAs. A MoA analysis facilitates the transparent reporting and assessment of data, requiring explicit consideration of the strengths and weaknesses of the available database including inconsistencies, highlighting qualitative and quantitative similarities and differences across studies/species/strains/sex and related uncertainties, as well as helping to identify and define critical data gaps (Boobis et al., 2008). The OECD Adverse Outcome Pathway (AOP) activity (OECD, 2016c, 2017b) also provides a similar structured framework and weight of evidence methodology, to integrate the evidence. In the weight of evidence considerations in the IPCS MoA framework (adopted also by the AOP framework) both biological plausibility and empirical support are weighted, however, biological plausibility is the most influential consideration (Meek et al., 2014a,b).

> In the weight of evidence considerations in the MoA framework (adopted also by the AOP framework), both biological plausibility and empirical support are weighted, however, biological plausibility is the most influential consideration.

A MoA can be described as a series of biological events, i.e. key events (KEs) that result in the specific adverse effect. The MoA of an endocrine modality will normally contain some earlier KEs (which provide mechanistic information at the molecular or cellular level) and some later KEs (which provide mechanistic information at the organ or system level, including the adverse effect). In the case of endocrine disruption, this sequence at least includes one endocrine-mediated KE which may or may not also be adverse.

KEs are those events that are considered essential to the induction of the (eco)toxicological response as outlined in the postulated MoA. They are empirically observable and measurable steps and can be placed at different levels of biological organisation (at cell, tissue, organ, and individual or population level; see Figure 6). To support an event as key, there needs to be a sufficient body of experimental data in which the event is characterised and consistently measured. KEs are connected to one another and this linkage is termed a key event relationship (KER).

Some concern has been expressed that the level of evidence required by these frameworks to support the sequence of events leading to adversity might be too high for the hazard identification of an ED substance for regulatory purposes (JRC, 2013). However, to conclude on the biological plausibility of the link, it may not be necessary to have demonstrated for the substance under evaluation the whole sequence of events leading to the adverse effect. Existing knowledge from endocrinology and/or toxicology may be sufficient to assess the link and come to a conclusion on the biological plausibility between adverse effects and the endocrine activity.

Figure 5 illustrates the necessary steps in the MoA analysis, which are explained below. The first step of the MoA analysis is to postulate one or more MoA(s) (see Section 3.5.1) by linking the available lines of evidence to (each of) the postulated MoA(s). In this step, it is also necessary to assess whether the available information is sufficient to substantiate the postulated MoA(s). In the second step, it needs to be assessed whether there is a biologically plausible link between endocrine activity and observed adversity (see Section 3.5.2).

Considerations related to human relevance are given in Section 3.5.4.4.

All available data should be reported by following the steps of the MoA analysis described in the following sections in order to transparently document the assessment.

The steps outlined below are generic and apply for both the MoA analysis with respect to humans and with respect to non-target organisms.

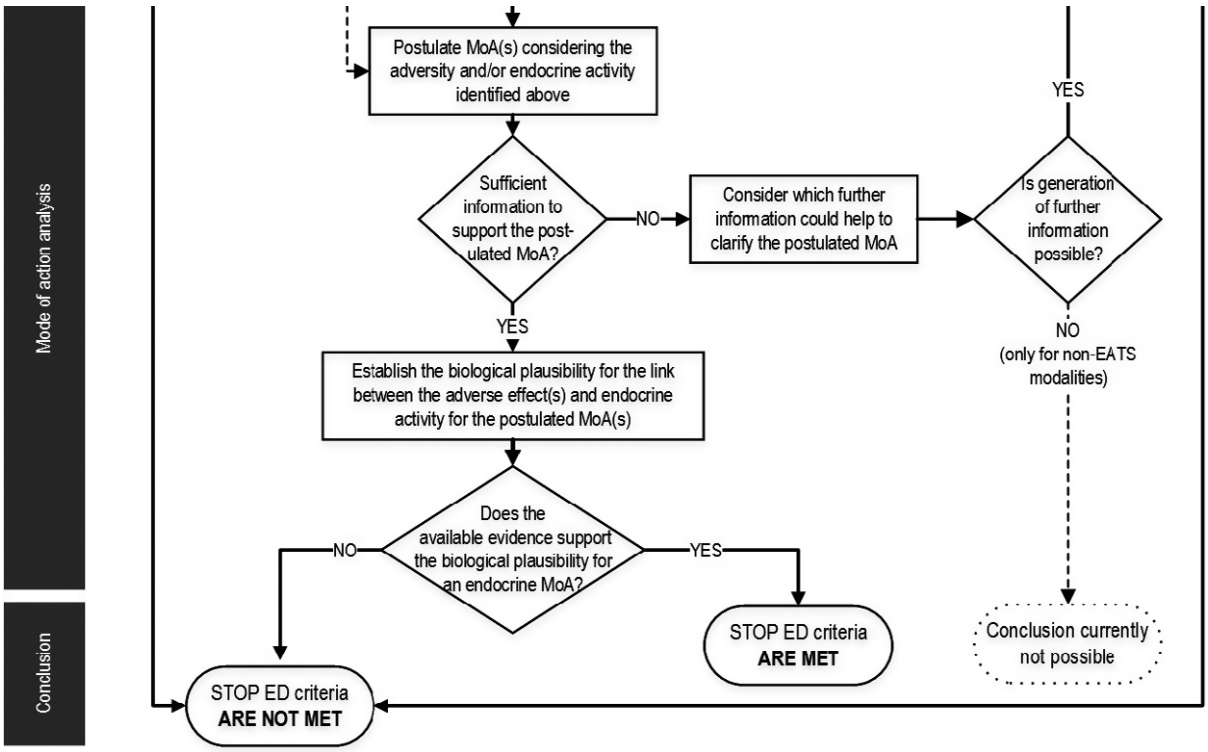
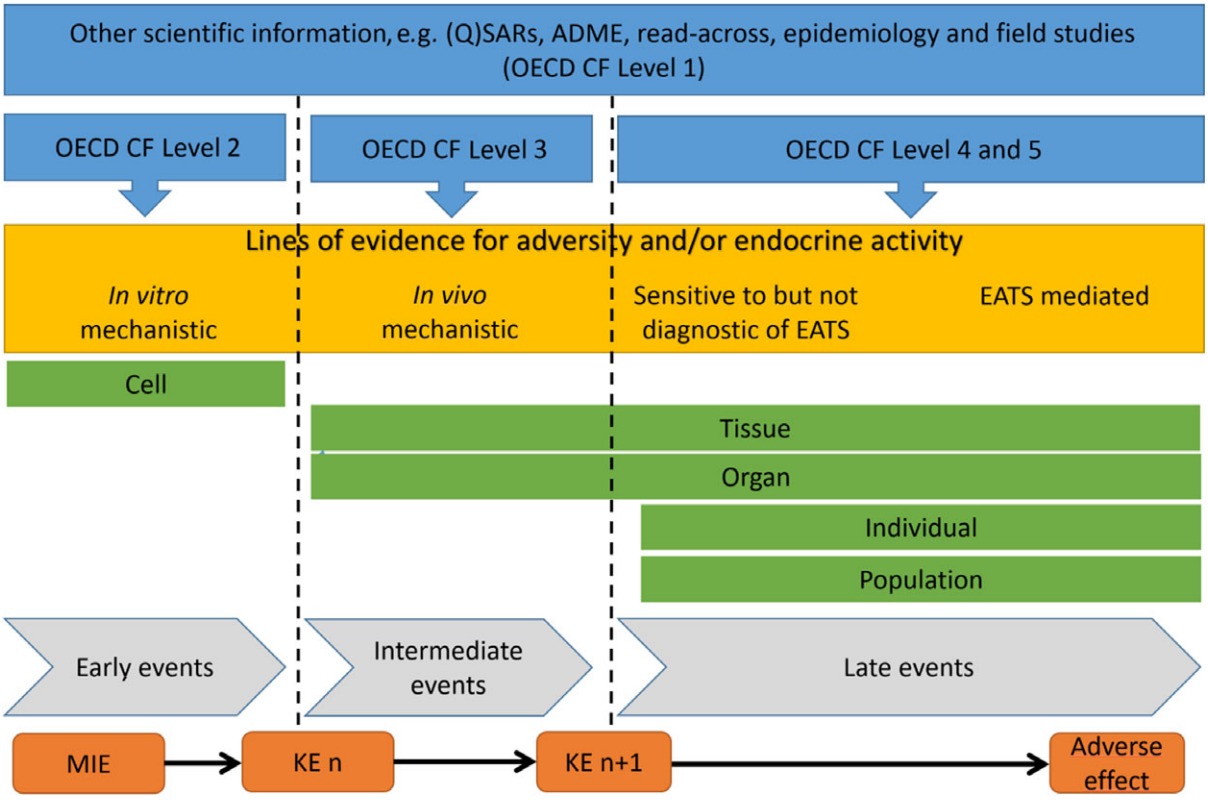


Figure 5: Zoom in on MoA analysis and conclusion steps from the flow chart in Figure 1

3.5.1. Postulate MoA(s) considering the adversity and/or endocrine activity

Either an adverse effect or an endocrine activity (or both) can trigger the MoA analysis (i.e. postulate a MoA and consider if available data are sufficient or which further data would be necessary to support/clarify the postulated MoA, see Section 3.5.3). For this purpose, one or more hypotheses for postulating a MoA(s) could be developed, covering the observed adverse effect(s) and/or endocrine activity that have triggered the assessment.

From the available information assembled and integrated into lines of evidence, there will be indications that suggest whether the substance acts via one or more of the modalities as well as information on potential KEs. In order to postulate a MoA, the information in the lines of evidence is ordered and mapped to the corresponding level of biological organisation (see Figure 6). Subsequently, the KEs in the postulated MoA are identified and briefly described, together with the supporting evidence (i.e. the list of lines of evidence that support each KE) (see Table 6).



KE: key event; MIE: molecular initiating event.

Figure 6: Scheme illustrating how the lines of evidence can be organised to support the postulated mode of action. The arrows linking KEs represent the KE relationships

3.5.2. Establish the biologically plausible link

According to the ED criteria, the biological plausibility of the link between an adverse effect and endocrine activity has to be demonstrated and to do this, this guidance is recommending to use a MoA analysis.

However, there may be situations where a MoA analysis is not needed, due to lack of EATS-mediated adversity in a data set where 'EATS-mediated' parameters have been fully investigated (see scenario 1a in Section 3.4.1). In this scenario, adversity based on 'sensitive to, but not diagnostic of, EATS parameters' is considered as not likely to be caused by alterations in the EATS modalities, because all the 'EATS-mediated' parameters investigated in the same higher tier studies were negative. A MoA analysis is also not needed in scenario 2a(ii) where the EATS-related endocrine activity has been sufficiently investigated and found negative. In scenarios 1b, 2a(i) and 2b, a MoA analysis is required (see Section 3.4.2). Depending on which scenario applies the extent of the MoA analysis may vary.

EATS-mediated adversity

For example in the scenarios 1b and 2b, where adversity is based on 'EATS-mediated' parameters (see also Table 2), the underlying knowledge of the likely endocrine nature of the effects may be such that judgement can be reached on the biological plausibility of a link without recourse to a detailed MoA analysis.

In such cases, the MoA analysis could be very simple; when an adverse effect is 'EATS-mediated', the biologically plausible link is already pre-established in the absence of information proving the contrary (i.e. a fully developed non-ED MoA). This is because, in the case of 'EATS-mediated' parameters, where the pattern of effects is deemed adverse, the biological plausibility that the adverse effects are caused via an EATS-mediated MoA is high, based on existing knowledge and theory (i.e. coherence analysis), and as such, it may not be necessary to generate further empirical data on the substance under evaluation to substantiate the link between the observed adverse effect(s) and an endocrine-mediated MoA.

For example, when performing the assessment with regard to humans, hypospadias accompanied by decreased anogenital distance (AGD) and nipple retention in male rats would be indicative of anti-androgenic activity and adverse at the same time; or when performing the assessment of non-target organisms, a change in sex ratio of fish accompanied by gonad histopathological findings is seen as both adverse and highly likely to be EAS mediated. A detailed MoA analysis is not required in such cases, see above.

> In the case of adversity based on 'EATS-mediated' parameters, the underlying knowledge (i.e. by coherence analysis (Susser, 1991)) of the likely endocrine nature of the effects may be such that judgement can be reached on the biological plausibility of a link without recourse to a detailed MoA analysis.

Adversity based on 'sensitive to, but not diagnostic of 'EATS' parameters

In the scenario 2 (a)(i) where endocrine activity has been observed and where 'sensitive to, but not diagnostic of 'EATS' parameters' are observed and where the pattern of effects is deemed adverse, the biological plausibility that the adverse effects are (exclusively) caused via an endocrine-mediated MoA is not as strong as for the 'EATS-mediated' parameters. Nevertheless, these effects might provide indications of an endocrine MoA which warrant further investigation; in these cases, it is likely that further empirical data will need to be generated, e.g. levels 3, 4 and/or 5 on the substance under evaluation to demonstrate the link between the observed adverse effect and an endocrine MoA.

For non-target organisms (i.e. fish), the most common situation might be that adversity is identified on the basis of 'sensitive to, but not diagnostic of, EATS parameters'. Therefore, to enable a MoA analysis, additional information on intermediate KEs is needed. The decision of which additional study to perform will depend on the available data set. For example, if there is evidence of aromatase inhibition and in addition, a FLCTT is available where only 'sensitive to, but not diagnostic of, EATS' parameters, e.g. fecundity were measured, additional level 3 tests such as the Fish Short-Term Reproduction Assay (OECD TG 229 (OECD, 2012a)) or the 21-day Fish Assay (OECD TG 230 (OECD, 2009b)) may be sufficient to further elucidate the intermediate KEs (e.g. oestradiol level and VTG) (see Appendix G).

Adversity based on non-EATS endocrine parameters

In cases where non-EATS-mediated endocrine MoAs are suspected, although somewhat out of the main scope of this guidance, a MoA analysis should be conducted in order to investigate the biological plausibility between the adverse effects and non-EATS endocrine activity. For example, histopathological findings in the pancreas warrant additional mechanistic investigations targeted on insulin signalling.

Multiple MoAs

A substance may have a single MoA or more than one MoA, which can be endocrine or non-endocrine. The potential of a substance to elicit more than one MoA can obviously lead to difficulties in the interpretation of assay data. If there are indications that a substance may act via multiple MoAs, then the investigations should start with the MoA for which the most convincing evidence is available. The nature of the outlined approach is such that only one MoA is analysed at a time. If several adverse effects are observed, even if recorded in the same organism, which cannot be explained by the same endocrine modality, then each adverse effect will require a separate analysis to discern each MoA leading to the adverse effects. Furthermore, there may be more than one MoA which could cause similar effects; hence, it may be necessary to undertake an analysis of each postulated MoA for a particular adverse effect.

There may be also situations where an adverse effect has been identified which, based on current knowledge, is highly likely to be E, A or S but due to the complexity and cross-talk of the endocrine system it is difficult to identify the specific modality. In such cases, this should be considered an ED regardless through which modality the substance causes adversity.

Alternative non-endocrine MoA

In cases where an applicant considers to postulate an alternative non-endocrine MoA for adverse effects based on EATS-mediated parameters, the level of empirical support and biological plausibility would need to be very strong to demonstrate that the alternative MoA was the more likely explanation of the adverse effects observed. In such cases, a comparative MoA analysis will need to be applied when postulating and substantiating an alternative non-endocrine MoA (Meek et al., 2014b) (see Section 3.5.2). Such an alternative non-endocrine MoA may be postulated where the potentially endocrine-related adverse effects are considered secondary to other non-endocrine related toxic effects (see Section 3.3.1).

Table 6: Example of table summarising the key events based on EATS-mediated parameters. This example shows an EATS-mediated MoA; however, the same table should be used for non-EATS endocrine and for non-endocrine MoA

[Summary of the hypothesis] **The molecular initiating event is unknown; however, the substance increases serum oestradiol in a dose-dependent manner. This results in continuous estrogen receptor 1 activation in estrogen sensitive tissues (numerous tissues are affected however this mode of action focuses on the uterus). The increased estrogen signalling ultimately results in cancer**		
	Brief description of key event (KE)	**Supporting evidence**
Molecular initiating event (MIE)	Inhibition of androgen synthesis (postulated MIE)	None (no data provided, but hypothesised based on current knowledge and former experience with chemicals)
KE 1	Increased serum oestradiol	Increased serum estradiol (OECD TG 407)
KE 2	Uterine hypertrophy	Increased uterine weight (OECD TG 407 and 408)
KE 3	Uterine hyperplasia	Histopathology (OECD TG 408 and 453)
Adverse effect (AE)	Uterine neoplasia	Histopathology (OECD TG 453)

3.5.3. Consider which further information could help to clarify the postulated MoA(s)

If the available information is not sufficient to support the postulated MoA, the generation of further information is needed to substantiate the postulated MoA(s). In principle, any suitable source of information reported in Section 4 could be considered to generate the specific additional information necessary; however, specifically designed mechanistic studies (e.g. hormonal investigations) could also represent a relevant source of information.

On a case-by-case basis, when adversity is indicated by 'EATS-mediated' parameters, and the conclusion on the biological plausibility for the link between adverse effects and endocrine activity for the postulated MoA is challenged by the applicant (refer to previous section Alternative non-endocrine MoA) further data must be generated, in order to substantiate the alternative non-endocrine MoAs.

In some cases, only evidence on endocrine activity may be available (i.e. scenario 2a(i)). In this case, the MoA can be postulated; however, additional information would be needed in order to further develop it including related adversity. For example, if there is mechanistic information indicating endocrine activity, but 'EATS-mediated' parameters have not been sufficiently investigated (i.e. the data set is not sufficient) to empirically support the postulated MoA, it may be necessary to generate *in vivo* level 3, 4 or 5 studies. If no *in vivo* endocrine activity (level 3) or adversity (level 4 or 5) is observed, this would support the lack of an endocrine MoA; if *in vivo* endocrine activity or adversity are observed, the endocrine MoA would need to be substantiated through a MoA analysis. It should however be noted, that in some specific situations, like for aromatase inhibitors and inducers, only level 4 and 5 studies would be applicable to further investigate positive endocrine activity.

Targeted mechanistic studies (e.g. level 2 studies or specifically designed mechanistic studies) may also be of value to address a specific question to either substantiate or remove the concern that the adverse effect arises from an endocrine MoA.

When further information needs to be generated to support the postulated MoA, the applicant should determine the information needed to clarify the concern and thereafter agree this with the evaluating risk assessors. Agreeing with the risk assessors on what type of information is needed may avoid additional data requests later in the process, and thus facilitate decision making.

3.5.4. Additional considerations for non-endocrine or non-EATS mediated MoA(s)

In some cases, i.e. when a non-endocrine or non-EATS-mediated MoA is postulated, it will be necessary to develop a MoA to substantiate the biologically plausible link between the observed adverse effect and the early KE(s) (ideally including the molecular initiating event (MIE)) consequent to the exposure to the specific substance. The available frameworks suggested before are still valid here (Boobis et al., 2006, 2008; Meek et al., 2014b; OECD, 2016c, 2017b). In the case of non-endocrine MoA(s), a comparative WoE analysis will be necessary to increase transparency, consistency and understanding when evaluating the confidence in the WoE supporting the postulated (and competing)

MoAs (Meek et al., 2014b). To determine the biological plausibility for the link between the KEs outlined in the postulated MoA(s) and the adverse effects observed, WoE consideration should be given to a number of elements (modified Bradford Hill considerations; (Becker et al., 2015; Meek et al., 2014a) such as biological plausibility for the KERs (see Section 3.5.4.1), the empirical support for the KERs (see Section 3.5.4.2), i.e. dose–response and temporal concordance, and essentiality for each KE.

Additional elements to support the strength of the postulated MoA are analogy, consistency and specificity (see Section 3.5.4.3). Additionally, human relevance needs to be considered (see Section 3.5.4.4).

It is acknowledged that it may not be possible to address all the elements listed above (e.g. for lack of information). In principle, biological plausibility is weighted more heavily than empirical support. However, there may be cases where the empirical evidence is quite strong, whereas the biological plausibility has not been firmly established (Edwards et al., 2016). Consequently, in such cases, biological plausibility and empirical support related to KERs, or the MoA as a whole, should be considered in combination. As a minimum, the empirical support should provide a clear understanding of the evidence leading to the adverse effect. Although this exercise is expected to be also conducted at the step of assembling and assessing all the evidence for adversity, the same evidence could be used for the empirical support in the MoA context (e.g. time and dose concordance for a known/observed evolution of histological changes like increase in organ weight, follicular cell hypertrophy, hyperplasia, neoplasm in the thyroid; effect observed in multiple species; coherent pattern of effects observed).

3.5.4.1. Biological plausibility for the key event relationships

The assessment should consider whether the KER is consistent with what is known in general (biological plausibility) and also what is known for the substance specifically.

> Biological plausibility of each of the KERs in the MoA is the most influential consideration in assessing weight of evidence or degree of confidence in an overall postulated MoA for establishing the link between the adverse effect and the molecular initiating event (Meek et al., 2014a,b).

The analysis of the biological plausibility for the KER refers only to the broader knowledge of biology. The postulated endocrine MoA and the KEs need to be consistent with the current understanding of physiology, endocrinology and toxicology by addressing structural and/or functional relationships between KEs. In addition to the information that can be directly retrieved from the indications provided in Section 4 or from other *ad-hoc* designed mechanistic studies, the following questions may be helpful to address this element:

- Is the hypothesis consistent with the broader knowledge of biology?
- Is the mechanistic relationship between the KE up and the KE own consistent with established biological knowledge?

Information on biological plausibility for the KERs will come mostly from scientific literature (e.g. endocrinology textbooks, scientific journals and case studies on related topics and associated diseases/ syndromes). It is recommended that supporting references justifying the biological plausibility for the KERs are considered as part of the WoE assessment. It is recognised that there may be cases where the biological relationship between two KEs may be very well established. In such cases, it may be impractical to exhaustively cite the relevant primary literature.

The biological plausibility is weighted as follows:

- Strong: if is there is extensive understanding of the KER based on extensive previous documentation and broad acceptance
- Moderate: if the KER is plausible based on analogy with accepted biological relationships, but scientific understanding is not completely established
- Weak: the structural or functional relationship between the KEs is not understood.

3.5.4.2. Empirical support for dose–response/incidence and temporal concordance for the key event relationship

Dose and temporal concordance are important elements which must be addressed when determining the empirical support for KERs. Comparative tabular presentation of the KEs, including information on the time point of the observations and the severity/incidence of the effects observed is essential in examining both dose-effect and temporal concordance (see Table 7 and (OECD, 2016c)).

Table 7: Example of a table which allows analysis of both dose–response and temporal concordance between the key events (KEs).This example shows an EATS-mediated MoA; however, the same table should be used for non-EATS endocrine and for non-endocrine MoA

***[Species X]* dose–response and temporal concordance between the key events**				
	KE 1	**KE 2**	**KE 3**	**Adverse effect**
	Increased serum oestradiol	**Uterine hypertrophy**	**Uterine hyperplasia**	**Uterine neoplasia**
Dose (mg/kg/day)				
10		– (90 days)	– (90 days)	
30	+ (28 days)	+ (28 days)		– (2 years)
90	++ (28 days)	++ (28 days) +++ (90 days)	+ (90 days)	+ (2 years)
180		+++ (28 days)	++ (90 days and 2 years)	++ (2 years)
360	+++ (28 days)	+++ (90 days)	+++ (90 days)	

Only key events with available data for dose-response and temporal concordance are included.
– indicates no effect; +, ++ and +++ indicates the effect size, i.e. severity.

The dose–response and temporal concordance can be used either within one specific study, where parameters associated with different KEs are measured, or across studies. Most often, the complete data set needed to fully address temporal concordance is not available and this should be considered in the WoE.

Dose–response/incidence concordance. This analysis focuses on the characterisation of the dose–response/incidence concordance for the KEs. The following questions may be helpful to address this element:

- Are the KEs observed at doses below or similar to those associated with the adverse effect?
- Are the earlier KEs observed at doses below or similar to the doses of later KEs?
- Is the incidence of the adverse effect consistent with the incidence of each KE? (e.g. at similar doses the incidence of the adverse effect would not be expected to be greater than that of earlier KEs but can/should be lower, or may not be observed at all in some studies).

Temporal concordance. This analysis focuses on the temporal relationships of the KEs to each other and the adverse effect. The temporal sequence of the KEs leading to the adverse effect should be established. The following questions may be helpful to address this element:

- Are the KEs observed in the hypothesised order?
- Are the earlier KEs observed in studies of similar or shorter duration of later KEs?

KEs should occur before the adverse effect and should be consistent temporally with each other (e.g. receptor activation followed by cellular/tissue response which progresses to adversity). This is essential in order to determine whether or not the available evidence supports the postulated MoA.

In those cases where temporal concordance cannot be demonstrated, the existing biological knowledge of the sequence of the events, if supported, may be considered sufficient.

The empirical support is weighted as follows:

- Strong: if there is extensive evidence for temporal, dose–response and incidence concordance and no or few critical data gaps or conflicting data.
- Moderate: if there is evidence inconsistent with the expected pattern for which, however, an explanation can be found (e.g. based on experimental design, technical considerations, differences among laboratories).
- Weak: if there are significant inconsistencies in the empirical support (e.g. no dose-response and temporal concordance, inconsistencies among studies) that cannot be explained.

3.5.4.3. Essentiality, consistency, analogy and specificity of the evidence for the association of the KEs with the adverse effect

This section focuses on the evidence for linking the KEs in the postulated endocrine MoA to the adverse effect by analysing the elements of essentiality, consistency, analogy and specificity. Table 8 gives an example of how to transparently document these elements.

Essentiality. This is an important aspect to consider for all hypothesised MoAs (although it is recognised that information is not always available to assess it). Stop/recovery studies (if available), or experiments conducted in knock-out animal models for a postulated KE, showing absence or reduction of subsequent KEs or the adverse effect when a KE is blocked or diminished are an important test for demonstration of essentiality. The following question may be helpful to address this element:

- Is the sequence of events reversible if dosing is stopped or a KE prevented?

The essentiality is weighted as follows:

- Strong: if there is direct evidence from specifically designed experimental studies illustrating essentiality for at least one of the KEs (e.g. stop/reversibility studies, antagonism, knock-out models, etc.).
- Moderate: if there is indirect evidence that sufficient modification of an expected modulating factor attenuates or augments a KE.
- Weak: if there is contradictory experimental evidence of the essentiality of any of the KEs or there is evidence for no reversibility.

> Essentiality is an important aspect to consider for all postulated MoAs although it is recognised that information is not always available to assess it.

Consistency. This analysis addresses the repeatability of the KEs in the postulated MoA in different studies/species/strains/systems. For example, consistent observation of the same KE(s) in a number of studies with different study design increases the support, since different designs may reduce the potential for unknown biases and/or confounding factors. Both positive and negative results should be considered. The following questions may be helpful to address this element:

- Is there consistency across studies for the relevant parameters?
- Is the pattern of effects across studies/species/strains/systems consistent with the hypothesised MoA?

Analogy. This analysis addresses whether or not the postulated KEs also occur for other substances for which the same MoA has already been established. The following question may be helpful to address this element:

- Is the same sequence of KEs observed with other substances for which the same MoA has been established?
- Would the MOA be anticipated based on broader chemical specific knowledge?

Specificity. This analysis looks at whether the MoA for the adverse effect is endocrine-related, i.e. if an adverse effect is a consequence of the hypothesised endocrine MoA, and not an indirect result of other non-endocrine-mediated toxicity. The following questions may be helpful to address this element:

- Could the adverse effect be the result of a different MoA (i.e. non-endocrine-mediated)?

In the context of this guidance, consistency, analogy and specificity are important elements that support the strength of the MoA. This is because these elements mainly refer to individual KE(s) and not to the KER(s).

3.5.4.4. Human relevance

The criteria clarify that relevance to humans should be assumed by default in the absence of appropriate scientific data demonstrating non-relevance. The IPCS MoA and human relevance framework (Meek et al., 2014b) provides guidance on how to establish and demonstrate non-relevance to humans of the adverse effects observed in animal models. It should however be noted, that such a framework is considering both qualitative as well as quantitative aspects to define human relevance, whereas this guidance is focussing on hazard identification and, as such, is mainly focusing on the qualitative aspects described by the framework.

A substantial amount of information is therefore required to conclude that the given endocrine MoA is not relevant to humans. If such a conclusion is strongly supported by the data, then a substance producing endocrine disruption in animals only by that endocrine MoA would not be considered to pose an ED hazard to humans. It is worth noting that where an endocrine MoA is considered not to be

relevant for humans, absence of other/concomitant endocrine MoAs leading to the same adverse effect in humans should also be excluded.

3.5.5. Extent of support for the overall assessment of the MoA analysis

The result of the analysis conducted for the elements in Sections 3.5.4.1, 3.5.4.2 and 3.5.4.3 should be transparently documented by the applicant. The proposed documentation is applicable to any MoA analysis. Tables 6 and 7 give an example of how this information could be summarised as a minimum. It is noted that elements in Sections 3.5.4.1, 3.5.4.2 and 3.5.4.3 may be not needed in the case of EATS-mediated adversity. An example on how to deal with adversity based on 'sensitive to but not diagnostic of EATS' parameters is reported in Appendix G (MoA for fish is reported).

To increase transparency, the rationales for the assignment of the scores based on the specified questions/considerations should be justified. The rationales should explicitly provide the reasoning for assignment of the score, based on the considerations for strong, moderate or weak weight of evidence. Therefore, the outcome of the analysis should always be reported and should include, as a minimum, the postulated MoA and at least a qualitative justification of the assessment.

Biological plausibility of each of the KERs in the MoA is the most influential consideration in assessing weight of evidence or degree of confidence in an overall postulated MoA for the application of the MoA analysis (Meek et al., 2014a,b). The assessment of the overall biological plausibility should also identify the KEs for which confidence in the relationship with the adverse effect is greatest (i.e. to facilitate determining the most sensitive predictor of the adverse effect).

It is important to recognise that, where possible, empirical support relates to 'concordance' of dose response, temporal and incidence relationships for KERs rather than the KEs; the defining question is not whether or not there is a dose response relationship for an associated KE but rather, whether there is expected concordance with the dose-response relationships for earlier and later KEs.

The essentiality, if experimentally provided, of the KEs is influential in considering confidence in an overall postulated MoA being secondary only to biological plausibility of KERs (Meek et al., 2014a,b). It is assessed, generally, on the basis of direct experimental evidence of the absence/reduction of downstream KEs when an upstream KE is blocked or diminished (e.g. in null animal models or reversibility studies).

Identified limitations of the database to address the biological plausibility of the KERs, the essentiality of the KEs and empirical support for the KERs are influential in assigning the scores for degree of confidence (i.e. strong, moderate or weak).

Where the MoA has not previously been established (i.e. non-endocrine MoA or non-EATS mediated endocrine MoAs), the possibility that a plausible case can be made because of existing biological understanding should be transparently addressed. Due to the complexity of this process, the focus of the analysis should not only address the sufficiency of underlying data to support a particular MoA conclusion but, specifically for non-endocrine MoA, also to illustrate a comparative analysis for increasing transparency in the data. The comparative analysis should assess the WoE of alternative MoA(s) for the specific substance (based on modified Bradford Hill considerations), to more explicitly indicate and document the degree of confidence in the postulated (and competing) MoA vs an endocrine-mediated MoA. Separate conclusions should be made, based on the extent of supporting WoE for the postulated MoAs for the same substance, using the experimental evidence and articulated and explicit considerations (Meek et al., 2014a,b).

3.5.6. Conclusion on the MoA analysis

The possibility of concluding on the ED properties of a substance by applying the MoA framework depends on whether there is sufficient evidence to establish the biological plausibility of the link between the observed adverse effect and the endocrine activity.

The overall conclusion is based on the WoE elaborated to substantiate the postulated MoA.

Following the assessment, a statement of confidence on the overall conclusion is necessary to address the strength of the evidence for the postulated MoA. A clear statement on the extent to which the KEs fit the postulated MoA(s) should be given, reflecting the biological plausibility for the KERs, the empirical support for the KERs, and the essentiality for the KEs. When essentiality data are available they should be considered using a WoE approach. If essentiality is proven, it should be considered as relevant information to strengthen the MoA. Similarly, consistency, analogy and specificity are important elements to substantiate the strength of the postulated MoA.

The documentation of the remaining uncertainties should include any uncertainties associated with the selection of the evidence, reliability and relevance, and the WoE method. Additionally, any

uncertainties stemming from the use of expert knowledge should be listed. Furthermore, if an additional conclusion is possible, this should be also listed as an uncertainty. It is recommended that the uncertainties are reported in a tabular form as exemplified in Table 8.

Table 8: Example summarising the conclusions on the biological plausibility of the link between the adverse effect and the endocrine activity for a postulated MoA. This example shows an EATS-mediated MoA; however, the same table should be used for non-EATS endocrine and for non-endocrine MoA

	Key event relationships (KERs)			
	MIE to KE 1	**KE 1 to KE 2**	**KE 2 to KE 3**	**KE 3 to AE**
Biological plausibility for the KERs	MODERATE – It is known that chemically induced inhibition of androgen synthesis can increase the oestradiol/ testosterone ratio with a significant elevation of total or free hormone. Although this is plausible, the scientific understanding is still incomplete and/or different MIE can be postulated	STRONG – It is well documented and mechanistically accepted that unopposed estrogen action results in hypertrophy, hyperplasia and ultimately cancer	See KE 1 to KE 2	See KE 1 to KE 2
Empirical support for the KERs	MODERATE – The substance clearly increases serum oestradiol in a dose-dependent manner; however a dependent change in both key events following perturbation of the MIE is not data supported	STRONG – substance increases uterine weight (KE 2) following hormonal perturbation (KE 1) with dose–response and temporal concordance	STRONG – dose/ incidence and time concordance is observed for the relationship between KE 2 and KE 3	STRONG – It is known that a continuum exists between uterine epithelial cell hyperplasia and adenoma and the relationship between the two KEs is showing incidence and time concordance

	MIE	**KE 1**	**KE 2**	**KE 3**	**AE**
Essentiality of KEs	No data				
		MODERATE – There are no stop-recovery studies available. However, based on human clinical experience (provide references) an unopposed estrogen action is essential for the tumour development			
			See KE 1		
				See KE 1	
					See KE 1
Consistency	The KEs have been observed consistently in three different studies with different duration. The pattern of effects is consistent between the studies there are no conflicting observations. Consistency across species cannot be assessed because there are only rat studies available				
Analogy	No information. Increase in oestradiol is reported for some antifungal agent, but a full MoA was not developed				
Specificity	In this case, the MIE is unknown; however, the substance clearly increases the levels of oestradiol at doses well below those which induce general systemic toxicity.				

Identified uncertainties [for guidance see (EFSA Scientific Committee, 2018)]	Comment
Uncertainty 1 *[Lack of a clear understanding of the MIE]*	Increase in oestradiol can be consequent to many MIE
Uncertainty 2 *[For the empirical support for the KER between the MIE and the KE 1, data are only available for the perturbation of the KE down]*	A clear dose and temporal concordance cannot be established
Uncertainty 3 *[Effect only observed in one species]*	
Uncertainty 4 *[Hormonal assessment only performed for oestradiol]*	A more comprehensive hormonal study, measuring testosterone or additional steroid hormones would be beneficial for postulate more precisely the MIE
Overall conclusion on the postulated MoA	
The MIE is unknown, however, the overall biological plausibility is strong and substantiated by a strong empirical support for the majority of postulated KEs. The substance increases estrogen activity through increased serum oestradiol which ultimately results in cancer. It is considered likely that this is an endocrine MoA as no alternative non-endocrine mode of action has been identified	

MIE: molecular initiating event; KE: key event; AE: adverse effect; MoA: mode of action.

3.6. Overall conclusion on the ED criteria

In line with the criteria, the conclusions should answer the two problem formulations identified within this guidance and a conclusion should be drawn for humans and non-target organisms:

- Is there a biologically plausible link between endocrine activity and observed adverse effect(s) that are relevant for humans?
- Is there a biologically plausible link between endocrine activity and observed adverse effect(s) that are relevant for non-target organisms at population level?

It is sufficient that the substance meets the ED criteria for one group of non-target organisms in order to be identified as ED.

Where, based on a sufficient data set, no 'EATS-mediated' adversity was observed or where endocrine activity was found negative, it is possible to by-pass the MoA analysis and to conclude that the criteria are not met.

Where a MoA is based on 'EATS-mediated' adversity the ED criteria are considered met; unless an alternative non-endocrine MoA is demonstrated and in a comparative analysis found to be the most likely explanation.

Where a MoA is based on 'sensitive to but not diagnostic of EATS' adversity and the MoA supports the biological plausibility of the link between the observed adverse effects and endocrine activity for at least one postulated MoA(s), the substance is considered to meet the ED criteria, unless an alternative non-endocrine MoA is demonstrated and in a comparative analysis found to be the most likely explanation.

Where the available information is sufficient to postulate a non-EATS endocrine MoA, it is possible that, the supporting available information would be not sufficient to develop the MoA. In these situations, an analysis of the available testing methodologies should be carried out by the applicant in order to justify that the generation of further scientific information suitable for the identification of a non-'EATS-mediated' endocrine MoA is not feasible based on the available scientific knowledge and that the biological plausibility is highly uncertain, and therefore, a conclusion is currently not possible.

There may be cases where data are not provided for performing the ED assessment according to this Guidance and this is not considered justifiable. For example, failure to perform the MoA analysis as required, failure to generate the information needed to sufficiently investigate endocrine activity and/or endocrine related adversity (despite the fact that appropriate test methods are available), and failure to provide adequate scientific justifications for omission of information. In all those cases, the assessors shall clearly indicate which missing information should have been provided by the applicant when following the present Guidance and to which extent this information is critical to allow a conclusion to be reached on the ED properties of a substance.

The conclusion on the ED criteria needs to be transparently documented, including the remaining uncertainties.

4. Information sources for endocrine disruptor identification

In this section, the sources of information that may be used and helpful for the assessment and identification of the ED properties of a substance are described. These information sources comprise non-test methods, *in vitro* and *in vivo* test methods, and other data (as described in Section 3.2.1.2).

OECD Conceptual Framework and OECD GD 150

This section is largely based on the 'Guidance document on standardised test guidelines for evaluating chemicals for endocrine disruption' provided by the Organisation for Economic Co-operation and Development (OECD GD 150; OECD, 2018b). The OECD GD 150 provides widely accepted consensus guidance on the interpretation of effects measured in relevant OECD test Guidelines (OECD test guidelines), which may arise as a consequence of perturbations of EATS modalities, and how these effects might be evaluated to support ED identification.

OECD GD 150 includes the OECD Conceptual Framework for Testing and Assessment of Endocrine Disrupting Chemicals (OECD CF, see Table 9). The OECD CF lists the OECD test Guidelines and standardised test methods available, under development or proposed, that can be used to evaluate chemicals for endocrine disruption.

The OECD CF is not intended to be a testing strategy but to provide a guide to the tests available and what type of information the tests generally provide. It is important to bear in mind that as the field of endocrine disruption is still developing, the OECD CF and its associated guidance will be subject to periodic revisions.

Table 9: OECD Conceptual Framework for Testing and Assessment of Endocrine Disrupting Chemicals, revised 2018 (OECD, 2018b)

Mammalian and non mammalian toxicology		
Level 1 Existing data and existing or new non-test information	• Physical & chemical properties, e.g., MW reactivity, volatility, biodegradability • All available (eco)toxicological data from standardised or non-standardised tests. • Read-across, chemical categories, QSARs and other *in silico* predictions, and ADME model predictions	
Level 2 *In vitro* assays providing data about selected endocrine mechanism(s)/ pathways(s) (Mammalian and non mammalian methods)	• Estrogen (OECD TG 493) or androgen receptor binding affinity (US EPA TG OPPTS 890.1150) • Estrogen receptor transactivation (OECD TG 455, ISO 19040-3), yeast estrogen screen (ISO 19040-1 & 2) • Androgen receptor transactivation (OECD TG 458) • Steroidogenesis *in vitro* (OECD TG 456) • Aromatase Assay (US EPA TG OPPTS 890.1200) • Thyroid disruption assays (e.g. thyroperoxidase inhibition, transthyretin binding) • Retinoid receptor transactivation assays • Other hormone receptors assays as appropriate • High-Throughput Screens	
	Mammalian toxicology[(c)]	**Non-mammalian toxicology**[(c)]
Level 3 *In vivo* assays providing data about selected endocrine mechanism(s)/ pathway(s)[(a)]	• Uterotrophic assay (OECD TG 440) • Hershberger assay (OECD TG 441)	• Amphibian metamorphosis assay (AMA) (OECD TG 231) • Fish short-term reproduction assay (FSTRA) (OECD TG 229)[(b)] • 21-day fish assay (OECD TG 230) • Androgenised female stickleback screen (AFSS) (GD 148) • EASZY assay. Detection of Substances Acting Through Estrogen Receptors Using Transgenic cyp19a1b GFP Zebrafish Embryos. (draft OECD TG) • *Xenopus* embryonic thyroid signalling assay (XETA) (draft OECD TG)

		• Juvenile Medaka Anti-Androgen Screening Assay (JMASA) (draft OECD GD) • Short-Term Juvenile Hormone Activity Screening Assay Using *Daphnia magna* (draft OECD TG) • Rapid Androgen Disruption Adverse Outcome Reporter (RADAR) Assay (draft OECD TG)
Level 4 *In vivo* assays providing data on adverse effects on endocrine relevant endpoints[(b)]	• Repeated dose 28-day study (OECD TG 407) • Repeated dose 90-day study (OECD TG 408) • Pubertal development and thyroid Function assay in peripubertal male rats (PP male Assay) (US EPA TG OPPTS 890.1500) • Pubertal development and thyroid function assay in peripubertal female Rats (PP female assay) (US EPA TG OPPTS 890.1450) • Prenatal developmental toxicity study (OECD TG 414) • Combined chronic toxicity and carcinogenicity studies (OECD TG 451-3) • Reproduction/developmental toxicity screening test (OECD TG 421) • Combined repeated dose toxicity study with the reproduction/ developmental toxicity screening test (OECD TG 422) • Developmental neurotoxicity study (OECD TG 426) • Repeated Dose Dermal Toxicity: 21/28-day Study (OECD TG 410) • Subchronic dermal toxicity: 90-day study (OECD TG 411) • 28-Day (Subacute) Inhalation Toxicity Study (OECD TG 412) • Subchronic inhalation toxicity: 90-day study (OECD TG 413) • Repeated dose 90-day oral toxicity study in non-rodents (OECD TG 409)	• Fish sexual development test (FSDT) (OECD TG 234) • Larval amphibian growth & development assay (LAGDA) (OECD TG 241) • Avian reproduction test (OECD TG 206) • Fish early life stage (ELS) toxicity test (OECD TG 210) • New guidance document on harpacticoid copepod development and reproduction test with *Amphiascus* (OECD GD 201)[(b)] • *Potamopyrgus antipodarum* reproduction test (OECD TG 242)[(d)] • *Lymnaea stagnalis* reproduction test (OECD TG 243)[(d)] • Chironomid toxicity test (OECD TG 218-219)[(d)] • *Daphnia magna* reproduction test (with male induction) (OECD TG 211)[(d)] • Earthworm reproduction test (OECD TG 222)[(d)] • Enchytraeid reproduction test (OECD TG 220)[(d)] • Sediment water *Lumbriculus* toxicity test using spiked sediment (OECD TG 225)[(d)] • Predatory mite reproduction test in soil (OECD TG 226)[(d)] • Collembolan reproduction test in soil (TG OECD 232)[(d)]
Level 5 *In vivo* assays providing more comprehensive data on adverse effects on endocrine relevant endpoints over more extensive parts of the life cycle of the organism[(b)]	• Extended one-generation reproductive toxicity study (OECD TG 443)[(e)] • Two-Generation reproduction toxicity study (OECD TG 416 most recent update)	• Fish lifecycle toxicity test (FLCTT) (US EPA TG OPPTS 850.1500) • Medaka extended one-generation reproduction test (MEOGRT) (OECD TG 240) • Avian Two-generation toxicity test in the Japanese quail (ATGT) (US EPA TG OCSPP 890.2100/740-C-15-003) • Sediment water chironomid life cycle toxicity test (OECD TG 233)[(d)] • *Daphnia* multigeneration test for assessment of endocrine disrupting chemicals (draft OECD TG)[(d)] • Zebrafish extended one-generation reproduction test (ZEOGRT) (draft OECD TG)

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

Notes to the OECD Revised Conceptual Framework:
Note 1: Entering at all levels and exiting at all levels is possible and depends upon the nature of existing information and needs for testing and assessment.
Note 2: The assessment of each chemical should be made on a case-by-case basis, taking into account all available information.
Note 3: The framework should not be considered as all inclusive at the present time. It includes assays that are either available, or for which validation is under way. With respect to the latter, these are provisionally included, and a few assays (e.g. the ATGT) have only been validated at national level. At level 2, some assays are not (yet) proposed for validation but are included because they may provide information on important molecular interactions.

OECD Conceptual Framework level 1 refers to existing data and non-test information such as read-across and category approaches, (Q)SAR and other *in silico* approaches. *In silico* predictions may be used as supporting information for EATS modalities, e.g. in relation to the MIE, when assembling lines of evidence. The evidence from *in silico* predictions is strengthened if the same result is obtained with independent *in silico* models ((Q)SAR and/or read-across). *In vitro* mechanistic screening assays are placed at level 2. Assays placed at level 3 of the OECD CF are *in vivo* screening assays designed to provide information about whether a compound has the ability to act via specific endocrine-mediated modalities. If no effects are observed in a level 3 study, it cannot be concluded that the substance has no ED effects, both due to the small group sizes used in these screening studies (i.e. low power to detect effects), lack of testing of sensitive life stages and since the substance may act through other ED MoAs than the one investigated by the assays. Assays from CF level 3 may also provide some evidence of adversity in certain circumstances. *In vivo* assays that may provide data on adverse effects on endocrine relevant parameters are listed at levels 4 and 5 of the OECD CF. All assays at these levels primarily measure apical endpoints that are potentially adverse and in some cases may be indicative of an endocrine activity (i.e. EATS-mediated). Some of these assays have been, or are in the process of being, validated with the inclusion of additional endocrine parameters.

In the OECD GD 150, all test methods are sorted according to which level of the OECD CF they occupy. In addition, in the OECD GD 150, the test methods are grouped in two parts (A and B) according to the extent of guidance provided for effects interpretation. The test methods listed under Part A are established test methods which have been validated and published as OECD test guidelines, whereas the test methods listed under Part B have not received full validation by OECD, or are in the process of OECD validation, or are guidelines which have been validated and published by non-OECD organisations. As more ED-relevant test methods are developed into Test guidelines or endocrine parameters added to existing test guidelines, it is anticipated that both the OECD GD 150 and this guidance will need to be updated.

All the parameters, reported in OECD GD 150 and in Sections 4.2 and 4.3 of this guidance and considered to be relevant to support ED identification, are mainly derived from guideline studies, i.e. standardised test methods validated for regulatory decision making (e.g. EU test methods/OECD test guidelines or United States Environmental Protection Agency (US EPA)/Food and Drug Administration (FDA) test guidelines). However, guideline studies, other than those listed in OECD GD 150, may also include apical endpoints that can be affected by endocrine modes of action, and therefore may provide relevant information. Furthermore, information on the broader toxicological profile of the substance may provide better understanding of potential indirect effects on the endocrine system.

In addition, non-standardised test methods can also be used to derive relevant information provided that they are appropriately designed and judged to be of acceptable quality (see Section 3.2.1). In general, any non-standard study providing information on relevant EATS-mediated effects similar to those derived from standardised test methods (see Sections 4.2 and 4.3 for a more detailed list) should be considered. In addition, some non-standard studies may provide information on non-EATS endocrine modalities such as those involving the corticosteroid axis, somatotropic axis, and the retinoid, vitamin D and peroxisome proliferator-activated receptor signalling modalities (see OECD Detailed Review Paper 178 (OECD, 2018b)).

Finally, it is important to bear in mind while carrying out the ED assessment (Section 3), that some parameters (such as decreased body weight consequent to a decrease of food consumption) do not

necessarily reflect an endocrine MoA and are not included in OECD GD 150, but are nevertheless important for the interpretation of whether observed effects, which may potentially arise through EATS modalities, are possibly a non-specific secondary consequence of other toxic effects.

Other sources of information

The primary data sources will be the data generated using standardised test methods (see Sections 4.2 and 4.3) and the systematic literature review, which allow retrieving published literature (see Section 3.2.1.2) according to the data requirements of the specific regulatory framework. Human (epidemiological) data (see Section 4.4.1) and Field studies, from controlled field experiments (see Section 4.4.2) are retrieved from the data requirements of the specific regulatory framework and, when available, also from the published literature.

In addition, if a substance is regulated under other EU Regulations (e.g. REACH[12] and Cosmetic Product Regulation[13]), the available dossier could provide additional relevant information.

Other sources and types of information to be considered include the following:

- Databases of compiled data (see Table 10)
- (Q)SAR model outputs (see Section 4.1)
- Read-across and category approaches (see Section 4.1).

A general overview of some relevant public databases of compiled data (not exhaustive) is given in Table 10. It is worth noting that the data contained in these databases is not exclusive to EATS and the criteria used to consider a chemical as endocrine active may vary among databases. More information can be found in Appendix D.

Table 10: Databases of compiled data (not exhaustive)

Databases which can provide information on endocrine MoA	ToxCast (US EPA)
	ToxCast ER prediction model (US EPA)
	Endocrine disruptor screening program, EDSP21 (US EPA)
	OECD (Q)SAR toolbox (OECD, ECHA)
	AOP knowledge base (OECD)
	ToxRefDB (US EPA)

4.1. Non-test methods

The assessment of ED properties has been traditionally carried out with vertebrates and *in vitro* testing. Experience gained through testing has been used to build models that predict endocrine activity. The OECD CF for the screening and testing of endocrine-disrupting chemicals lists non-test information such as read-across, chemical categories, (Q)SARs and other *in silico* predictions, including predictions of ADME properties at level 1.

Several software tools to predict ED-related properties/activities of substances and databases containing information on endocrine-active or endocrine-disrupting properties are available. A brief overview of available software tools for predicting endocrine activity is given in Table 11. Most of these software systems are commercially available, although some can be used for free. Databases that contain relevant information on endocrine-active or endocrine-disrupting properties are listed in Table 10. A more detailed description of the software tools as well as the databases is provided in Appendix D. It is important to note that the list of databases, tools and models in Appendix D is not exhaustive and that the applicability (e.g. applicability domain) of the models should be obtained from more detailed description in the literature.

[12] Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC. OJ L 396, 30.12.2006, p. 1–849. http://data.europa.eu/eli/reg/2006/1907/oj

[13] Regulation (EC) No 1223/2009 of the European Parliament and of the Council of 30 November 2009 on cosmetic products. OJ L 342, 22.12.2009, p. 59–209. http://data.europa.eu/eli/reg/2009/1223/oj

In silico prediction methods

A range of *in silico* predictive methods related to ED have been described in previous reviews (Benigni et al., 2017; Cronin and Worth, 2008; EFSA Scientific Committee, 2013; JRC, 2014; Lo Piparo and Worth, 2010).

In silico predictions may be used as a means of generating supporting information for endocrine modalities within a WoE approach. In particular, by providing information on the MIE, *in silico* predictions can be used to support the identification of endocrine modes of action and contribute to informing the decision on the most appropriate testing strategy when generation of new data is required.

Whenever *in silico* methods are used, the general provisions outlined in ECHA Guidance R6 should be followed (ECHA, 2008).

The different types of *in silico* prediction methods can be grouped as:

Molecular modelling of receptor interactions

These models make use of the 3D structure of the receptor and/or ligand to determine EAS. Molecular dynamics (McGee et al., 2008), docking studies (Warren et al., 2006) and 3D-(Q)SARs like the comparative molecular field analysis (CoMFA) (Cramer et al., 1988) are examples of receptor interaction models in decreasing level of complexity and detail provided.

More specialised expertise and computational power may be needed to apply these approaches. For example, precise knowledge about the receptor structure, presteps for the selection of the 'active' conformers, or supercomputers to carry out molecular dynamics may be needed. Therefore, these methods are less likely to be routinely used for regulatory purposes. However, information and mechanistic understanding derived from such models may be useful in supporting the identification of MoA.

(Q)SAR modelling of receptor-based activity

These models correspond to mathematical relations between the structural and/or physicochemical properties of chemicals and their receptor-related effects (e.g. binding affinities to nuclear receptors (NR)) or more downstream effects (e.g. transcriptional activation of NR pathways, developmental toxicity). These models cover different types of receptors (e.g. ER, AR, TR) and affinities (agonist/antagonist) and provide qualitative or quantitative binding information (Kleinstreuer et al., 2017; Li and Gramatica, 2010; Panaye et al., 2008; Renjith and Jegatheesan, 2015; Ribay et al., 2016; Vedani et al., 2012; Zhang et al., 2013; Zhao et al., 2005). An extensive (but not exhaustive) list of models from the literature for the prediction of nuclear receptor binding is provided in Appendix D. Unlike some molecular modelling approaches, (Q)SARs are in general very easy to use, especially when already implemented in software (see Table 11).

Profilers based on structural alerts and decision trees

These types of models are simple algorithms that search for predefined structural motifs which indicate a probable activity such as protein binding or ER activation. They are usually based on existing structure–activity relationships (SARs) or chemotypes (property-enhanced alerts). They can be derived from statistical modelling or mechanistic considerations. These models may also include decision trees based on multiple structural alerts and/or properties.

These approaches are very valuable as profilers to support the grouping of chemicals for read-across (Wu et al., 2013; JRC, 2014). For ease of use, profilers are typically implemented in software tools, such as the OECD (Q)SAR Toolbox (OECD, 2014; Dimitrov et al., 2016) and the Chemotyper (Yang et al., 2015).

Table 11: Software tools for predicting endocrine activity

Software tool	Effect addressed				
	E	A	T	S	Other
EDKB	X	X			
ADMET Predictor	X				
ACD/Labs Percepta – Toxicity Module	X				
Derek	X				
MolCode Toolbox	X				X[(a)]
CASE Ultra	X	X			
TIMES	X	X			X[(a)]
VirtualToxLab	X	X	X	X[(b)]	X[(c)]
OECD (Q)SAR Toolbox	X				
Endocrine Disruptome	X	X	X	X[(d)]	X[(e)]
COSMOS KNIME workflow	X	X	X	X[(d)]	X[(f)]
Danish (Q)SAR DB	X	X	X		X[(g)]
(Q)SAR Data Bank	X				
VEGA platform	X				

AhR: aryl hydrocarbon receptor; GR: glucocorticoid receptor; LXR: liver X receptor; PPAR: peroxisome proliferator-activated receptor; RXR: retinoic acid receptor; AR: androgen receptor; ER: estrogen receptor; GR: glucocorticoid receptor; PR: progesterone receptor; FXR: farnesoid X receptor; PXR: pregnane X receptor; TR: thyroid hormone receptor.
(a): AhR.
(b): GR, mineralocorticoid.
(c): AhR, LXR, PPAR γ, enzymes CYP450 3A4 and 2A13.
(d): GR.
(e): LXR, PPAR, RXR.
(f): PPAR, AhR, PR, FXR, LXR, PXR, TR, VDR, RXR.
(g): PXR.

Attention should be paid in the interpretation of results to understand the specific basis and scope of the prediction for each endocrine pathway, taking into account the performance and the applicability domain of each *in silico* predictive model when drawing conclusions. For more details on the software/expert systems, see Appendix D.

Read-across approaches and categories

Substances that have physicochemical, toxicological and ecotoxicological properties that are similar or follow a regular pattern as a result of structural similarity, may be considered as a group, or 'category' of substances. These similarities may be due to a number of factors:

- Common functional group (i.e. chemical similarity within the group).
- Common precursors and/or likelihood of common breakdown products through physical and/or biological processes which result in structurally similar degradation products (i.e. similarity through (bio)transformation).
- A constant pattern in the changing of the potency of the properties across the group (i.e. of physicochemical and/or biological properties).

Thus, read-across is a data-gap filling technique that uses known endpoint data of a substance (source substance(s)) for inferring the same type of endpoint data for a similar substance (target substance(s)). In principle, there is no particular aspect of read-across for predicting ED activities that needs to be addressed differently from other read-across as the key point remains a robust justification, see ECHA Guidance (ECHA, 2008, 2017c,d). One of the main applications of read-across within the field of ED may correspond to the inference of a putative MoA from other substances within a group of substances which have the same MoA (e.g. aromatase inhibition), or even to infer adverse effects from one chemical to another. This type of read-across may be useful when assessing the overall coherence of the data set or when determining the KEs in a postulated MoA.

As an adaptation of the data requirements according to Annex IV, Section 1.5 of the BP Regulation,[1] read-across approaches can use relevant information from analogous ('source') substances to predict the properties of 'target' substances. If the grouping and read-across approach is

applied correctly, experimental testing can be reduced as there is no need to test every target substance.

If a read-across approach is successful, the study conducted with the source substance is read across as a whole to the target substance. In such cases, relevance and reliability for the source study should be assessed as if the study was conducted with the target substance. In addition, the uncertainty related to the use of an alternative method should be separately addressed.

4.2. *In vitro* test methods

Disruption of the endocrine system can be a consequence of interference with hormone receptors, their downstream signalling, their transporters, non-classical receptors or interaction with key enzymes involved in the regulation of hormone levels. *In vitro* assays can provide valuable information on potential interference at the cellular level (e.g. by responding to chemicals that bind to these receptors), on the regulation of the downstream signalling or on change in hormone production and conversion, assuming that the compound can reach the cellular target *in vivo* in a relevant amount. *In vitro* assays can also support the strength of the evidence that an observed adverse effect *in vivo* might be produced via a particular endocrine MoA. The results obtained from validated and non-validated *in vitro* test methods can be used in combination with other data in a WoE approach. Specifically, *in vitro* tests can provide mechanistic information when assessing the toxicological properties of chemicals. Positive *in vitro* results indicate a potential of ED concern *in vivo* and may inform whether further (targeted) testing may be necessary. In addition, positive and negative findings can be used when considering the grouping of chemicals in read-across and category approaches (see Section 4.1).

In vitro assays providing data about the selected endocrine pathways are captured under level 2 of the OECD CF for the testing and assessment of ED (OECD, 2018b). The assays currently listed in the OECD CF level 2 are specifically those that detect one particular endocrine modality only, focusing on the estrogenic or androgenic pathway, as well as impacts on steroidogenesis (see Table 12). However, compounds might be able to act via more than one mechanism. Therefore, no single *in vitro* test can be expected to detect all types of endocrine activity and a battery of tests would usually be carried out.

Defined approaches are a particular case of combining tests and/or non-test methods in which the tests that need to be carried out and the way in which the data is interpreted are predefined. Defined approaches provide a means of integrating multiple sources of data, including non-test methods. One example of a particular defined approach is the *ToxCast Estrogen Receptor (ER) model* which integrates 18 *in vitro*, high-throughput ToxCast screenings assays measuring receptor (ER) binding, dimerisation, chromatin binding, transcriptional activation, and ER-dependent cell proliferation (Judson et al., 2015). The 18 *in vitro* assays provide comprehensive pathway coverage for the biology of the ER signalling pathway (Browne et al., 2015). US-EPA is accepting the ToxCast ER model for 1812 chemicals as alternatives for EDSP tier 1 ER binding, ER transactivation, and uterotrophic assays. The ToxCast ER model scores ≥ 0.1 were considered positive, negative scores $= 0$, and model scores ($0 <$ AUC < 0.1) were considered inconclusive (Browne et al., 2015). Given that ToxCast raw data are publically available the bioactivity summary of the 18 *in vitro* ToxCast screenings assays and the result of the ToxCast Estrogen Receptor (ER) model should be included the assessment report.

Assays that are designed to detect estrogens and androgens usually detect activation of (one or more of) the receptor(s) involved. These assays can generally be divided into three main categories, according to their working principle: binding assays, proliferation assays and transactivation assays. Binding assays reflect the ligand-receptor interaction which is the initial step of the signalling pathway, and allow a quantification of the direct interaction of a substance to specific receptors. However, binding assays cannot determine whether the binding of the ligand to the receptor will result in activation or inhibition of receptor activity. In proliferation assays, cells grow (proliferate) as a consequence of activity on a specific (endocrine) pathway. Transactivation assays can identify chemicals that can bind to and consequently activate a specific receptor, as the cells produce a reporter gene product that can easily be quantified (e.g. luciferase, a fluorescent protein or β-galactosidase) following the activation of a specific receptor. Proliferation assays and transactivation assays can in principle differentiate between (partial) agonists (when tested in isolation) and antagonists (when tested in combination with a known agonist) although the *in vivo* (ant)agonistic effect might differ due to, for example, receptor subtypes, receptor tissue distribution or background activity.

Assays that provide information on steroidogenesis are not based on activation of a specific receptor. These assays either utilise cells that express one or more of the enzymes involved in steroidogenesis or

utilise, for example, microsomes that contain these enzymes. By chemically analysing the conversion rate of specific steroids, information can be obtained on the potential interference. While the current assays utilise human H295R cells (OECD TG 456; OECD, 2011c) and/or enzymes (US EPA OPPTS 890.1200; US EPA, 2009b), the key steps in the steroidogenic pathways relevant for androgen and estrogen synthesis are well conserved across taxa and therefore results in this assay are likely to be relevant to other vertebrate species. However, differences exist in the (preferred) steps making up the steroidogenesis pathway across species and stages of development (Payne and Hales, 2004; Scott et al., 2009), which should also be taken in account.

Different types of assays are available to study thyroid hormone dysregulation, although none of these assays is currently available as a test guideline. These assays target specific aspects of thyroid action, including assays addressing thyroid hormone production (e.g. interference with the sodium–iodide symporter, thyroperoxidase or iodothyronine deiodinases), transport (e.g. binding to thyroid hormone transport proteins like transthyretin or thyroxine-binding globulin) or the cellular response (e.g. thyroid receptor transactivation assays).

Many of the *in vitro* assays that are designed to provide information on an endocrine MoA utilise human or mammalian cell lines, although other cell lines (e.g. yeast, fish) are also used. Due to the high level of conservation of the endocrine system and receptor homology across the vertebrates, as well as the key enzymes involved, it is assumed that results of such *in vitro* assays, while often based on mammalian cells, can generally provide information applicable to both humans and other vertebrates. This assumption has been shown true especially for estrogenic compounds of moderate to high affinity. However, some compounds exhibiting low binding affinity to human estrogen receptor 1 ($< 0.001\%$) showed higher binding affinity for ERα from poikilothermic vertebrates, specifically fish and reptiles (Ankley et al., 2016).

Currently, only a few assays have OECD-adopted test guidelines, although several relevant assays are under consideration for test guideline development. It is therefore expected that much of the *in vitro* data will be obtained from the scientific literature and will be from non-test guideline methods. While preference might be given to studies which are shown to be reliable (e.g. test guideline studies), data generated by other relevant *in vitro* assays should always be considered, provided that the principle of the assay is clearly described. However, it is acknowledged that in many cases, information on the robustness and reproducibility will be unavailable (e.g. by using the criteria set out in the performance-based test guidelines for transactivation assays or validation principle as addressed in the OECD guidance document on good *in vitro* method and practices (GIVIMP (OECD, 2018a)). An OECD guidance document is in place on the reporting of non-standardised *in vitro* assays (i.e. non-test guidelines) (OECD, 2017a) in order to encourage the provision of all relevant data to allow, as far as possible, an independent evaluation of the reliability and relevance of a particular assay. Such an evaluation might be based on the OECD performance-based test guidelines that are valid for, and can more easily be extended to encompass, multiple assays. Performance-based test guidelines are now in place for ER binding assays (OECD TG 493 (OECD, 2015e)) and ER transactivation assays (OECD TG 455 (OECD, 2012c)), while a performance-based test guideline for AR transactivation assays is in development.

Table 12: Parameters in OECD CF level 2 '*in vitro* mechanistic', for which guidance is provided in OECD GD 150

Test guideline	**OECD TG**	**455**	**493**		**458**		**456**
	US EPA OPPTS		**890.1250**	**890.1150**		**890.1200**	
Species/*in vitro* test system		ER TA (human) cells expressing ERα	Binding to rat (EPA) or human (OECD) estrogen receptor	Binding to rat androgen receptor	AR TA (human AR-EcoScreen™) cell line	Human recombinant microsomes	Human H295R cells
Indicative of: (a)		E	E	A	A	S	S
Androgen receptor binding/transactivation				X	X		
Aromatase						X	
Estrogen receptor binding/transactivation		X	X				
Steroidogenesis (oestradiol and/or testosterone synthesis)							X

(a): Based on OECD GD 150, indicative of: the (E)strogen- (A)ndrogen- (S)teroidogenesis- or (T)hyroid modalities.

4.3. *In vivo* test methods

This section describes the *in vivo* test methods and the parameters measured with these test methods which are relevant to support the identification of ED-relevant effects. Based on the grouping of parameters explained in Section 3.1.1, the parameters considered in this section are those from the following groups:

- *In vivo* mechanistic
- 'EATS-mediated'
- 'sensitive to, but not diagnostic of, EATS'.

A list of relevant parameters and the corresponding *in vivo* test methods where these parameters are measured is provided in Sections 4.3.1 and 4.3.2, depending if a parameter is measured in a mammalian or non-mammalian test, and is tabulated in Tables 13, 14, 15, 16 and 17.

The list of parameters related to general adversity, which are not listed in OECD GD 150, mainly comprises parameters indicative of general systemic toxicity, e.g. signs of animal stress, mortality, changes in body weight and food consumption.

The relevant standard *in vivo* test methods are described in the levels 3–5 of OECD CF. Level 3 assays are screening assays designed to detect possible endocrine-disrupting activity and to provide clear answers about the ability of a chemical to interact with 'EATS-mediated' modalities in the life stage tested, e.g. by looking at alterations in endocrine-sensitive tissues. They are designed to be highly responsive; in some cases castrated or ovariectomised rats without an intact hypothalamic–pituitary–gonadal (HPG) axis or other immature animal models are used, which are therefore unable to compensate fully for endocrine perturbations.

However, these level 3 assays are incapable of revealing the full spectrum of possible ED effects, since animals with minimal endogenous oestrogen/androgen production are exposed during a short period of time, covering only a limited part of their life cycle, which may not cover the most sensitive window of exposure, and do not allow for examination of delayed effects.

Regarding methods at levels 4 and 5, they are mainly non-acute test methods and especially test methods on developmental toxicity, reproductive toxicity, carcinogenicity and (sub)acute and (sub) chronic repeated dose toxicity for human health evaluation and chronic toxicity tests on fish, amphibians and birds for non-target organism evaluation.

Some limitations of these test guidelines may be due to their design, such as: lack of exposure during sensitive window(s), difficulty to detect delayed effects, (too) short exposure duration, or low statistical power due to a low number of animals.

The focus of this GD is on EATS modalities, however, it should be acknowledged that certain test guidelines allow for the detection of other endocrine modalities (e.g. disruption of the pancreas can be detected in the OECD TG 408 based on the analysis of organ weight, pathology and histopathology).

4.3.1. Mammalian

4.3.1.1. OECD CF level 3 tests

Information on a possible MoA of endocrine-disrupting compounds can be obtained by using mechanistic assays, i.e. assays that are designed to provide information on a specific endocrine axis. In general, these assays are designed to provide simple yes/no answers to the ability of a compound to interact with a specific endocrine pathway (EATS).

Two methods are currently listed regarding mammalian toxicology: the uterotrophic assay (OECD TG 440 on estrogenic effects (OECD, 2007d) and OECD GD 71 on anti-estrogenic effects (OECD, 2007b)); and the Hershberger assay (OECD TG 441 (OECD, 2009d) and OECD GD 115 (OECD, 2009a) on the weanling Hershberger assay for (anti-) androgenic properties (OECD, 2009a)).

The list of relevant parameters, based on OECD GD 150 and JRC screening methodology, is shown in Table 13.

It should be noted that level 3 tests using intact (immature) animals might also provide (additional) evidence of adverse effects relevant for individuals before puberty.

Uterotrophic assay (OECD TG 440, OECD GD 71, CF level 3)

The uterotrophic assay is designed to detect estrogenic and anti-estrogenic modalities OECD (2006c). The parameters measured are: uterine weight (wet and dry), as well as (optional) histopathological changes in the uterus and vagina. The assay is run on ovariectomised young adult female rats (with adequate time for uterine tissues to regress, and acclimatisation after surgery) or immature (after weaning and prior to puberty) ones, and allows the detection of weak and strong estrogens as well as anti-estrogens. The use of immature animals may allow the detection of substances acting via mechanisms other than ER-mediated ones, as the animals have an intact HPG axis, but the ability to detect these is limited. This test can also detect androgenic modalities. Indeed, aromatisable and non-aromatisable androgens have also been shown to increase uterine weight. It should be noted that progesterone and synthetic progestins may also give a positive response. Another important aspect to consider is that the immature model is more sensitive to body weight effects on the uterus than the ovariectomised one (where the uterine weight is affected by hormones like estrogen but not by the growth factors that regulate body size).

The uterotrophic assay is a short-term assay (3 days), using oral gavage or subcutaneous routes. For ovariectomised female rats, longer exposures are acceptable and may improve the detection of weakly active substances. The choice of the administration route should reflect the most relevant one for human exposure, and should be taken into account when interpreting results (considering ADME, e.g. considering by-pass of first pass hepatic metabolism in case of subcutaneous injection).

Both methods (intact and ovariectomised animals) have been shown to be reliable and repeatable in intra- and interlaboratory studies, presenting comparable sensitivity and reproducibility (OECD, 2006a,b; Schapaugh et al., 2015). When using the ovariectomised animals, care should be given to ensure that no ovarian tissue remains, as it can produce endogenous estrogen and retard the regression of the uterine weight.

Hershberger assay (OECD TG 441, OECD GD 115, CF level 3)

The Hershberger assay detects androgenic and anti-androgenic modalities. The detection of (anti-) androgenic activity is based on the measurement of the weights of ventral prostate, seminal vesicles (plus fluids and coagulating glands), Levator ani/bulbocavernosus muscle complex (LABC), paired Cowper's glands and glans penis. In the intact weanling assay, the weight of epididymides should also be measured.

Other optional organ weight measurements are, for example, paired adrenal and testis weights. Serum hormones can also be optionally measured, informing on other modalities, such as the thyroid hormones (triiodothyronine (T3) and thyroxine (T4)), luteinising hormone (LH), follicle-stimulating hormone (FSH) and testosterone. The weanling assay does not include glans penis.

The assay uses immature weanling (OECD GD 115) or castrated peripubertal (OECD TG 441) male rats. It has been designed to be sensitive, and can detect weak and strong AR modulators and 5-alpha-reductase inhibitors. However, it has been shown that the use of immature rats seems not to consistently detect weak anti-androgenic chemicals (OECD, 2009a,d).

The intact HPG axis of immature animals could allow the detection of substances acting through this axis. However, the immaturity of the animals added to the co-administration of testosterone in the anti-androgen test, makes this unlikely (OECD GD 150).

The Hershberger assay can discriminate between anti-androgens acting through AR antagonism or through inhibition of the 5-α-reductase. The enzyme inhibitors will have a more pronounced effect on the ventral prostate. It should be noted that the growth of sex accessory tissues can also be induced by non-androgenic modalities, such as through potent estrogens or chemicals affecting steroid metabolism. However, these non-androgenic modalities are unlikely to affect the five male accessory tissues concomitantly. For a substance to be considered as a positive androgen agonist or antagonist, two or more target organ weights should be statistically significantly increased (or decreased, in the case of antagonism), and all the target tissues should display some degree of increased (or reduced, for antagonism) growth.

The weights of the optional organs (adrenal) provide information not only on androgen modality, but also on systemic toxicity. Measurement of LH and FSH levels provide indication of disturbance of the hypothalamic-pituitary function. Serum T4 and T3 measures would provide useful supplemental information about the ability to disrupt thyroid hormone homeostasis. Although the test guideline states that, 'with regard to serum hormone level, testosterone levels are useful to determine whether the test substance induces liver metabolism of testosterone, lowering serum levels, which could otherwise be misinterpreted as an anti-androgenic effect', in the context of this guidance, a decrease in hormone level occurring through induced liver metabolism is considered endocrine related.

The Hershberger assay is a short-term assay (10 days), using oral gavage or subcutaneous injection. The choice of the administration route should reflect the most relevant one for human exposure, and should be taken into account when interpreting results (considering adsorption distribution metabolism and excretion).

Guidance on the interpretation of the parameters measured in the uterotrophic and Hershberger assays as provided by OECD GD 150 is presented in Table 13. All of the relevant parameters listed from all the assays have been categorised according to one or more of the EATS pathways on which they are informative.

Table 13: Mammalian – parameters '*in vivo* mechanistic' (highlighted in orange)

Test guideline		**OECD TG 440 + OECD GD 71 (Level 3)**	**OECD TG 441 + OECD GD 115 (Level 3)**
Test duration		3 days	10 days
Life stages		Immature females (after weaning and prior to puberty) or young adult females after ovariectomy	Immature males (after weaning and prior to puberty) or young adult males after castration
Species/*in vitro* test system		Rat	Rat
Parameter name	**Indicative of:** [a]		
Adrenals weight[b]	N		x (optional)
Cowper's glands weight	A		X
Epididymis weight[b]	E, A, S		X
Oestradiol level[c]	E, A, S		X
FSH level[b]	E, A, S		x (optional)
Glans penis weight	A		X
Keratinisation and cornification of vagina	E	X	
LABC weight[b]	A		X
LH level[b]	E, A, S		x (optional)
Liver weight[d]	T		x (optional)
Proliferation of endometrial epithelium	E	X	
Prostate weight[b]	A		X
Seminal vesicles weight[b]	A		X
Steroidogenesis (genes/ enzyme changes)[c]	E, A, S		X
T3 and T4 level[b]	T		x (optional)
Testis weight[b]	E, A, S		X
Testosterone level[b]	E, A, S		x (optional)
Uterus histopathology[b]	E	X	
Uterus weight[b]	E, A	X	
Vaginal opening	E, A	X	

(a): Based on OECD GD 150, indicative of: the (E)strogen-; (A)ndrogen-; (S)teroidogenesis- or (T)hyroid modalities; (N)ot assignable to a specific modality.

(b): These parameters are also listed in Table 14, which lists 'EATS-mediated' parameters. The reason is that these parameters are measured in tests which are part of OECD CF level 3 (which provide '*in vivo* mechanistic' information) and in tests from OECD CF level 4/5 (which provide 'EATS-mediated' information).

(c): These parameters are not listed in OECD GD 150. They have been reported based on the JRC screening methodology to identify potential ED (JRC, 2016). The reason they are included in this table is that these parameters are frequently measured in studies available in scientific literature and they provide information relevant to endocrine activity through EATS modalities.

(d): This parameter is considered T-mediated, only when a change is observed in combination with other thyroid-related endpoints.

4.3.1.2. OECD CF level 4 and 5 tests

Many effects relevant for humans and wild mammals are identified using mammalian assays that are listed under levels 4 and 5 in the OECD CF. Assays at level 4 can provide a more comprehensive assessment of the potential or actual endocrine-disrupting effect than the level 3 assays (see Section 4.3.1.1), because they are designed to be sensitive to more than one MoA (whereas level 3 assays (even if sensitive to several MoAs) have been developed to specifically investigate a selected modality). All these assays cover different periods of susceptibility, but no current guideline covers the full lifecycle from *in utero* to old age, to allow investigation of early life exposure on effects manifested only later in life. The developmental and reproductive toxicity studies at level 5 are considered to provide more comprehensive data on adverse effects on endocrine relevant endpoints over more extensive parts of the life cycle of the organism, adding weight to the overall WoE obtained from level 3 and 4 assays. In

addition, some level 4 and 5 tests also include parameters indicative of endocrine activity. The list of relevant parameters, based on OECD GD 150 and the JRC screening methodology, is shown in Table 14.

Repeated dose 28-day oral toxicity study in rodents (OECD TG 407, OECD CF level 4)

The 28-day repeat dose toxicity test (OECD TG 407 (OECD, 2008)) has been validated using young adult animals. It was revised in 2008 to include some endocrine parameters. However, the sensitivity of the assay is not sufficient to identify all 'EATS-mediated' parameters or parameters 'sensitive to but not diagnostic of, EATS modalities'.

According to OECD GD 150, the validation of the assay showed that it identified strong and moderate ED acting through the ER and AR, and ED weakly and strongly affecting thyroid function, as well as steroidogenesis inhibition. It was relatively insensitive to weak ED acting through the ER and AR. In any case, it has to be borne in mind that owing to the low power of the study (5 animals/group), the window of exposure and the parameters tested, only positive results can be interpreted as being indicative, whereas a negative outcome is not conclusive of no effect. Dosing should begin as soon as possible after weaning and, in any case, before the animals are 9 weeks old.

When interpreting the histopathological data of the ovaries (follicular, thecal and granulosa cells), uterus, cervix and vagina, possible asynchrony of the oestrus cycle should be taken into account. Indeed, subtle endocrine effects by chemicals with a low potency for affecting sex hormone homeostasis may be identified by disturbance of the synchronisation of the oestrus cycle in different tissues and not so much by frank histopathological alterations in female sex organs (OECD, 2008).

Two similar tests exist using dermal (repeated dose dermal toxicity: 21/28-day study, OECD TG 410 (OECD, 1981a)) or inhalation (subacute inhalation toxicity: 28-day study, OECD TG 412 (OECD, 2017c)) exposures.

Preferred species: rat.

Repeated dose 90-day oral toxicity study in rodents (OECD TG 408, CF level 4)

Originally, the assay has not been designed to detect ED. However, the aim of the update of 2018 was to add endocrine-sensitive endpoints to improve the detection of potential endocrine activity of test chemicals. The parameters added are mainly related to thyroid: measurements of thyroid weight, T3, T4 and thyroid-stimulating hormone (TSH), low-density lipoproteins (LDL) and high-density lipoproteins (HDL) (as these parameters are directly controlled by thyroid hormone action and contribute to evidence of thyroid effects). Other parameters have been added as optional: sperm parameters (sperm morphology, sperm motility, sperm number) and hormone measurements (FSH, LH, oestradiol, testosterone). Assessment of the optional measurements may be considered if existing information for the test chemical or similar chemicals suggests potential to influence these or can be triggered by observations from required measurements collected as part of this guideline.

Dosing should begin as soon as possible after weaning and, in any case, before the animals are nine weeks old. As the dosing period is longer than in the OECD TG 407, and the number of animals per group is larger, OECD TG 408 (OECD, 2018c) is likely to be more sensitive than OECD TG 407.

Preferred species: rat.

In addition, three other tests (not in the OECD CF as published in 2012) cover some of the above-mentioned parameters: repeated dose 90-day oral toxicity study in non-rodents (OECD TG 409 (OECD, 1998), subchronic dermal toxicity: 90-day study (OECD TG 411 (OECD, 1981b)), and subchronic inhalation toxicity: 90-day study (OECD TG 413 (OECD, 2017d)).

Prenatal developmental toxicity study (OECD TG 414, CF level 4)

The prenatal developmental toxicity study (OECD TG 414 (OECD, 2018d)) involves repeated dosing of pregnant females and therefore potential exposure of the developing fetus. This test was not specifically designed to detect EDs; however, a recent update (2018) has added various endocrine-related parameters, including EATS-mediated parameters such as AGD measurement or thyroid hormones measurement, thyroid weight and histopathology. It should be noted that these parameters are meant to be measured in rats (and not in rabbits), and that thyroid measurements are intended in the dams (and not the fetuses).

In this study, the test substance is administered daily from implantation (e.g. day 5 post-mating) to the day prior to scheduled caesarean section (treatment may be extended to include the entire period of gestation).

Preferred species: rat (rodent) and rabbit (non-rodent).

One-generation reproduction toxicity study (OECD TG 415, CF level 4)

With respect to apical endpoints, this assay provides a more thorough assessment of effects on reproduction and development than OECD TG 421/422, but is not as comprehensive as the reproductive studies in level 5. Moreover, it has also not been updated with endocrine-sensitive endpoints. For example, it does not include 'EATS-mediated' parameters such as sexual maturation; vaginal opening or preputial separation.

This test can detect adverse apical effects which may be caused by endocrine modalities other than EATS, such as disruption of the HPG axis or other hormone systems.

The dosage period in this assay is longer than the OECD TG 421 and 422, starting 10 weeks prior to mating for male rats (8 weeks for mice), representing one complete spermatogenic cycle, and from at least 2 weeks prior to mating up to weaning for females.

The OECD TG 415 (OECD, 1983) includes only one cycle of mating. It is intended to be used with the rat or mouse.

It should be noted that this test guideline has been deleted from the OECD list (as considered obsolete). However, it is mentioned in this guidance as it can still be found in some dossiers.

Reproduction/developmental toxicity screening test (OECD TG 421) and combined repeated dose toxicity study with the reproduction/developmental toxicity screening test (OECD TG 422) (CF level 4)

The reproduction/developmental screening tests OECD TG 421 (OECD, 2016a) and 422 (OECD, 2016b) are included in level 4 as supplemental tests because they give limited but useful information on interaction with endocrine systems. Both test guidelines were updated in 2016 to incorporate parameters suitable to detect 'EATS-mediated' parameters as well as parameters 'sensitive to, but not diagnostic of, EATS', in particular because of the sensitive periods during development (pre- or early postnatal periods) covered by these test guidelines. In these tests, males are dosed for a minimum of 4 weeks (including 2 weeks prior to mating), and females from 2 weeks prior to mating up to 13 days post-delivery. In view of the limited premating dosing period in males, fertility may not be a particularly sensitive indicator of testicular toxicity. Therefore, a detailed histological examination of the testes (i.e. staging) is essential.

Regarding TH, measurement of T4 is mandatory in the male parent animals. In pups, T4 should be measured at postnatal day (PND) 4 (if number of pups allows) and at PND 13. Other hormones may be measured if relevant. Preferably, T4 and TSH should be measured as 'total' hormone (free and bound).

Preferred species: rat.

Developmental neurotoxicity study (OECD TG 426, CF level 4)

The developmental neurotoxicity study (OECD TG 426 (OECD, 2007c)) involves repeated dosing of pregnant females and therefore potential exposure of the developing fetus. It has not been specifically designed to detect EDs, but it includes parameters relevant to endocrine disruption.

The developmental neurotoxicity assay specifies a dosing period of the dam from time of implantation (gestational day 6) throughout lactation (PND 21). It is generally assumed that exposure of the pups occurs through the maternal milk; however, direct dosing of pups should be considered in those cases where there is a lack of evidence of continued exposure to offspring. Evidence of continuous exposure can be retrieved from, for example, pharmacokinetic information, offspring toxicity or changes in biomarkers.

The assay provides data, on the potential functional and morphological effects on the developing nervous system of the offspring that may arise from exposure *in utero* and during early life. Dams are tested to assess effects in pregnant and lactating females and may also provide comparative information (dams vs offspring). Offspring are tested during postnatal development and adulthood for gross neurologic and behavioural abnormalities, physical development, behavioural ontogeny, motor activity, motor and sensory function; learning and memory; brain weights and neuropathology.

It has been shown that developmental neurotoxicity can arise via thyroid disruption (Fan and Wu, 2016; Ghassabian et al., 2014). Furthermore, sex hormones play an important role in development of sexual dimorphism of the brain. Substances interfering with the sex hormone balance may therefore also affect the developing brain. In this test, the exposure of the fetus (which may be a sensitive life-stage for endocrine-disrupting effects) and the duration of dosing make it an assay that can be used

when assessing effects relevant to endocrine disruption. In addition, it provides data on effects related to reproduction and development, in particular the EATS-mediated parameters of sexual maturation.

Preferred species: rat.

Combined chronic toxicity/carcinogenicity studies (OECD TG 451-3, CF level 4)

These three tests measure chronic toxicity (general toxicity and carcinogenicity), dosing animals between 12 months and most of lifespan (18 months mouse, 24 months rat). These tests have not been designed to detect ED, but do measure some 'EATS-mediated' parameters and some parameters 'sensitive to, but not diagnostic of, EATS' modalities. OECD TG 453 (OECD, 2009g) was revised in 2009 and replaced OECD TG 451 (OECD, 2009e). TG 452 (OECD, 2009f) (chronic toxicity study) and TG 453 are likely to be more sensitive than the 28-day and 90-day tests because of the extended dosing period and the larger number of animals per group. However, they do not include some sensitive endpoints (e.g. thyroid hormones, functional measurement of oestrous cyclicity) included in the updated 28-day test. In any case, attention must be paid to dose levels and dose spacing between the different study types.

All tests should preferably use rodent species. Dosing of animals should start as soon as possible after weaning, and preferably before they are 8 weeks old. These tests are the only ones that cover the ageing of animals; however, the dosing period does not include early life stages.

Peripubertal male and female assays (OPPTS 890.1500 and 890.1450, CF level 4)

The pubertal development and thyroid function assay in peripubertal male (OPPTS 890.1500 (US EPA, 1996)) or female (OPPTS 890.1450 (US EPA, 2009c)) rats are designed to detect chemicals interfering with the androgen (male test), estrogen (female test) and thyroid pathways, as well as steroidogenesis and the HPG axis. The male assay can also detect ER-mediated effects, but the accuracy of this is unknown (OECD, 2018b).

Both tests will also detect chemicals that alter pubertal development via changes in the HPG axis.

In these assays, the animals are dosed during their sexual maturation. The limitations of these assays, noticed during their validation, are that no chemical was shown to be completely negative in the assay, and that it does not detect specific aromatase inhibitors. The sensitivity of the assays for ER/AR agonists and antagonists is less than that of the uterotrophic and Hershberger assays.

Two-generation reproduction toxicity test (OECD TG 416, CF level 5)

The two-generation reproduction toxicity test (OECD TG 416 (OECD, 2001)) assesses endocrine-related parameters in a less comprehensive way that the other level 5 assay (OECD TG 443 (OECD, 2012b)), and although some 'EATS-mediated' parameters like oestrous cyclicity and primordial follicle counts were included in the 2002 version, it does not include 'EATS-mediated' parameters like nipple retention. The full list of measured parameters can be found in Table 14.

This test can detect effects resulting from (anti-)estrogenic, (anti-)androgenic, thyroid and steroidogenic modalities. However, other endocrine modalities can also be detected, such as chemicals acting on the HPG axis or other hormone systems.

Males of the parental generation are dosed during growth, and for at least one complete spermatogenic cycle to allow adverse effects on spermatogenesis to be more easily detected. Females of the parental generation are dosed during growth and for several complete oestrus cycles (in order to detect any adverse effects on oestrus cyclicity), throughout pregnancy until weaning of offspring. Dosing of F1 offspring continues during their growth into adulthood, mating and production of an F2 generation, until the F2 generation is weaned. Offspring are exposed during all vulnerable periods of development. Late effects becoming manifest after weaning are partly covered in young adults, especially in relation to reproductive function, but later ones (e.g. premature reproductive senescence) are not.

Preferred species: rat.

Extended one-generation reproductive toxicity study (OECD TG 443, CF level 5)

The extended one-generation reproductive toxicity study (OECD, 2012b) has been designed to cover specific life stages rarely covered by other assays (with the exception of OECD TG 416) and to test for effects that may occur as a result of pre- and postnatal exposure to chemicals. It incorporates additional EATS-sensitive parameters, when compared to the OECD TG 416. The dosing is continuous, prior to and during mating, and throughout production of the subsequent generation(s). Although the study was developed to cover apical effects arising from either endocrine or non-endocrine activities, it has also been designed to include some endocrine parameters ('EATS-mediated', and 'sensitive to, but

not diagnostic of, EATS') in the F1 generation (in both juvenile and adult life stages) such as nipple retention, the AGD index at birth, age of vaginal opening and preputial separation. According to the test guideline, the study design should include by default the evaluation of the fertility of parental animals and postnatal development of F1 animals until adulthood, as well as cohorts specifically for the investigation of developmental neurotoxicity (DNT) or developmental immunotoxicity (DIT). The rationale for omission of these cohorts should be given. An option for extending the assay to include an F2 generation by mating the F1 animals is included in the test guideline. Selection of this option should reflect current knowledge for the chemical being evaluated, as well as the needs of various regulatory authorities. Additional clinical chemistry endpoints (such as measurement of thyroid hormones and TSH levels) usually measured in repeat dose studies have also been included in the study design.

The parental (P) generation is dosed for a defined premating period (minimum of 2 weeks) and a 2-week mating period. P males are further treated at least until weaning of the F1, for a minimum of 10 weeks in total. Treatment of the P females is continued during pregnancy and lactation until termination after the weaning of their litters (i.e. 8–10 weeks of treatment). The F1 offspring is further dosed from weaning to adulthood. Therefore, OECD TG 443 (together with the older OECD TG 416) is the only current OECD guideline that can provide information on the effects of ED exposure during the post-natal (juvenile) development, from weaning through to puberty and sexual maturity. If a second generation is assessed, the F1 offspring will be maintained on treatment until weaning of the F2, or until termination of the study. The pups will normally receive the test substance indirectly through the milk, until direct dosing commences for them at weaning. In diet or drinking water studies, the pups will additionally receive the test substance directly when they start to feed themselves during the last week of the lactation period. Modifications to the study design should be considered when excretion of the test substance in milk is poor and where there is lack of evidence for continuous exposure of the offspring. Therefore, analytical determination of the test substance in the dams' milk or its accumulation in certain regions of the pups, i.e. brain, and direct dosing of pups during the lactation period should be considered.

OECD GD 151 (OECD, 2013a) provides guidance on the design, conduct and interpretation of results of OECD TG 443. Guidance specifically related to endocrine disruption is given for some parameters, as described below.

TH levels have been demonstrated as critical for the maturation and function of the central nervous system. Measurement of T4 and/or TSH in parental and F1 offspring at various life stages to assess direct effects on thyroid function or indirect effects via the HPT axis is required. The measurement of both T4 and TSH can provide information on the MoA of the test chemical and its potential effect. The diurnal fluctuations of thyroid hormone levels should be taken into account, and appropriate measurement method should be used (see Appendices A and B). Changes in hormone levels should be evaluated in conjunction with any changes in thyroid gland weight and histopathology, as well as neurological or other developmental adverse effects.

The mammary gland has been shown to be estrogen-sensitive, particularly in males, and histopathological examination is among the parameters to be checked in adults and weanlings of both sexes. Development of the terminal end buds into differentiated structures is of particular interest (OECD GD 151). The test guideline recommends that parameters involving pup mammary glands of both sexes be included, when validated.

Decrease of anogenital distance and increase of nipple retention in male rats have been associated with exposure to an anti-androgen. Interpretation of anogenital distance should take (cube root of) body weight into account, through the calculation of anogenital distance index.

Vaginal opening and first vaginal oestrus are parameters sensitive to estrogen disruption. Exposure of the developing female to an estrogenic substance will likely cause a significant advancement of the age of vaginal opening, but not necessarily advance first ovulation. The same holds true for prepubertal androgen exposure, due to the presence of aromatase activity in the vaginal epithelium of immature rats. In most cases, environmental estrogens will cause early vaginal opening and a pattern of persistent vaginal oestrus, (i.e. pseudo-precocious puberty) which may or may not continue as the animal matures. Thus, evaluating the first vaginal oestrus following vaginal opening will provide information as to whether there are group/dose differences in the timing of these two events that would signal an abnormal progression through puberty. As indicated above, first oestrus may be affected in time proportional to the appearance of vaginal opening, or the two may be disconnected, indicating independent alterations in response to a test chemical within the vagina and the hypothalamic-pituitary control of first ovulation at puberty (OECD GD 151). It should be kept in mind when interpreting results of vaginal opening and first oestrus measurements, that body weight can influence these parameters. Another parameter which should be investigated in relation to effect on

oestrus cyclicity is uterus weight. Indeed, compounds that cause loss of cyclicity (e.g. estrogen antagonists, steroidogenesis inhibitors) may cause uterus atrophy and weight reduction.

The data from the DNT and DIT cohorts are also relevant to endocrine disruption. Indeed, it has been shown that the developing brain is a classical target of thyroid hormones (Fan and Wu, 2016; Ghassabian et al., 2014) while interaction of chemicals with the hypothalamic–pituitary–adrenal axis may affect both the developing immune and nervous systems. Furthermore, sex hormones play an important role in development of sexual dimorphism of the brain. Substances interfering with the sex hormonal signalling may therefore also affect the developing brain. Moreover, estrogens and androgens are involved in the development and regulation of immunity, as well as in sex-based disparities in immune responses (Cutolo et al., 2002; Adori et al., 2010; Arredouani, 2014; Trigunaite et al., 2015).

Preferred species: rat.

Table 14: Mammalian *in vivo* parameters – parameters '*in vivo* mechanistic' (highlighted in orange), parameters 'EATS-mediated' (highlighted in blue) and parameters 'sensitive to, but not diagnostic of, EATS' (highlighted in purple)

		Section A										Section B	
Test guideline		**OECD TG 407**	**OECD TG 408**	**OECD TG 414**	**OECD TG 415[f]**	**OECD TG 421**	**OECD TG 422**	**OECD TG 426**	**OECD TG 451-3**	**OECD TG 416[a]**	**OECD TG 443[a]**	**OPPTS 890.1500[a]**	**OPPTS 890.1450**
Test duration		28 days (plus 14 days recovery period)	90 days	From implantation to the day prior to the scheduled caesarean section (GD 6–20 in rodent, GD 7–28 in rabbits)	16–19 weeks	28 days in males and approximately 63 days in females	28 days in males and approximately 63 days in females	From GD 6 to PND 21	Between 12 and 18 months in mouse or 24 in rat	29 weeks	30 weeks	30 days	20 days
Life stages		Adult (P)	Adult (P)	Fetus	Adult (P) and F1	Adult (P) and F1	Adult (P) and F1	Fetus and F1	Adult (P)	Adult (P), F1 and F2	Adult (P), F1 and eventually also F2	Juvenile male	Juvenile female
Species/*in vitro* test system		Rat	Rat	Rat, rabbit	Mouse, rat	Rat	Rat	Rat	Mouse, rat	Mouse, rat	Rat	Rat	Rat
Parameter name	**Indicative of:** [a]												
Parameter name	**Indicative of:** [a]	**OECD TG 407**	**OECD TG 408**	**OECD TG 414**	**OECD TG 415[f]**	**OECD TG 421**	**OECD TG 422**	**OECD TG 426**	**OECD TG 451-3**	**OECD TG 416[e]**	**OECD TG 443[e]**	**OPPTS 890.1500[e]**	**OPPTS 890.1450**
Oestradiol level	E, A, S		x (optional)										
Follicle stimulating hormone (FSH) level[b]	E, A, S		x (optional)										
Luteinising hormone (LH) level[b]	E, A, S		x (optional)										
T3 and/or T4 level[b]	T	x (optional)	x	x (dams, rat)		x	x				x	x	X
Testosterone level[b]	E, A, S		x (optional)									x	
Thyroid-stimulating hormone level (TSH)	T	x (optional)	x	X (dams, rat)		x	x				x	x	X

Accessory sex organs histopathology	E, A, S		x			x			x				
Age at first oestrus	E, A												X
Age at balanopreputial separation	E, A, S							X		x	x	x	
Age at vaginal opening	E, A, S							X		X	x		X
Anogenital distance	E, A, S			X (rat)		X	x	X		X	x		
Cervix histopathology	E, A, S	X	x		X		x		X	X	x		
Coagulating gland histopathology	E, A, S	X	x		X		x		x	X	x		
Coagulating gland weight	E, A, S	X				x	x			X	x	x	
Colloid area (thyroid histopathology)	T	X					x (optional)					x	X
Cowper's gland weight	E, A, S					x (optional)	(optional)						
Epididymis histopathology	E, A, S	X	x		(optional)	X	X		X	X	x	x	
Epididymis weight [(b)]	E, A, S	X	x			X	X		X	X	x	x	
Oestrus cyclicity	E, A, S	X (optional; at necropsy by vaginal smears)	x			x	x			X	x		X

Parameter name	Indicative of: [a]	OECD TG 407	OECD TG 408	OECD TG 414	OECD TG 415[f]	OECD TG 421	OECD TG 422	OECD TG 426	OECD TG 451-3	OECD TG 416[e]	OECD TG 443[e]	OPPTS 890.1500[e]	OPPTS 890.1450
Follicular cell height (thyroid histopathology)	T	X					X			X		x	X
Glans penis weight	E, A, S					x (optional)	x (optional)						
Genital abnormalities	E, A, S			X	x	x	X			X	x		
HDL/LDL ratio[c]	T		x										
LABC weight[b]	E, A, S					x (optional)	x (optional)					x	
Liver weight[c]	T	X	x				X		x	x	x	x	X
Mammary gland histopathology (male)	E, A, S	x (optional)	x				X		x (optional)		x		
Mammary gland histopathology (female)	E, A, S	X	x						X		x		
Nipple development	A					x	X				x		
Ovary histopathology	E, A, S	X	x		X (optional)	x	X		X	X	x		X
Ovary weight	E, A, S	x (paired) (optional)	x			x (optional)	X		X	X	x		X
Oviduct histopathology	E, A, S		x		X (optional)						x		
Prostate histopathology (with seminal vesicles and coagulating glands)	E, A, S	X	x		X (optional)	x	X		x	X	x		
Prostate weight [b]	E, A, S	X	x			x	X			X	x	x	
Seminal vesicles histopathology	E, A, S	X	x		x (optional)		X		x	X	x		
Seminal vesicles weight[b]	E, A, S	X	x			x	X			X	x	x	
Sperm morphology	E, A, S		x (optional)							X	x		

Sperm motility	E, A, S		x (optional)							X	x		
Sperm numbers	E, A, S		x (optional)							X	x		
Testis histopathology	E, A, S	X	x		x (optional)	x	X		X	X	x	x	
Parameter name	**Indicative of:** [a]	**OECD TG 407**	**OECD TG 408**	**OECD TG 414**	**OECD TG 415**[f]	**OECD TG 421**	**OECD TG 422**	**OECD TG 426**	**OECD TG 451-3**	**OECD TG 416**[e]	**OECD TG 443**[e]	**OPPTS 890.1500**[e]	**OPPTS 890.1450**
Testis weight [b]	E, A, S	X	x			x	X		X	X	x	x	
Thyroid histopathology	T	X	x	X (dams, rat)		x (optional)	x (optional)		X	x (optional)	x	x	X
Thyroid weight	T	x (optional)	x	X (dams, rat)		x (optional)	x (optional)		X	x	x	x	X
Uterus histopathology (with cervix) [b]	E, A, S	X	x		X (optional)	x (optional)	X		X	x	x		X
Uterus weight (with cervix) [b]	E, A, S	X (optional)	x	x (gravid uterus)	X	x (optional)	X		X	x	x		X
Vagina histopathology	E, A, S	X	x		x (optional)		X		X	x	x		
Vaginal smears	E, A, S	x (optional)	x			X	X			x	x		
Adrenals histopathology	N	X	x				X		X		x		
Adrenals weight [b]	N	X	x				X		x	x	x	x	X
Auditory startle	N										x (DNT cohort)		
Brain histopathological examination	N							x			x (DNT cohort)		
Brain morphometric (quantitative) evaluation	N							x			x (DNT cohort)		
Brain weight	N	X	x				X		X	x	x		
Dystocia	N				X	X				x	x		
Fertility	N				X	X	X			x	x		
Fetal development (or physical development of the fetuses?)	N			X	X	X	X	x					

Functional observation battery [(d)]	N										x (DNT cohort)		
Gestation length	N			X	X	X	X	x		X	X		
Learning and memory in offspring	N							x					
Parameter name	**Indicative of:** [(a)]	**OECD TG 407**	**OECD TG 408**	**OECD TG 414**	**OECD TG 415**[(f)]	**OECD TG 421**	**OECD TG 422**	**OECD TG 426**	**OECD TG 451-3**	**OECD TG 416**[(e)]	**OECD TG 443**[(e)]	**OPPTS 890.1500**[(e)]	**OPPTS 890.1450**
Litter size	N			X	X	x	X	x		X	X		
Litter viability	N				x	x	x			x	X		
Litter/pup weight	N			X	x	X	X	x		X	X		
Motor activity	N							x			x (DNT cohort)		
Motor and sensory function	N							x					
Number of implantations, corpora lutea	N			X		X	X			X	X		
Number of live births	N				X	X	X			x	X		
Numbers of embryonic or fetal deaths and viable fetuses	N			X									
Number of ovarian follicles	N										X		
Pituitary histopathology	N	x (optional)	X		x (optional)		x		x		X		
Pituitary weight	N		X						x	x	X	x	X
Post-implantation loss	N			X		x	x			x	X		
Pre-implantation loss	N			X		x	x			x			
Presence of anomalies (external,	N			X	X	x	x			x	X		

visceral, skeletal)													
Pup development	N									x	X		
Pup survival index	N				X			x		x	X		
Reproduction	N					x	x						
Sex ratio	N			X	X	x	x	x		x	X		
Time to mating	N									x	X		
Tumour types	N								x				

The table is divided into two sections as reflected in OECD GD 150: Section A lists parameters from established tests which have been validated and published as OECD test guidelines; Section B lists parameters from tests that have not received full validation by OECD, or are in the process of OECD validation, or are guidelines which have been validated and published by non-OECD organisations.

GD: gestational day; LABC: levator ani-bulbocavernosus muscle; LDL/HDL ratio: low-density lipoprotein/high-density lipoprotein ratio; T3: triiodothyronine; T4: thyroxine.

(a): Based on OECD GD 150, indicative of: the (E)strogen-, (A)ndrogen-, (S)teroidogenesis- or (T)hyroid modalities; (N)ot assignable to a specific modality.

(b): These parameters are also listed in Table 13, which lists '*in vivo* mechanistic' parameters. The reason is that these parameters are measured in tests which are part of OECD CF level 3 (which provide '*in vivo* mechanistic' information) and in tests from OECD CF level 4/5 (which provide 'EATS-mediated' information).

(c): These parameters are considered T-mediated, only when a change is observed in combination with other thyroid-related endpoints.

(d): as described in Appendix A of the OECD TG 443.

(e): For OECD TG 416, OECD TG 443 and EPA OPPTS 890.1500, it should be noted that coagulating gland weight is in combination with seminal vesicles. Furthermore, for EPA OPPTS 890.1500, the prostate weight is provided for two separate sections (dorsolateral and ventral).

(f): OECD TG 415 is not listed anymore in OECD GD 150 because this test guideline has been deleted from the Test Guideline programme. However, this test guideline is kept in the table because it can still be found in some dossiers.

4.3.2. Non-mammalian

This section describes the *in vivo* test methods and the parameters measured with these test methods which are relevant to support the identification of ED for non-target organisms.

4.3.2.1. Parameters

Some parameters such as growth, sexual maturity, reproduction parameters (fecundity, gonadosomatic index (GSI)) and behavioural parameter are known to be sensitive to substances interfering with the sex hormone system or the thyroid hormone system (WHO/IPCS, 2002; OECD, 2004, 2011a). These parameters are not 'EATS–mediated' as they might be influenced by other endocrine and non-endocrine factors such as systemic toxicity or dietary influences, but can be used in a WoE approach to draw a conclusion on a specific endocrine pathway. It is therefore important to consider possible confounding factors and use a WoE approach when interpreting changes in a single or several studies.

Fecundity, for example, measured in terms of number of eggs/surviving female/day, is 'sensitive to, but not diagnostic of EATS' modalities. Changes in fecundity inform about apical effects on reproduction, which consequently inform about potential adverse effects at the population level. Abnormal behaviour or appearance might also be endocrine-mediated, i.e. territorial aggressiveness in genetic males or masculinised females has been observed in fathead minnows under androgenic exposure, and in zebrafish, the characteristic mating and spawning behaviour after the dawn onset of light is reduced or hindered by estrogenic or anti-androgenic exposure (OECD, 2009b, 2012a). However, abnormal behaviour or appearance could also be clinical signs of general toxicity, or due to other MoAs. Therefore, any adverse behavioural effects need to be assessed in a weight of evidence in order to ascertain if they are linked to an endocrine activity.

The parameters normally measured in non-mammalian *in vivo* test methods are detailed below.

Vitellogenin

Vitellogenin (VTG) is normally produced by the liver under estrogenic regulation as a precursor of yolk proteins in female fish, amphibians and birds (Slater et al., 1991). VTG is only produced at very low level in immature female and male fish under natural conditions, because they lack sufficient circulating estrogen; therefore, VTG measurement has been developed as a biomarker for endocrine activity. Induction of VTG production in male is a biomarker used to detect estrogenic compounds, whereas reduction of VTG in female may be indicative of sexual steroid synthesis modulation. VTG modulation can also be triggered by chemicals that interfere with the AR-mediated pathway (Kwon et al., 2005) (https://aopwiki.org/aops/23) and chemicals disrupting steroidogenesis activities. Therefore, changes in this biomarker are a well-established method that can be used to detect chemicals potentially interfering with the endocrine system, especially in fish, and has been integrated in several OECD test guidelines.

However, it should be kept in mind that a decrease in VTG may also be caused by overt or systemic toxicity and non-endocrine MoAs (e.g. hepatotoxicity) or by confounding factors such as diet or infection (Dang, 2016). Consequently, a decrease in VTG, while generally considered EAS-mediated, needs to be interpreted with caution in combination with other observations.

Spiggin

Spiggin is a glycoprotein produced in the kidneys of sexually mature male three-spined sticklebacks (*Gasterosteus aculeatus*) under androgen stimulation during their breeding season. It is the only known androgen-induced protein produced by the three-spined sticklebacks (Östlund-Nilsson et al., 2006). It is stored in the urinary bladder from which it is excreted and used as a cement to build up a nest in which the female lays her eggs. It is therefore not present in the kidneys of female fish under natural conditions, and its production in females means that they have been exposed to substances with androgenic properties (Andersson et al., 2007). This was the basis for the development of a screening test for androgen antagonism (OECD GD 148 (OECD, 2011a)), and for the development of another method based on the use of genetically modified medaka eleuthero-embryos (Sebillot et al., 2014). This method has recently been submitted for validation at the OECD (see the RADAR assay).

Secondary sex characteristics

Another parameter is the detection of male secondary sex characteristics (SSC) in female fish. In male fathead minnows (*Pimephales promelas*) and Japanese medaka (*Oryzias latipes*), SSC are

externally visible, quantifiable and responsive to chemicals interfering with the EAS pathways. When females are exposed to androgenic substances, they can develop male SSC. In particular, in fathead minnows, the number and rating of nuptial tubercles located on the snout of the female fish is recorded, while in females of medaka, the main marker of exogenous exposure to androgenic compounds is the number of papillary processes on the anal fin. Zebrafish (*Danio rerio*) also possess quantifiable SSC-like urogenital papillae and change in body colour but these characteristics have not been validated in standardised tests. A decrease in SSC in males may indicate an estrogenic or anti-androgenic MoA but can also be influenced by non-endocrine MoA; it should therefore be interpreted with caution and based on WoE and expert judgement (OECD, 2009b). There is ongoing debate on the consideration of SSC as an apical endpoint and about the relevance of this endpoint at the population level.

Sex ratio

There are two types of sex ratio: phenotypic and genetic sex ratio. The phenotypic sex ratio is determined in individual fish via the histological examination of the gonads and it is defined as female, male, intersex (both oocytes and spermatogenetic cells in one gonad) or undifferentiated (fish with gonads exhibiting no discernible germ cells). Change in the phenotypic sex ratio is a parameter reflecting sex reversal, and can in principle be affected by estrogens, anti-estrogens, androgens, anti-androgens and steroidogenesis inhibiting chemicals (Scholz and Kluver, 2009). The ability of a substance with a suspected specific endocrine MoA to change the sex ratio of fish should be considered during the choice of fish test species because some species are more susceptible to sex ratio changes caused by a specific endocrine mechanism than others.

The genetic sex is examined via genetic markers and can be determined in fish species such as Japanese medaka and the three-spined stickleback where this marker is present, as well as in the African clawed frog (*Xenopus laevis*). The presence of a genetic sex marker is a considerable advantage where the genetic sex can be individually linked to the phenotypic sex, because it allows individual phenotypic sex reversal to be confirmed, which increases the power of the sex ratio statistics. However, in some strains of medaka, the existence of some XX (genetic female) individuals has been shown to perfectly function as (phenotypic) male (Nanda et al., 2003). It has to be kept in mind that in some species, temperature (i.e. zebrafish) or other type of general stressors (Matthiessen and Weltje, 2015; Ribas et al., 2017) can also play a role in the sex determination and the sex ratio, which should be taken into account when interpreting the results, however this should not be an issue when testing under controlled laboratory condition.

Sex ratio determination is also foreseen in amphibians and birds test guidelines.

Gonadosomatic index

The gonadosomatic index (GSI) is the calculation of the gonad mass as a proportion of the total body mass. Changes in the GSI may provide additional information about the gonad maturation and spawning readiness (OECD, 2004). Reduction of the GSI in male fish is regarded as a sensitive parameter in reproductive studies with estrogenic substances (OECD, 2004). However, the GSI might also be influenced by androgenic, anti-estrogenic and anti-androgenic MoAs, and might also be influenced by non-EATS modalities. This parameter can also be impacted secondarily through the cortisol-mediated stress response endocrine pathway as it has been observed that female Mozambique tilapia (*Oreochromis mossambicus*) implanted with cortisol to simulate chronic stress had reduced oocyte size and the GSI (Foo and Lam, 1993). It should therefore not be considered as specifically 'EATS-mediated'. In addition, it must be considered that the GSI may substantially increase during a spawning season (Helfman et al., 1997), and that interindividual variation in ovarian weight can be high during the spawning cycle (OECD, 2004). The GSI is therefore a highly variable measure in fish and should be interpreted with caution. The GSI might also be relevant for amphibians (Polzonetti-Magni et al., 2004).

Gonad histopathology

Gonad histology can help to interpret effects on reproduction and can be performed on amphibians (OECD, 2015a,b) and fish (OECD GD 123 (OECD, 2010)) and could be relevant for birds.

With respect to the histological changes, according to the guidance document (OECD GD 123) on the diagnosis of endocrine-related histopathology in fish gonads (OECD, 2010), the following parameters are of primary diagnostic interest:

- In males: increased proportion of spermatogonia (early sperm cells), the presence of testis-ova, increased testicular degeneration, interstitial (Leydig) cell hyperplasia/hypertrophy.
- In females: increased oocyte atresia, perifollicular cell hyperplasia/hypertrophy, decreased yolk formation (aromatase inhibition and non-aromatisable androgens), changes in gonadal staging.

Although it has not been demonstrated that these parameters are specific to a particular endocrine MoA, increased spermatogonia in males have been associated with exposure to estrogenic compounds and perifollicular cell hyperplasia/hypertrophy in females has been associated with exposure to aromatase inhibitors and non-aromatisable androgen. Leydig cell hyperplasia in males has been associated with steroidogenesis-related activity (OECD, 2010, 2018b).

Other effects (such as a decreased proportion of spermatogonia, altered proportions of spermatozoa (mature sperm cells) and gonadal staging in males, or interstitial fibrosis, granulomatous inflammation in females) are of secondary diagnostic interest. Parameters of both primary and secondary interest may also be influenced by non-endocrine-mediated MoAs.

Thyroid histopathology

Thyroid histology is a valuable and sensitive diagnostic parameter for detecting the ability of a substance to interact with the HPT axis, particularly for thyroid system antagonism (Grim et al., 2009). With respect to the histological changes, according to the guidance document on amphibian thyroid histology (OECD, 2015a,b), the core criteria are the following: thyroid gland hypertrophy/atrophy, follicular cell hypertrophy and follicular cell hyperplasia. The severity grading scheme is semi-quantitative and employs a four-grade approach describing ranges of variation within assigned ordinal classes: not remarkable, mild, moderate and severe. The purpose of this severity grading approach is to provide an efficient, semi-objective tool for comparing changes (compound-related effects) among animals, treatment groups, and studies (Grim et al., 2009). The descriptors are based on relative differences from thyroid glands in control animals, and/or on the percentage of cells or tissue affected. In addition to the severity grade, qualitative changes associated with the lesions should be documented.

4.3.2.2. Fish

When choosing a study or interpreting the results, differences in the developmental biology of species must be considered. This is particularly true for fish, as various species with different sexual determination/differentiation process can be used for testing. Japanese medaka, for example, is a differentiated gonochorist that develops early directly to either male or female gonads and sex does not change after gonadal development. Hormonal influence (especially that of female hormones) in this species starts very early during prehatch development (OECD, 2004) and thus life stages under exposure need to be considered carefully while analysing test results. If effects on gonadal staging are analysed, the reproductive cycle of a species should be considered. Especially for fish that have only one breeding season, such as rainbow trout (*Oncorhynchus mykiss*), endocrine effects may be observed only during the process of maturing prior to spawning and may be missed at other times of the year.

Moreover, effects potentially related to EATS modalities may be only observable during specific windows of exposure like specific life stage (e.g. larvae, juvenile, adult) and/or during specific stages of the reproductive cycle (e.g. gonadal development and differentiation, recrudescence, oocyte growth, final maturation). Therefore, whether or not endocrine-mediated effects are observable highly depends on the life stage tested. For example, testis-ova might be induced in adult males as, at least in some species, the gonads remain bipotent, but sensitivity to testis-ova is usually highest during sexual differentiation of the gonad (Nakamura et al., 1998).

OECD CF level 3 tests

There are three fish *in vivo* assays which are placed at level 3 of the OECD CF that include both apical endpoint and information on the endocrine activity. These are the fish short-term reproduction assay (OECD TG 229 (OECD, 2012a)), the 21-day fish assay (OECD TG 230 (OECD, 2009b)) and its variant the androgenised female stickleback screen, published by OECD as a guidance document (OECD GD 148 (OECD, 2011a)). It should be noted that all three fish tests primarily give information on potential endocrine activity in adult fish, although some of those tests can also give information on relevant adverse effect (e.g. fecundity). Test conditions and measured parameters are briefly described below and summarised in Table 15. In addition, three other tests are currently under validation at the

OECD level, the EASZY test, an *in vivo* fish-based assay designed to quantify the estrogenic effect on fish in early life stages, the juvenile medaka anti-androgen screening assay (JMASA) and the RADAR assay, an *in vivo* fish-based assay designed to quantify the androgen axis activity in early life stages.

Fish short-term reproduction assay (OECD TG 229, CF Level 3)

In the OECD TG 229 fish short-term reproduction assay (OECD, 2012a), sexually mature male and spawning female fish are exposed to a chemical for 21 days after a recommended pre-exposure period of 7–14 days. Two parameters are measured in both males and females: VTG (*in vivo* mechanistic) and SSC (EATS-mediated). Induction of plasma VTG levels in male fish allows to detect chemicals with an estrogenic MoA. SSC are responsive to androgenic compounds; however, this assay may have low sensitivity to detect anti-androgenic activity through effects on this parameter. Gonad histopathology can be evaluated to help assessing the reproductive fitness of the test animals and to add to the WoE of other parameters if needed. Additionally, quantitative fecundity is monitored daily, as well as behaviour and morphological abnormalities.

Even though the OECD TG 229 test is considered to be a level 3 test for endocrine MoA, it is considered both as a screen and as a test in the OECD Conceptual Framework, because of the fecundity parameter, which can show adverse effects. It has to be highlighted that the OECD TG 229 does not cover the juvenile life stage, so it will be insensitive to 'EATS-mediated' MoAs targeting specifically this sensitive window.

Validated species: All parameters have been validated on the fathead minnow (*Pimephales promelas*); a subset of parameters have been validated in the Japanese medaka (*Oryzias latipes*, i.e. VTG and secondary sex characteristics), and the zebrafish (*Danio rerio;* i.e. VTG).

21-day fish assay: a short-term screening for estrogenic and androgenic activity and aromatase inhibition (OECD TG 230, CF level 3)

The OECD TG 230, 21-day fish assay: a short-term screening for estrogenic and androgenic activity and aromatase inhibition (OECD, 2009b) has a similar test design and includes the same parameters as OECD TG 229, except for fecundity and gonad histopathology changes.

Validated species: Fathead minnow (*Pimephales promelas*); Japanese medaka (*Oryzias latipes),* partially validated for the zebrafish (*Danio rerio;* VTG).

Androgenised female stickleback screen (OECD GD 148, CF level 3)

A variant of OECD TG 230 is the androgenised female stickleback screen (OECD GD 148 (OECD, 2011a)). OECD declined to adopt this test as a test guideline, due to the modified nature of the test organism (androgenised females) *via* exposure to the potent androgen dihydrotestosterone. This is a 21-day *in vivo* assay for identifying endocrine active chemicals with (anti-)androgenic activity in fish using sexually mature female sticklebacks. Its usefulness is greater to detect androgen antagonists; however, its ability to detect anti-androgens is relevant only for chemicals that interact with the AR because females are specifically dosed with dihydrotestosterone to induce a moderate level of spiggin production and co-exposure to chemicals blocking the AR receptor will reduce spiggin production, indicating anti-androgenic effect. Compounds that display anti-androgenic activity *via* other mechanisms (i.e. disruption of steroidogenesis) will not be identified as such. In this test, spiggin is the only mechanistic parameter to be assessed. Additionally, survival, behaviour, morphological abnormalities should be monitored as well as body weight, in order to calculate the biomarker level (spiggin/g body weight).

Validated species: three-spined stickleback (*Gasterosteus aculeatus*).

EASZY assay detection of substances acting through estrogen receptors using transgenic cyp19a1bGFP zebrafish embryos (CF level 3)

This 96-h assay is currently under validation by the OECD. The test uses a transgenic zebrafish line expressing green fluorescent protein (GFP) under the control of the promoter of the ER-regulated *cyp19a1b* gene coding for brain aromatase. After 96 h of exposure, the embryos are scanned using a fluorescence imaging microscope, and the intensity of fluorescence recorded. This assay identifies whether estrogens may be produced from aromatisable androgens in certain parts of the brain sensitive to ER agonists; pro-estrogens that can be metabolised to become ER agonists; androgens that can be aromatised to ER agonists; and some non-aromatisable androgens.

Caution should be used with chemicals with a molecular weight $\geq$ 3 kDa and/or a very bulky molecular structure because absorption into the embryo via the chorion may have been impeded.

Moreover, although fish embryos have been shown to have metabolic capacities, it should be kept in mind that they might have a less efficient metabolism than juveniles and adults, i.e. that the use of this test with EDCs that require metabolic activation may give some false negatives (OECD, 2018b).

Species: cyp19a1bGFP zebrafish (*Danio rerio*).

Juvenile medaka anti-androgen screening assay (JMASA) (CF level 3)

This test is being drafted at the OECD as a Guidance Document. It is designed to identify androgen antagonists and chemicals interfering with androgen biosynthesis. No validation data have yet been produced, but some developmental data are available (OECD, 2018b).

The assay is based on male juvenile medaka (*Oryzias latipes*), which develops papillary processes as SSC under androgenic control. Anti-androgens or chemicals which interfere with androgen biosynthesis can prevent their appearance or limit their number. Juvenile medakas (both sexes) are exposed to the test chemical from 42 to 70 days post-fertilisation (28 days). Their genotypic sex is then determined and the males are evaluated for the presence, reduction or absence of papillary processes. It is optionally possible to measure VTG, so the assay can in principle also be used to detect estrogen agonists and antagonists and aromatase inhibitors, although those modalities are not currently under validation.

Species: Japanese medaka (*Oryzias latipes*).

Rapid androgen disruption adverse outcome reporter assay (RADAR) (CF level 3)

This 72- or 96-h assay is currently under validation by the OECD, for the detection of androgen receptor agonists and antagonists and chemicals interfering with androgen biosynthesis. No validation data have yet been produced but some published developmental data are available (Sebillot et al., 2014).

The test uses a transgenic medaka line expressing GFP under the control of the promoter of the AR-regulated three-spined stickleback spiggin1 gene coding for spiggin glue protein. After 72 or 96 h of exposure, the mesonephros of the embryos are imaged using a fluorescence imaging microscope, and the intensity of fluorescence is recorded to quantify androgen axis signalling.

Species: spg1-gfp medaka (*Oryzias latipes*).

OECD CF level 4 and 5 tests

There are three *in vivo* test guidelines for identification of endocrine-related adverse effects in fish at the level 4 and 5 of the OECD CF: the fish sexual development test or FSDT (OECD TG 234 (OECD, 2011b)) at level 4, the medaka extended one-generation reproduction test or MEOGRT (OECD TG 240 (OECD, 2015c)) and the fish life cycle toxicity test or FLCTT (US EPA OPPTS 850.1500 (US EPA, 2009d), which has not been validated by OECD) at level 5. Additionally, there is also the reproduction partial life cycle test at level 4, although no guideline is available for this test. Moreover, the fish early life stage test (OECD TG 210 (OECD, 2013b), which is proposed to be placed in Level 4 of the revised version of the OECD CF), although not being designed to give information on endocrine effects, should be considered as this test guideline is included in the standard information requirement for PPPs, might be required for BPs (see Appendix C), and gives information on both general toxicity (information which is necessary for a reliable interpretation of ED effect) and on parameters that might be sensitive to endocrine disruption such as hatchability and development (OECD TG 210).

The list of relevant parameters that give indications on the ED properties, based on OECD GD 150 and the JRC screening methodology, is shown in Table 15.

Fish sexual development test (OECD TG 234, CF level 4)

The OECD TG 234 fish sexual development test (FSDT, OECD 2011b) assesses early life stage effects and potential adverse consequences of endocrine-disrupting chemicals (e.g. estrogens, androgens and steroidogenesis inhibitors) on sexual development. It is an enhancement of the OECD TG 210 (OECD, 2013b), the fish early life stage toxicity test, with exposure from newly fertilised eggs until completion of sexual differentiation. The protocol is applicable to Japanese medaka, three-spined sticklebacks and zebrafish. The fathead minnow was also partially validated. Regarding endocrine activity, two main parameters are measured: VTG concentration and sex ratio. In Japanese medaka and three-spined sticklebacks, the sex ratio can be determined based on the genetic sex, which increases the power of the sex ratio statistics because it enables the detection of individual phenotypic sex reversal. Phenotypic sex is determined by gonadal histology examination, and it is a required parameter. Gonadal histopathology (evaluation and staging of oocytes and spermatogenetic cells) is an

optional measurement in this test guideline, which should be considered as it gives additional information for ED identification. SSC are also analysed in Japanese medaka. It has to be noted that the Japanese medaka (*Oryzias latipes*) is the species that can give the maximum information (fully validated species with both the genetic sex marker to identify individual sex reversal and analysable SSC). However, before choosing the species, the species sensitivity to sex ratio changes should be considered because some species are more susceptible than others to sex ratio changes caused by a specific endocrine mechanism. As an example, the validation data available so far showed that alterations of phenotypic sex ratio by the test substances were uncommon in sticklebacks (OECD 2011b). Therefore, the absence of observed changed in sex ratio in this species would not be sufficient to disregard a substance's endocrine potential in fish and in general, sticklebacks should not be used for conducting a new study. In contrast, the zebrafish sex ratio is very sensitive, more particularly to androgen agonists (OECD, 2018b).

An effect on sex ratio shows that the test chemical causes an adverse apical effect, is a developmental toxicant, and is probably also an ED, in the absence of general systemic toxicity at the same concentration (OECD GD 150). The combined measurement of VTG and sex ratio also gives, in the same test, information on both mechanism and adverse effect relevant at the population level, and can demonstrate the endocrine MoA. Additionally, gonadal histopathology is an optional 'EATS-mediated' parameter; body length and weight should be measured and survival, hatching success, abnormal behaviour and morphological abnormalities should be monitored.

Validated species: Japanese medaka (*Oryzias latipes*), zebrafish (*Danio rerio*), three-spined stickleback (*Gasterosteus aculeatus*), fathead minnow (*Pimephales promelas*) partially validated.

Medaka extended one-generation reproduction test (OECD TG 240, CF level 5)

The OECD TG 240 medaka extended one-generation reproduction test (MEOGRT (OECD, 2015c)) is a level 5 test method of the OECD CF, designed to evaluate the potential chronic effects of chemicals on fish, including potential endocrine effects. Fish are exposed over multiple generations, starting with the exposure of sexually mature males and females (F0), through development and reproduction in the F1 generation, until hatching in the F2 generation.

This test guideline includes various 'EATS-mediated' and '*in vivo* mechanistic' parameters such as hepatic VTG mRNA or VTG protein, phenotypic SSC characteristic (e.g. male anal fin papillae as related to genetic sex) and gonad histopathology which should be measured when this study is performed in the context of this guidance. In addition, this test guideline recommends measuring additional parameters like survival, behaviour, morphological abnormalities, gross development, hatching, time to spawn and reproduction, kidney and liver histopathology which are relevant for the ED assessment.

It is noted that this test is not expected to detect modest deviation of the sex ratio parameter because of the relatively small numbers of fish per replicate, i.e. low statistical power.

The Japanese medaka is the appropriate species for use in this test guideline, because of the possibility to determine its genetic sex.

A similar extended one-generation toxicity test on zebrafish is currently under development at the OECD, as an alternative species to the medaka. The endocrine-sensitive parameters would be the same, taking into account the biological differences between the species (e.g. the absence of validated SSC or sex probe for genetic sex determination in zebrafish). Ultimately, the choice of the species should depend on the sensitivity of each test species to a given parameter and species-specific characteristics.

Validated species: Japanese medaka (*Oryzias latipes*).

Fish life cycle toxicity test (OPPTS 850.1500, CF level 5)

The fish life cycle toxicity test (FLCTT) is placed at level 5 of the OECD CF. This method has not been adopted as an OECD guideline, and it is a draft US EPA method (OPPTS 850.1500 (US EPA, 2009d)). This method is used to investigate adverse apical effects on development, growth or reproduction over an entire lifecycle. The test should last from a given life stage in F0 to at least the same life stage in F1 (e.g. egg to egg) and the fish should be continuously exposed through reproductive maturity, followed by assessment of the early development of the F1 generation. It has been developed for use with fathead minnows and for the sheepshead minnow, although other species, such as medaka or zebrafish can be used, with minor changes to the protocol. Although the test is well recognised, it has not been validated by OECD. As the published test protocol contains limited details, any decision to perform the test should require further protocol specification (particularly if using other species, such as medaka or zebrafish). In the context of this guidance, as

this test does not include parameters specific to a particular EATS modality, it is recommended that those parameters should be added and that the test design of the FLCTT is adapted to include all the parameters covered by the MEOGRT. Limited data are obtained from the F1 generation in the test. The parameters of particular interest in the context of estrogens, androgens and steroidogenesis disruptors identification are time to sexual maturity, sex ratio of adults, fecundity and fertility, but other parameters may also be responsive to other endocrine modes of action (e.g. growth may respond to some thyroid disruptors).

Species: fathead minnow (*Pimephales promelas*), sheepshead minnow *(Cyprinodon variegatus),* but any other species could be used if the protocol is modified accordingly.

Fish reproduction partial lifecycle test (no guideline available, CF level 4)

A fish reproduction partial lifecycle test that would cover exposure of sexually mature adults in the F0 generation, through spawning, followed by a short-term exposure of F1 embryos and juveniles might give useful information on 'EATS-mediated' effects. Currently, there is no validated guideline for such a test. If such data are already available they can be taken into account. However, if a new study has to be carried out, a validated guideline should be used.

Validated species: none.

Fish early life stage toxicity test (OECD TG 210, CF level 4)

This test is designed to define the chronic lethal and sub-lethal effects of chemicals on fish early life stage. The duration of the test varies between 28 and 68 days post-hatch, depending on the species, and covers the life stages from immediately after fertilisation, larvae and juvenile fish.

Although there are no 'EATS-mediated' parameters measured in this test, it gives information on general toxicity that can help with the interpretation of data for ED identification and include parameters that might be 'sensitive to, but not diagnostic of, EATS' such as hatchability and development. Moreover, there is limited evidence to suggest that some thyroid system disruptors are able to interfere with the metamorphosis of the fish embryo to the larvae (Nelson et al., 2016; Stinckens et al., 2016). It has to be noted that this test does not cover the reproductive life stage of the fish; therefore, chemicals that are suspected to affect reproduction should be examined in a test that covers it. This test guideline was not reported in Table 15 since it includes only 'sensitive to, but not diagnostic of, EATS' parameters.

Validated species: rainbow trout (*onchorhynchus mykiss*), fathead minnow, (*Pimephales promelas*), zebrafish (*Danio rerio*), medaka (*Oryzias latipes*), and also sheepshead minnow (*Cyprinodon variegatus)* and silverside (*Menidia* spp.).

Table 15: Fish: main investigated parameters – parameters '*in vivo* mechanistic' (highlighted in orange); 'EATS-mediated' (highlighted in blue) and parameters 'sensitive to, but not diagnostic of, EATS' (highlighted in purple)

		Section A				Section B				
Test guideline		**OECD TG 229[c]**	**OECD TG 230**	**OECD TG 234**	**OECD TG 240[d]**	**OPPTS 850.1500[e]**	**OECD GD 148**	**EASZY[f]**	**RADAR[f]**	**JMASA[f]**
Test duration		21 days	21 days	60 days post-hatch	133 days	100-190 days	21 days	96 h	72 or 96 h	28 days
Life stages		Sexually mature male and spawning female (F0)	Sexually mature male and spawning female (F0)	From newly fertilised egg until completion of sexual differentiation (F0)	From sexually mature males and females of F0 to hatching of the F2	Freshly fertilised eggs of F0 to juvenile stage of F1	Sexually mature female (F0)	Embryonic	Embryonic	Juveniles
Species		Fathead minnow, Japanese medaka, zebrafish	Fathead minnow, Japanese medaka, zebrafish	Japanese medaka, three-spined stickleback, zebrafish, fathead minnow (partially validated)	Medaka; can be adapted to zebrafish (ZEOGRT, under validation)	Fathead minnow or sheepshead minnow (marine). Can be adapted to medaka and zebrafish	Stickleback	Zebrafish	Medaka	Medaka
Parameter name	**Indicative of:[a]**									
VTG in females	E, A, S	X	X	X	X					X
VTG in males	E, A, S	X	X	X	X					X
Spiggin	A					X		X		
Male SSC in females	A	X	X	X[g]						
Male SSC in males	E, A, S	X	X	X[g]					X	
Specific gonad histopathology[b]	E, A, S	X (optional)		X (optional)						
Sex ratio (female biased)	E, A			X	X					X
Sex ratio (male biased)	E, A, S			X	X					X

Test guideline		Section A				Section B				
		OECD TG 229[c]	OECD TG 230	OECD TG 234	OECD TG 240[d]	OPPTS 850.1500[e]	OECD GD 148	EASZY[f]	RADAR[f]	JMASA[f]
Transcriptional activity of cyp19a1b	E						X			
Behaviour	N	X	X	X	X	X				X
Length	N			X	X					X
Morphological abnormalities	N	X	X	X		X				
Gonadosomatic index	N									
Embryo time to hatch	N									
Reproduction (fecundity, fertility)	N	X			X					X
Survival	N	X	X	X	X	X	X	X	X	X
Larval survival and length	N			X						
Survival of embryos	N			X						
Time to maturity (time to first spawn)	N				X					X
Hatching success	N			X	X					X
Histopathology (liver, kidney)	N									
Body weight	N			X	X	X				X

The table is divided into two sections: Section A lists parameters from tests for which guidance is provided in OECD GD 150; Section B lists parameters from tests that have not yet completed validation, or not primarily designed for detection of endocrine disruption, for which limited guidance is given in OECD GD 150.

(a): Based on OECD GD 150 (OECD, 2018b), indicative of: the (E)strogen- (A)ndrogen-, (S)teroidogenesis- or (T)hyroid modalities; (N)ot assignable to a specific modality.

(b): Histological examination of the gonads should enable identification of intersex (presence of testis-ova) and undifferentiated fish. It should be noted that some specific gonad histopathological findings are EATS-mediated but some other are not (i.e. oocyte atresia). More detailed guidance on specific gonad histopathology examination in fish is given in (OECD, 2010).

(c): The USEPA FSTRA guideline (OPPTS 890.1350) is considered equivalent if all the endpoint of the OECD TG 229 have been investigated. Additionally, the gonadosomatic index should be reported and plasma sex steroid concentration might be reported (optional).
(d): The USEPA MEOGRT guideline (OCSPP 890.2200) is considered equivalent if all the endpoint of the OECD TG 240 have been investigated. A similar guideline to TG 240 is currently under validation by OECD on zebrafish (ZEOGRT) and could be used instead of the MEOGRT, once validated. The choice between those two test guidelines should be made based on the species sensitivity and the chemicals being test.
(e): As this test does not include parameters specific to a particular EATS modality, those parameters should be added and the test design of the FLCTT should be adapted to include all the parameters covered by the MEOGRT in order to be considered equivalent.
(f): This guideline is currently under validation, and has been included for the sake of completeness. The assignment of parameters to the different groups should be applied in accordance to the final guideline.
(g): When medaka is the test species.

4.3.2.3. Amphibians

Two standardised tests, the amphibian metamorphosis assay (AMA (OECD, 2009c)) and the larval growth and development assay (LAGDA (OECD, 2015d)) can be used to investigate potential endocrine adverse effects in amphibians. The AMA (OECD TG 231, level 3 of the OECD CF) is a validated amphibian mechanistic *in vivo* assay designed as a screening assay for potential thyroidal effects. The LAGDA (OECD TG 241, level 4 of the OECD CF) is more comprehensive, covering, in addition to thyroidal effects, other endocrine-disrupting effects on the development of the reproductive system, and allowing the evaluation of other types of developmental and reproductive toxicants. Test conditions and measured parameters are briefly described below and summarised in Table 16. Moreover, those tests also include the investigation of parameters that are not mechanistically specific for thyroid effects and might be sensitive to general toxicity. It has to be noted that water quality could impact the results, as common water pollutants like nitrates may also have thyroid effects in amphibians (Wang et al., 2015). Another level 3 test, the *Xenopus* embryonic thyroid signalling assay (XETA) is currently under validation for the detection of thyroid active substances.

OECD CF level 3 tests

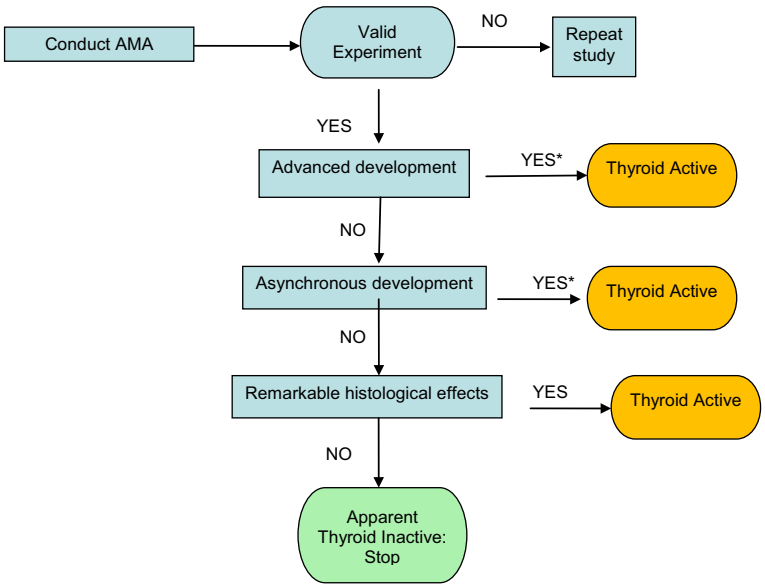
Amphibian metamorphosis assay (OECD TG 231; OPPTS 891100, CF level 3)

The AMA was developed to identify substances affecting the function of the HPT axis in vertebrates. The test is conducted with larval stages (tadpoles) of *Xenopus laevis* exposed for 21 days. The developmental stage, hindlimb length, snout to vent length measurement and wet weight are the apical endpoints of the AMA.

The apical endpoint hindlimb length as well as thyroid histological changes are mediated by endocrine effects on the thyroid axis. Snout–vent length and wet weight are measured to assess growth and are useful in detecting generalised toxicity of the test compound, although they can also be affected by thyroid disturbance. Abnormal behaviour (floating on the surface, lying on the bottom of the tank, irregular swimming, etc.) and gross malformations (morphological abnormalities, haemorrhagic lesions, bacterial or fungal infection) should be recorded.

Accelerated development is assessed via hindlimb length measurement normalised by snout–vent length and occurs through effects which are thyroid hormone related. These can be either from direct interaction with thyroid hormone receptors or effects which alter circulating thyroid hormone levels. Accelerated and asynchronous development (characterised by disruption of the relative timing of the morphogenesis or development of different tissues and the inability to clearly establish the developmental stage of an animal by morphological landmarks) are thyroid-mediated effects. Delayed development is not by itself an indicator of anti-thyroidal activity and needs to be confirmed by histopathological analysis of the thyroid. A decision tree for the detection of thyroidal effects in the AMA is presented in Figure 7.

Validated species: African clawed frog (*Xenopus laevis*).



*Histology may be required by some regulatory authorities despite significant differences in advanced and asynchronous development. The entity performing this test is encouraged to consult the competent authorities prior to performing the test to determine which parameters are required.

Figure 7: Decision tree for evaluating thyroidal effects in the AMA (from OECD TG 231 (OECD, 2009c))

Xenopus embryonic thyroid signalling assay XETA (CF level 3)

This 72-h *in vivo* transcriptional assay is currently under validation by the OECD. This assay requires the use of a transgenic *Xenopus laevis* at embryonic stages. This transgenic line can detect the activity of thyroid agonists that activate thyroid hormone receptors, as well as antagonists of the thyroid axis that work through various mechanisms. The principle of the assay is the measurement of GFP fluorescence in the tadpoles, each translucent tadpole expressing a basal fluorescence. In contact with a thyroid disruptor, the GFP is down- or upregulated, which allows the chemical effect on the thyroid system to be assessed.

Species: African clawed frog (*Xenopus laevis*).

OECD CF level 4 and 5 tests

Larval amphibian growth and development assay (OECD TG 241; OCSPP 890.2300 CF level 4)

The LAGDA was designed to detect apical adverse effects resulting from endocrine and non-endocrine mechanisms covering all early life stages of amphibians from embryo to larva to early juvenile, and is placed at level 4 of the OECD CF.

It is possible to diagnose thyroidal effects following the same evaluation of test parameters and decision tree as in AMA (see Section OECD CF level 3 tests for details). In addition, the LAGDA allows the detection of endocrine effects on the development of the reproductive system and give emphasis to population relevance parameters. The HPG axis is particularly active during gonadal differentiation (which occurs during larval development), maturation of gonads and development of SSC (juvenile phase) and during functional reproduction of adults. The LAGDA covers the first two of these sensitive phases, but not the third phase. In order to cover the full reproductive cycle, it would be necessary to conduct a full life cycle test, which is currently not possible within a laboratory test, owing to the limitations of the model species.

Exposure of tadpoles to estrogens or androgens acting through the E, A and S pathway can lead to partial or full sex reversal and in some cases resulting in fully sexually functional adults (OECD, 2015a). Phenotypic sex ratio is an apical endpoint mediated by endocrine activity on the HPG axis. The histopathology of gonads and reproductive ducts give information on potential endocrine mechanisms, whereas change in levels of VTG provide information about a substance interfering with the sex hormone system (E, A, S).

The apical endpoints time to metamorphosis, as well as thyroid histological changes, are mediated by endocrine effects on the thyroid axis.

In addition, mortality, abnormal behaviour, growth (length and weight), histopathology examination of the liver (i.e. decreased glycogen vacuolation) and kidneys (i.e. mineralisation and tubule dilation)

as well as liver somatic index are useful in the context of interpreting the relevance of potentially ED-related effects as a secondary non-specific consequence of generalised systemic toxicity.

The potential relationship between the histological changes observed and the treatment on the one hand, and a potential endocrine-disrupting effect on the other hand should be considered on a case-by-case basis based on a WoE approach (OECD, 2015a) (OECD, 2015b).

Validated species: African clawed frog (*Xenopus laevis*).

Table 16: Amphibians: main investigated parameters for which guidance on the interpretation is provided in the OECD GD 150. Parameters '*in vivo* mechanistic' (highlighted in orange); 'EATS-mediated' (highlighted in blue) and parameters 'sensitive to, but not diagnostic of, EATS' (highlighted in purple)

		Section A		Section B
Test guideline		**OECD TG 231**	**OECD TG 241**	**XETA[f]**
Test duration		21 days	16 weeks	72 h
Life stages		Tadpole NF (NF 51)	Embryo, tadpoles, early juvenile	Tadpole (NF 45)
Species		*Xenopus laevis*	*Xenopus laevis*	*Xenopus laevis*
Parameter name	**Indicative of:** [a]			
Plasma level of VTG	E, A, S		X (optional)	
Developmental stage[b]	T	X		
Hindlimb length[c]	T	X		
Thyroid histopathology (amphibian)[d]	T	X	X	
Histopathology[d] (gonad[e], reproductive ducts)	E, A		X	
Sex ratio (phenotypic (gonad histology), genetic)	E, A		X	
Time to metamorphosis (NF stage 62)	T		X	
Transcriptional activity of THbZIP	T			X
Body weight	N	X	X	
Snout-vent length/ Growth	N	X	X	
Malformations	N	X	X	
Mortality	N	X	X	X
Behaviour	N	X	X	
Histopathology (liver, kidney)[d]	N		X	
Liver weight, liver somatic index	N		X	

(a): Based on OECD GD 150, indicative of: the (E)strogen-, (A)ndrogen-, (S)teroidogenesis- or (T)hyroid modalities; (N)ot assignable to a specific modality.

(b): The developmental stage is used to determine if the development is accelerated, asynchronous, delayed or unaffected. An accelerated development is considered as indicative of thyroid-related activity, whereas a delay in the development might be triggered by other non-endocrine pathways.

(c): Hindlimb length development is used qualitatively for the determination of developmental stage, and is also considered as a quantitative parameter to detect effect on the thyroid axis (increased hindlimb length).

(d): Histopathology changes criteria are detailed in OECD 2015a,b. As an example, decreased vacuolation (liver), gonadal stage, tubule development and germ cell degeneration (gonad); and mineralisation and tubule dilation (kidney) can be assessed.

(e): Some histopathologic findings in the gonad are EATS-mediated (i.e. intersex) but some other can be the result of other non-endocrine MoAs (i.e. oxidative stress can result in increased apoptosis).

(f): This guideline is currently under validation, and has been included for the sake of completeness. The assignment of parameters to the different groups should be applied in accordance to the final guideline.

4.3.2.4. Birds

For birds, only a limited number of standardised *in vivo* methods are available, and little information can be gained from those guidelines concerning potential ED-related effects. In general, little is known of the impact of endocrine disruptors in birds compared to other species, and more research is needed to develop responsive parameters and *in vitro* and *in vivo* protocols to specifically address the differences between birds and other vertebrate taxa. The avian reproduction test (OECD TG 206 (OECD), level 4 of the OECD CF) gives only apical endpoints while the avian two-generation toxicity test in the Japanese quail (OCSPP 890.2100, Level 5 of the OECD CF) (US EPA, 2009a) covers four different life stages of the quail and investigates some biochemical parameters. While the latter might have the capability to be responsive to most chemicals with EATS activities, during its validation the test design was considered unresponsive to EATS modalities with the tested chemicals and the undertaken validation process initiated by OECD was not completed. Therefore, the test has not been validated. A detailed OECD review paper on the avian two-generation study has nevertheless been published during the first phase of the validation process (OECD, 2007a). Table 17 sets out the parameters investigated according to the OECD TG 206 and OCSPP 890.2100, together with their relevance for identifying a substance with a potential for endocrine disruption according to the EATS modalities.

Avian reproduction test (OECD TG 206, CF level 4)

The avian reproduction toxicity test (OECD TG 206 (OECD, 1984)) gives a list of parameters that might be endocrine-sensitive but which cannot be considered specific for the identification of an endocrine MoA (i.e. 'sensitive to, but not diagnostic of, EATS'). For example, the effects of dichlorodiphenyldichloroethylene, DDT's metabolite, on eggshell thickness in birds, were considered in the past as being induced by increased liver metabolism of steroid hormones. However, the mechanisms underlying eggshell thickness are still not fully clarified, since different species show differing effects on eggshells. Therefore, the link to endocrine disruption is not completely clear (Lundholm, 1997; Berg et al., 2004; De Wit, 2006). It is noted that OECD TG 206 recommends gross pathology examinations, although further guidance on this assessment are not given in this test guideline. Nevertheless, the OECD provides recommendations on how this assessment should be performed (OECD, 2002). It is recommended that gross pathology findings are reported when available with particular reference to potential endocrine target organs (thyroid and gonads/ reproductive organs).

Validated species: mallard duck (*Anas platyrhynchos*), bobwhite quail (*Colinus virginiatus*) and Japanese quail (*Coturnix coturnix japonica*).

US EPA avian two-generation study (OCSPP 890.2100, CF Level 5)

The avian two-generation study developed at the US EPA was designed to investigate the impact of a chemical upon Japanese quail and includes chemical exposure at four life stages: *in ovo*, juvenile, sub-adults and adults (US EPA, 2009a). The test is specifically designed to investigate the health and reproductive fitness of the first filial (F1) generation following parental (F0) dietary exposure to the tested chemical. Survival of the F2 generation at 14 days post hatch is the primary parameter measured in this test. The test can also be extended until reproductive maturity of the second filial (F2) generation. To be valuable in assessing the potential for endocrine disruption the test should include measurement of thyroid and steroid hormones, histology and morphological parameters.

However, before conducting this test it has to be noted that it was considered insufficiently validated according to OECD standards, and that its use has considerable animal welfare implications. Therefore, as such this test should not currently be requested to address ED issues.

Species: Japanese quail (*Coturnix japonica*).

Table 17: Birds: main investigated parameters – parameters '*in vivo* mechanistic' (highlighted in orange); 'EATS-mediated' (highlighted in blue) and parameters 'sensitive to, but not diagnostic of, EATS' (highlighted in purple)

		Section A	Section B
Test guideline		**OECD TG 206 (level 4)**	**US EPA OCSPP 890.2100[c] (level 5)**
Test duration		At least 20 weeks	At least 33 weeks
Life stages		Adults (F0), *in ovo* (F1), chicks (F1 up to 14 days)	Adults (F0, F1), *in ovo* (F1, F2), juvenile (F1, F2), subadults (F1)
Species		Mallard duck, bobwhite quail, Japanese quail	Japanese quail
Parameter name	**Indicative of:** [a]		
Oestradiol, testosterone and thyroid hormone levels measurements (egg yolk, adult, thyroid hormone from thyroid gland)	E,A,T,S		X
Histopathology (thyroid gland, gonad)[b]	E,A,T		X
Phenotypic and genotypic sex ratio	E,A		X
Gross pathology	N	X	X
Hatchability	N	X	X
Egg fertility (embryonic day 8)	N		X
Eggshell thickness	N	X	X
Eggshell strength (Newton)	N		X
Egg viability (% viable embryo of egg set)	N	X	
Embryo viability (embryonic day 15)	N		X
Egg production	N	X	X
Cracked eggs	N	X	X
Body weight	N	X	X
Survival	N	X	X
Viable embryos	N	X	X
Number of 14-day old survivors	N	X	X
Time to female reproductive maturation (first egg production)	N		X
Time to male reproductive maturation (first foam production)	N		X
Histopathology (liver, kidney)[b]	N		X

The table is divided into two sections: Section A lists parameters from tests for which guidance is provided in OECD GD 150; Section B lists parameters from tests that have not yet completed validation, or not primarily designed for detection of endocrine disruption, for which limited guidance is given in OECD GD 150.

(a): Based on the OECD GD 150 (OECD, 2018b), indicative of. The (E)strogen, (A)ndrogen; (S)teroidogenesis; or (T)hyroid modalities; (N)ot assignable to a specific modality.

(b): Histopathology criteria are detailed in OCSPP 890.2100 (US EPA 2009a). If no signs of overt general toxicity are observed among F1 birds in the high treatment group, histopathological samples from F0, F1 and F2 birds will be limited to reproductive tissues and thyroid glands. If signs of overt toxicity are observed in the high-treatment group, the potential of overt toxicity mimicking or masking endocrine-related effects cannot be ruled out. Liver, kidney, adrenal, thyroid, reproductive tissues should be examined in the next highest dose until indications of overt toxicity are not observed.

(c): This test guideline is not validated by OECD.

4.4. Epidemiological data, field studies and population models

4.4.1. Epidemiological data

According to Regulation (EU) No 283/2013[8] setting out data requirements for active substances, the dossiers should include scientific peer-reviewed literature, notably 'relevant epidemiological (EPI) studies shall be submitted, where available'. Likewise, in the BP Regulation[1] concerning the making available on the market and use of BPs, the consideration of epidemiological data is part of Annex II (Information requirements for active substances; 8.12.4 Epidemiological studies on the general population) and Annex IV (General rules for the adaptation of the data requirements). The latter Annex states that the use of '*existing historical human data, such as epidemiological studies on exposed populations, accidental or occupational exposure data, biomonitoring studies, clinical studies and human volunteer studies performed in accordance with internationally accepted ethical standards shall be considered*'. However, it is clear that there is no obligation for the applicants to conduct epidemiological studies specifically for the active substance undergoing the approval or renewal process. Rather, according to the PPP Regulation[2], applicants submitting dossiers for approval of active substances should provide '*scientific peer-reviewed public available literature [...]. This should be on the active substance and its relevant metabolites dealing with side-effects on health [...] and published within the last 10 years before the date of submission of the dossier*'; in particular, epidemiological studies should be retrieved from the literature. As a literature search including epidemiological studies is mandatory and guidance is in place (EFSA, 2011); a consistent approach for inclusion of epidemiological studies in the dossier is expected.

4.4.2. Field studies and monitoring data

Field studies are described as experimental activities performed outside the laboratory environment, for instance on land plots or in outdoor micro/mesocosms, often in combination or in sequence with activities carried out in a laboratory (OECD, 1999). Mesocosms are complex systems, but are still experimental systems and more amenable to control of non-treatment factors when compared to field studies on land plots. It has to be noted, however, that fish and other vertebrates such as amphibians are usually not introduced into mesocosms because of their influence on other populations (e.g. invertebrates) (EFSA PPR Panel, 2013). Field studies are performed under more realistic environmental conditions when compared to the worst-case laboratory conditions, because the organisms interact with the abiotic and biotic factors and are also exposed to additional stressors and indirect effects occurring in their natural environment. Therefore, field studies might make it possible to better identify the impact of an adverse effect on a specific population. However, as already highlighted by the EFSA Scientific Committee (2013), one of the main issues of field experiments is the complexity of evaluating the results, the interpretation of which being affected by confounding factors (e.g. uncontrolled factors such as the weather conditions). Their interpretation requires therefore adequate and robust statistical analyses, and informed expert judgement. Extrapolation of observed study results under specific environmental conditions to different situations is uncertain. Field studies typically cover only a limited period of time and long-term population trends are usually not observed. Furthermore, with the exception of mesocosm studies, the field studies give a picture of a particular situation of use, but it is not possible to establish a dose–response relationship. Additionally, the design of this kind of study, in the case of vertebrates, is particularly complex. Due to the home range of these organisms, the choice of species that could be tested is limited, i.e. only species with manageable home range can be tested. This limitation also applies to the feeding guild; species representative of a certain feeding guild or feeding class may be difficult to test in the field, such as large predators (EEA, 2012). Furthermore, these issues could prevent the investigation of the potential impact on the most vulnerable species.

It is additionally noted that to ensure robustness of the results, field tests require a high number of animals/replicates to be tested and both the BP and PPP Regulations aim for a minimisation of animal (vertebrate) testing. Targeted experimental field studies may be useful to investigate adversity on vulnerable populations in relation to specific MoAs. Examples of the use of these studies in the assessment of endocrine-mediated effects at population level are reported in the scientific open literature, see e.g. (Caslin and Wolff, 1999; Palace et al., 2009). However, it must be noted that, in general, standard and validated methodologies to perform such studies are still missing.

Information on the potential effects at field level could also be deduced from monitoring studies. Field monitoring studies normally combine chemical monitoring in the environment (and in the food chain) with observation of effects on wildlife. Various examples of studies investigating endocrine-mediated effects in wildlife via monitoring are reported in the scientific open literature, e.g. in (EEA, 2012). Nevertheless, care must be taken in the interpretation of monitoring data when these studies are not designed to find the link between the exposure, the effects and the MoA of a specific chemical. In addition, the uncertainty around the exposure levels may hamper the interpretation of the results.

4.4.3. Population models

In addition to field data, computational methods (e.g. population modelling) could provide valid support in translating the effects observed in the laboratory to wild population level (Kohler and Triebskorn, 2013). A large number of population models are available for almost any taxonomic group. Typologies can be identified among those different models: (i) scalar or unstructured models which assess potential changes in the population over time (birth, death, immigration, emigration rates per unit of population such as the individual or biomass); (ii) structured demographic population models which incorporate the biological structure of the population by assessing demographic rates of a progression of cohorts usually classed by age or life stage (life history models); (iii) individual-based models which model the survival, productivity and movement of each individual in the population during its entire life span, in some cases also considering the physiological states of each individual; and (iv) dynamic energy budget models assessing the changes in bioenergetics at individual level (Kramer et al., 2011). The different models could then provide different answers and should be selected on the basis of the specific questions to be answered in the assessment.

5. Recommendations

5.1. Recommendations for applicants and assessors

***In vitro* assay interference**

It is recommended that assay interference is controlled by performing the *in vitro* method using suitable positive, negative, blank or vehicle controls. If the endpoints are of an analytical nature, the controls can also be spiked with the test item to verify that the test item does not in any way hinder the normal function of the test system or interfere with the readout.

Examples of readout-specific interference include:

- absorption, fluorescence or quenching of fluorescence at the evaluation wavelength;
- non-specific activation, prolonging or inhibition of the luciferase signal;
- alteration of enzyme function, or co-factor, or of other limiting reagents by test item;
- strongly reducing agents, reducing colour formation non-enzymatically.

***In vitro* cytotoxicity**

Non-cytotoxic concentrations should be considered for the assessment of the data. Different cells might behave differently, e.g. fungicides are more toxic to yeast cells than to mammalian cells. While cytotoxicity can be observed under the microscope, increasing use of high content, high throughput techniques makes the visual observation of cells more difficult. A measure of cytotoxicity can be obtained by specific methods assessing cell viability, e.g. by looking at cellular adenosine triphosphate content, lactate dehydrogenase release or at cellular (mitochondrial) metabolism.

Detailed histopathological evaluation of testis

Histopathological evaluation of testis in mammals is routinely performed in regulatory general toxicity studies. Detailed histopathological evaluation is considered a sensitive indicator of chemically induced effects. In the context of this guidance, 'detailed histopathological examination' (e.g. OECD TG 421/422) should be intended as a qualitative examination with an awareness of the spermatogenic cycle (staging). The reader should refer to the publication of Creasy for additional methodological and interpretative information (Creasy, 2003).

***In vivo* bioassays with fish and amphibians**

The current standard *in vitro* tests are only performed with mammalian cells. Some *in vivo* bioassays (e.g. RADAR, XETA, EASZY and JMASA) with fish and amphibians are currently in the

validation process (see OECD CF level 3 tests in sections 4.3.2.2 and 4.3.2.3). It is recommended that those under validation are performed together with the other (already validated) level 3 mechanistic assays reported in Table 15, once fully validated and when triggered based on the assessment strategy (see Section 3.1). In some cases, this will reduce the uncertainty linked to the extrapolation of mechanistic information from mammalian to other vertebrate species and from cells to whole organisms.

Fish chronic toxicity study

The OECD TG 234, 240 and fish life cycle toxicity test (OPPTS 850.1500) require, as optional, the assessment of gonad histopathology (e.g. staging of gonads, severity of intersex). It is recommended that this investigation is systematically performed each time that the study is carried out, see also OECD GD 123 (OECD, 2010).

Bird long-term toxicity studies

In the case of birds, it is noted that the avian reproduction test (OECD TG 206 (OECD, 1984)) recommends gross pathology examinations. However, further details on this assessment are not reported. Nevertheless, OECD provides recommendations on how this assessment should be performed (OECD, 2002). For the purpose of this guidance, it is recommended that gross pathology examinations' findings are reported when available with particular reference to ED's potential target organs (thyroid and gonads/reproductive organs).

Adverse outcome pathway for endocrine-related adverse outcomes

In the AOP Wiki,[14] a number of AOPs exist for endocrine- and non-endocrine- related adverse outcomes. They should be used in order to substantiate the biological plausibility in cases where the same pathway is investigated.

5.2. Recommendations for future research

It is recommended that more ED-related AOP should be developed by the scientific community; this will facilitate the applicability of the overall assessment and the interpretation of the outcome.

It is recommended that the possibility of including mechanistic parameters such as hormonal level measurements and histopathology in the OECD TG 206 is explored. Moreover, further guidance on the interpretation of data on histopathology on birds would be needed.

Considering the current knowledge in fish endocrinology and the availability of standard test methodologies, further investigations are recommended including the possibility of measuring additional parameters related to modalities other than EAS (e.g. thyroid hormones and histopathology) in the existing test guidelines.

Further exploration of the possibility of including measurements of thyroidal hormones in the OECD 241 is recommended.

Future research is recommended in order to better understand the endocrinology of reptiles and evaluate whether extrapolation from other vertebrates can be scientifically underpinned.

Further research is recommended for a better understanding of the endocrinology of invertebrates in the light of developing test guidelines for the identification of ED, including also mechanistic parameters.

Future research is needed for a better understanding of non-EATS modalities in light of developing a test strategy covering them.

Further research is needed to extrapolate the relevance at population level of adverse effects observed in laboratory studies.

References

Adori M, Kiss E, Barad Z, Barabas K, Kiszely E, Schneider A, Kovesdi D, Sziksz E, Abraham IM, Matko J and Sarmay G, 2010. Estrogen augments the T cell-dependent but not the T-independent immune response. Cellular and Molecular Life Sciences 67, 1661–1674. https://doi.org/10.1007/s00018-010-0270-5

Andersson N, Islander U, Egecioglu E, Löf E, Swanson C, Moverare-Skrtic S, Sjögren K, Lindberg MK, Carlsten H and Ohlsson C, 2005. Investigation of central versus peripheral effects of estradiol in ovariectomized mice. Journal of Endocrinology 187, 303–309. https://doi.org/10.1677/joe.1.06181

[14] https://aopwiki.org/

Andersson C, Katsiadaki I, Lundstedt-Enkel K and Orberg J, 2007. Effects of 17alpha-ethynylestradiol on EROD activity, spiggin and vitellogenin in three-spined stickleback (*Gasterosteus aculeatus*). Aquatic Toxicology 83, 33–42. https://doi.org/10.1016/j.aquatox.2007.03.008

Ankley GT and Jensen KM, 2014. A novel framework for interpretation of data from the fish short term reproduction assay (FSTRA) for the detection of endocrine-disrupting chemicals. Environmental Toxicology and Chemistry, 33, 2529–2540. https://doi.org/10.1002/etc.2708

Ankley GT, LaLone CA, Gray LE, Villeneuve DL and Hornung MW, 2016. Evaluation of the scientific underpinnings for identifying estrogenic chemicals in nonmammalian taxa using mammalian test systems. Environmental Toxicology and Chemistry 35, 2806–2816. https://doi.org/10.1002/etc.3456

Arredouani MS, 2014. New insights into androgenic immune regulation. Oncoimmunology 3. https://doi.org/10.4161/21624011.2014.954968

Beausoleil C, Ormsby JN, Gies A, Hass U, Heindel JJ, Holmer ML, Nielsen PJ, Munn S and Schoenfelder G, 2013. Low dose effects and non-monotonic dose responses for endocrine active chemicals: science to practice workshop: workshop summary. Chemosphere 93, 847–856. https://doi.org/10.1016/j.chemosphere.2013.06.043

Becker RA, Ankley GT, Edwards SW, Kennedy SW, Linkov I, Meek B, Sachana M, Segner H, Van Der Burg B, Villeneuve DL, Watanabe H and Barton-Maclaren TS, 2015. Increasing Scientific Confidence in Adverse Outcome Pathways: application of Tailored Bradford-Hill Considerations for Evaluating Weight of Evidence. Regulatory Toxicology and Pharmacology 72, 514–537. https://doi.org/10.1016/j.yrtph.2015.04.004

Benigni R, Battistelli CL, Bossa C, Giuliani A and Tcheremenskaia O, 2017. Endocrine Disruptors: data-based survey of in vivo tests, predictive models and the Adverse Outcome Pathway. Regulatory Toxicology and Pharmacology, 86, 18–24. https://doi.org/10.1016/j.yrtph.2017.02.013

Berg C, Blomqvist A, Holm L, Brandt I, Brunstrom B and Ridderstrale Y, 2004. Embryonic exposure to oestrogen causes eggshell thinning and altered shell gland carbonic anhydrase expression in the domestic hen. Reproduction 128, 455–461. https://doi.org/10.1530/rep.1.00211

Boobis AR, Cohen SM, Dellarco V, McGregor D, Meek ME, Vickers C, Willcocks D and Farland W, 2006. IPCS framework for analyzing the relevance of a cancer mode of action for humans. Critical Reviews in Toxicology 36, 781–792. https://doi.org/10.1080/10408440600977677

Boobis AR, Doe JE, Heinrich-Hirsch B, Meek ME, Munn S, Ruchirawat M, Schlatter J, Seed J and Vickers C, 2008. IPCS framework for analyzing the relevance of a noncancer mode of action for humans. Critical Reviews in Toxicology 38, 87–96. https://doi.org/10.1080/10408440701749421

Browne P, Judson RS, Casey WM, Kleinstreuer NC and Thomas RS, 2015. Screening Chemicals for Estrogen Receptor Bioactivity Using a Computational Model. Environmental Science and Technology 49, 8804–8814. https://doi.org/10.1021/acs.est.5b02641

Caslin TM and Wolff JO, 1999. Individual and demographic responses of the gray-tailed vole to vinclozolin. Environmental Toxicology and Chemistry 18, 1529–1533. https://doi.org/10.1897/1551-5028(1999)018

Cramer RD, Patterson DE and Bunce JD, 1988. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. Journal of the American Chemical Society 110, 5959–5967. https://doi.org/10.1021/ja00226a005

Creasy DM, 2003. Evaluation of testicular toxicology: a synopsis and discussion of the recommendations proposed by the Society of Toxicologic Pathology. Birth Defects Res B Dev Reprod Toxicol 68, 408–415. https://doi.org/10.1002/bdrb.10041

Cronin MTD and Worth AP, 2008. (Q)SARs for predicting effects relating to reproductive toxicity. Qsar & Combinatorial Science 27, 91–100. https://doi.org/10.1002/qsar.200710118

Cutolo M, Seriolo B, Villaggio B, Pizzorni C, Craviotto C and Sulli A, 2002. Androgens and estrogens modulate the immune and inflammatory responses in rheumatoid arthritis. Neuroendocrine Immune Basis of the Rheumatic Diseases Ii, Proceedings 966, 131–142. https://www.ncbi.nlm.nih.gov/pubmed/12114267

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

De Wit C, 2006. Effects of endocrine disrupters in wild birds and mammals. In: Grotmol T, Bernhoft A, Eriksen G and Flaten T (eds). *Endocrine Disrupters*. The Norwegian Academy of Science and Letters, Oslo.

Dimitrov SD, Diderich R, Sobanski T, Pavlov TS, Chankov GV, Chapkanov AS, Karakolev YH, Temelkov SG, Vasilev RA, Gerova KD, Kuseva CD, Todorova ND, Mehmed AM, Rasenberg M and Mekenyan OG, 2016. QSAR Toolbox – workflow and major functionalities. SAR and QSAR in Environmental Research, 2016, 1–17. https://doi.org/10.1080/1062936X.2015.1136680

ECHA (European Chemicals Agency), 2008. Guidance for the implementation of REACH, Chapter R.6: QSARs and grouping of chemicals. ECHA, Helsinki. 134 pp. Available online: https://echa.europa.eu/guidance-documents/guidance-on-reach

ECHA (European Chemicals Agency) 2011. Guidance for the implementation of REACH, Chapter R.4: Evaluation of available information. ECHA, Helsinki. 23 pp. Available online: https://echa.europa.eu/guidance-documents/guidance-on-reach

ECHA (European Chemicals Agency) 2017a. Guidance on the Biocidal Products Regulation: Volume III Human Health, Assessment + Evaluation (Parts B+C). ECHA, Helsinki. 361 pp. Available online: https://echa.europa.eu/guidance-documents/guidance-on-biocides-legislation

ECHA (European Chemicals Agency), 2017b. Guidance on the Biocidal Products Regulation: Volume IV Environment, Assessment & Evaluation (Parts B+C) ECHA, Helsinki. 416 pp. Available online: https://echa.europa.eu/guidance-documents/guidance-on-biocides-legislation

ECHA (European Chemicals Agency), 2017c. Read-Across Assessment Framework (RAAF). ECHA, Helsinki. 60 pp. Available online: https://echa.europa.eu/documents/10162/13628/raaf_en.pdf

ECHA (European Chemicals Agency), 2017d. Read-across assessment framework (RAAF) - considerations on multiconstituent substances and UVCBs. ECHA, Helsinki. 40 pp. Available online: https://echa.europa.eu/documents/10162/13630/raaf_uvcb_report_en.pdf/3f79684d-07a5-e439-16c3-d2c8da96a316

ECHA (European Chemicals Agency) and EFSA (European Food Safety Authority), 2018. Outcome of the different consultations with relevant parties on the Guidance for the identification of endocrine disruptors in the context of Regulations (EU) No. 528/2012 and (EC) No. 1107/2009. EFSA supporting publication 2018:EN-1447. 9 pp. https://doi.org/10.2903/sp.efsa.2018.EN-1447; ECHA-18-G-01-EN

Edwards SW, Tan YM, Villeneuve DL, Meek ME and McQueen CA, 2016. Adverse Outcome Pathways-Organizing Toxicological Information to Improve Decision Making. Journal of Pharmacology and Experimental Therapeutics 356, 170–181. https://doi.org/10.1124/jpet.115.228239

EEA (European Environment Agency), 2012. The Impacts of Endocrine Disrupters on Wildlife, People and their Environments – The Weybridge+15 (1996–2011) Report. In: EEA Technical Reports. Publications Office of the European Union, Luxembourg. 116 pp. https://doi.org/10.2800/41462

EFSA (European Food Safety Authority), 2009. Risk Assessment for Birds and Mammals. EFSA Journal 2009;7(12):1438, 358 pp. https://doi.org/10.2903/j.efsa.2009.1438

EFSA (European Food Safety Authority), 2010. Application of systematic review methodology to food and feed safety assessments to support decision making. EFSA Journal 2010;8(6):1637, 90 pp. https://doi.org/10.2903/j.efsa.2010.1637

EFSA (European Food Safety Authority), 2011. Submission of scientific peer-reviewed open literature for the approval of pesticide active substances under Regulation (EC) No 1107/2009 (OJ L 309, 24.11.2009, p. 1–50). EFSA Journal 2011;9(2):2092, 49 pp. https://doi.org/10.2903/j.efsa.2011.2092

EFSA (European Food Safety Authority), 2015. Principles and process for dealing with data and evidence in scientific assessments. EFSA Journal 2015;13(5):4121, 36 pp. https://doi.org/10.2903/j.efsa.2015.4121

EFSA PPR Panel (EFSA Panel on Plant Protection Products and their Residues), 2013. Guidance on tiered risk assessment for plant protection products for aquatic organisms in edge-of-field surface waters. EFSA Journal 2013; 11(7):3290, 268 pp. https://doi.org/10.2903/j.efsa.2013.3290

EFSA PPR Panel (EFSA Panel on Plant Protection Products and their Residues), 2014. Scientific Opinion on good modelling practice in the context of mechanistic effect models for risk assessment of plant protection products. EFSA Journal 2014;12(3):3589, 92 pp. https://doi.org/10.2903/j.efsa.2014.3589

EFSA PPR Panel (EFSA Panel on Plant Protection Products and their Residues), Ockleford C, Adriaanse P, Berny P, Brock T, Duquesne S, Grilli S, Hougaard S, Klein M, Kuhl T, Laskowski R, Machera K, Pelkonen O, Pieper S, Smith R, Stemmer M, Sundh I, Teodorovic I, Tiktak A, Topping CJ, Wolterink G, Bottai M, Halldorsson T, Hamey P, Rambourg M-O, Tzoulaki I, Court Marques D, Crivellente F, Deluyker H and Hernandez-Jerez AF, 2017. Scientific Opinion of the PPR Panel on the follow-up of the findings of the External Scientific Report 'Literature review of epidemiological studies linking exposure to pesticides and health effects'. EFSA Journal 2017;15(10):5007, 101 pp. https://doi.org/10.2903/j.efsa.2017.5007

EFSA Scientific Committee, 2013. Scientific Opinion on the hazard assessment of endocrine disruptors: Scientific criteria for identification of endocrine disruptors and appropriateness of existing test methods for assessing effects mediated by these substances on human health and the environment. EFSA Journal 2013;11(3):84, 3132 pp. https://doi.org/10.2903/j.efsa.2013.3132

EFSA Scientific Committee, Benford D, Halldorsson T, Jeger MJ, Knutsen HK, More S, Naegeli H, Noteborn H, Colin Ockleford C, Ricci A, Rychen G, Schlatter JR, Silano V, Solecki R, Turck D, Younes M, Craig P, Hart A, Von Goetz N, Koutsoumanis K, Mortensen A, Ossendorp B, Germini A, Martino L, Merten C, Mosbach-Schulz O, Smith A and Hardy A, 2018. The principles and methods behind EFSA's Guidance on Uncertainty Analysis in Scientific Assessment. EFSA Journal 2018;16(1):5122, 235 pp. https://doi.org/10.2903/j.efsa.2018.5122

EFSA Scientific Committee, Hardy A, Benford D, Halldorsson T, Jeger MJ, Knutsen HK, More S, Naegeli H, Noteborn H, Ockleford C, Ricci A, Rychen G, Schlatter JR, Silano V, Solecki R, Turck D, Benfenati E, Chaudhry QM, Craig P, Frampton G, Greiner M, Hart A, Hogstrand C, Lambre C, Luttik R, Makowski D, Siani A, Wahlstroem H, Aguilera J, Dorne J-L, Fernandez Dumont A, Hempen M, Valtue SSilvia Valtue Martínez, Martino L, Smeraldi C, Terron A, Georgiadis N and Younes M, 2017. Guidance on the use of the weight of evidence approach in scientific assessments. EFSA Journal 2017;15(8):4971, 69 pp. https://doi.org/10.2903/j.efsa.2017.4971

European Commission, (Directorate-General for the Environment), wca Environment, Risk & Policy Analysts Ltd, 2018. Setting priorities for further development and validation of test methods and testing approaches for evaluating endocrine disruptors- Final Report (study awarded under Framework Contract ENV.A.3/FRA/2014/0029 on implementation of the Community strategy on Endocrine Disruptors). Publications Office of the European Union, Luxembourg. Available online: https://doi.org/10.2779/21828

European Commission (Directorate-General for the Environment), Kortenkamp A, Martin O, Faust M, Evans R, McKinlay R, Frances Orton F and Rosivatz E, 2011. State of The Art Assessment of Endocrine Disrupters, Final Report (Project Contract Number 070307/2009/550687/SER/D3). 135 pp. Available online: http://ec.europa.eu/environment/chemicals/endocrine/pdf/sota_edc_final_report.pdf

Fan X and Wu L, 2016. The impact of thyroid abnormalities during pregnancy on subsequent neuropsychological development of the offspring: a meta-analysis. J Matern Fetal Neonatal Med 29, 3971–3976. https://doi.org/10.3109/14767058.2016.1152248

Foo J and Lam T, 1993. Serum cortisol response to handling stress and the effect of cortisol implantation on testosterone level in the tilapia. Oreochromis mossambicus. Aquaculture, 115, 145–158. https://doi.org/10.1016/0044-8486(93)90365-6

Ghassabian A, El Marroun H, Peeters RP, Jaddoe VW, Hofman A, Verhulst FC, Tiemeier H and White T, 2014. Downstream effects of maternal hypothyroxinemia in early pregnancy: nonverbal IQ and brain morphology in school-age children. Journal of Clinical Endocrinology and Metabolism 99, 2383–2390. https://doi.org/10.1210/jc.2013-4281

Grim KC, Wolfe M, Braunbeck T, Iguchi T, Ohta Y, Tooi O, Touart L, Wolf DC and Tietge J, 2009. Thyroid histopathology assessments for the amphibian metamorphosis assay to detect thyroid-active substances. Toxicologic Pathology 37, 415–424. https://doi.org/10.1177/0192623309335063

Helfman G, Collette B and Facey D, 1997. *The diversity of fishes*. John Wiley & Sons Ltd. (Blackwell Publishing), Chichester, West Sussex. 737 pp.

Hotchkiss AK, Lambright CS, Ostby JS, Parks-Saldutti L, Vandenbergh JG and Gray LE, Jr, 2007. Prenatal testosterone exposure permanently masculinizes anogenital distance, nipple development, and reproductive tract morphology in female Sprague-Dawley rats. Toxicological Sciences 96, 335–345. https://doi.org/10.1093/toxsci/kfm002

Hutchinson TH, Bogi C, Winter MJ and Owens JW, 2009. Benefits of the maximum tolerated dose (MTD) and maximum tolerated concentration (MTC) concept in aquatic toxicology. Aquatic Toxicology 91, 197–202. https://doi.org/10.1016/j.aquatox.2008.11.009

JRC (Joint Research Centre), 2013. Key scientific issues relevant to the identification and characterisation of endocrine disrupting substances - Report of the Endocrine Disrupters Expert Advisory Group. In: JRC Science and Policy Reports. Publications Office of the European Union, Luxembourg. 34 pp. https://doi.org/10.2788/8659

JRC (Joint Research Centre), 2014. Alternative methods for regulatory toxicology – a state-of-the-art review. In: JRC Science and Policy Reports. Publications Office of the European Union, Luxembourg. 470 pp. https://doi.org/doi:10.2788/11111

JRC (Joint Research Centre), 2016. Screening methodology to identify potential endocrine disruptors according to different options in the context of an impact assessment. In: JRC Science for policy reports. Publications Office of the European Union, Luxembourg. 54 pp. https://doi.org/10.2788/73203

Judson RS, Magpantay FM, Chickarmane V, Haskell C, Tania N, Taylor J, Xia MH, Huang RL, Rotroff DM, Filer DL, Houck KA, Martin MT, Sipes N, Richard AM, Mansouri K, Setzer RW, Knudsen TB, Crofton KM and Thomas RS, 2015. Integrated Model of Chemical Perturbations of a Biological Pathway Using 18 In Vitro High-Throughput Screening Assays for the Estrogen Receptor. Toxicological Sciences 148, 137–154. https://doi.org/10.1093/toxsci/kfv168

Kleinstreuer NC, Ceger P, Watt ED, Martin M, Houck K, Browne P, Thomas RS, Casey WM, Dix DJ, Allen D, Sakamuru S, Xia M, Huang R and Judson R, 2017. Development and Validation of a Computational Model for Androgen Receptor Activity. Chemical Research in Toxicology 30, 946–964. https://doi.org/10.1021/acs.chemrestox.6b00347

Klimisch HJ, Andreae M and Tillmann U, 1997. A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data. Regulatory Toxicology and Pharmacology 25, 1–5. https://doi.org/10.1006/rtph.1996.1076

Kohler HR and Triebskorn R, 2013. Wildlife Ecotoxicology of Pesticides: can We Track Effects to the Population Level and Beyond? Science 341, 759–765. https://doi.org/10.1126/science.1237591

Kramer VJ, Etterson MA, Hecker M, Murphy CA, Roesijadi G, Spade DJ, Spromberg JA, Wang M and Ankley GT, 2011. Adverse Outcome Pathways and Ecological Risk Assessment Bridging to Population-Level Effects. Environmental Toxicology and Chemistry 30, 64–76. https://doi.org/10.1002/etc.375

Kwon HC, Choi SH, Kim YU, Son SO and Kwon JY, 2005. Androgen action on hepatic vitellogenin synthesis in the eel, Anguilla japonica is suppressed by an androgen receptor antagonist. Journal of Steroid Biochemistry and Molecular Biology 96, 175–178. https://doi.org/10.1016/j.jsbmb.2005.02.010

Li JZ and Gramatica P, 2010. Classification and Virtual Screening of Androgen Receptor Antagonists. Journal of Chemical Information and Modeling 50, 861–874. https://doi.org/10.1021/ci100078u

Lo Piparo E and Worth A, 2010. Review of (Q)SAR Models and Software Tools for Predicting Developmental and Reproductive Toxicity. In: JRC Scientific and Technical Reports, Place: Published https://doi.org/10.2788/9628

Lundholm CD, 1997. DDE-induced eggshell thinning in birds: effects of p,p'-DDE on the calcium and prostaglandin metabolism of the eggshell gland. Comp Biochem Physiol C Pharmacol Toxicol Endocrinol 118, 113–128. https://www.ncbi.nlm.nih.gov/pubmed/9490182

Manibusan MK and Touart LW, 2017. A comprehensive review of regulatory test methods for endocrine adverse health effects. Critical Reviews in Toxicology 47, 433–481. https://doi.org/10.1080/10408444.2016.1272095

Marty MS, Blankinship A, Chambers J, Constantine L, Kloas W, Kumar A, Lagadic L, Meador J, Pickford D, Schwarz T and Verslycke T, 2017. Population-relevant endpoints in the evaluation of endocrine-active substances (EAS) for ecotoxicological hazard and risk assessment. Integrated Environmental Assessment and Management 13, 317–330. https://doi.org/10.1002/ieam.1887

Matthiessen P and Weltje L, 2015. A review of the effects of azole compounds in fish and their possible involvement in masculinization of wild fish populations. Critical Reviews in Toxicology 45, 453–467. https://doi.org/10.3109/10408444.2015.1018409

Matthiessen P, Wheeler JR and Weltje L, 2018. A review of the evidence for endocrine disrupting effects of current-use chemicals on wildlife populations. Critical Reviews in Toxicology 48, 195–216. https://doi.org/10.1080/10408444.2017.1397099

McGee TD, Edwards J and Roitberg AE, 2008. Preliminary molecular dynamic simulations of the estrogen receptor alpha ligand binding domain from antagonist to apo. Int J Environ Res Public Health 5, 111–114. https://www.ncbi.nlm.nih.gov/pubmed/18678925

Meek ME, Boobis A, Cote I, Dellarco V, Fotakis G, Munn S, Seed J and Vickers C, 2014a. New developments in the evolution and application of the WHO/IPCS framework on mode of action/species concordance analysis. Journal of Applied Toxicology 34, 1–18. https://doi.org/10.1002/jat.2949

Meek ME, Palermo CM, Bachman AN, North CM and Jeffrey Lewis R, 2014b. Mode of action human relevance (species concordance) framework: Evolution of the Bradford Hill considerations and comparative analysis of weight of evidence. Journal of Applied Toxicology 34, 595–606. https://doi.org/10.1002/jat.2984

Mintram KS, Brown AR, Maynard SK, Thorbek P and Tyler CR, 2018. Capturing ecology in modeling approaches applied to environmental risk assessment of endocrine active chemicals in fish. Critical Reviews in Toxicology 48, 109–120. https://doi.org/10.1080/10408444.2017.1367756

Nakamura M, Kobayashi T, Chang XT and Nagahama Y, 1998. Gonadal sex differentiation in teleost fish. Journal of Experimental Zoology 281, 362–372. https://doi.org/10.1002/(sici)1097-010x(19980801)281:5<362::aid-jez3>3.0.co, 2-M

Nanda I, Hornung U, Kondo M, Schmid M and Schartl M, 2003. Common spontaneous sex-reversed XX males of the medaka Oryzias latipes. Genetics 163, 245–251. https://www.ncbi.nlm.nih.gov/pubmed/12586712

Nelson KR, Schroeder AL, Ankley GT, Blackwell BR, Blanksma C, Degitz SJ, Flynn KM, Jensen KM, Johnson RD, Kahl MD, Knapen D, Kosian PA, Milsk RY, Randolph EC, Saari T, Stinckens E, Vergauwen L and Villeneuve DL, 2016. Impaired anterior swim bladder inflation following exposure to the thyroid peroxidase inhibitor 2-mercaptobenzothiazole part I: fathead minnow. Aquatic Toxicology, 173, 192–203. https://doi.org/10.1016/j.aquatox.2015.12.024

OECD (Organisation for Economic Co-operation and Development), 1981a. Test No. 410: Repeated Dose Dermal Toxicity: 21/28-day Study. In: OECD Guidelines for the Testing of Chemicals, Section 4. OECD Publishing, Paris. 8 pp. https://doi.org/10.1787/9789264070745-en

OECD (Organisation for Economic Co-operation and Development), 1981b. Test No. 411: Subchronic Dermal Toxicity: 90-day Study. In: OECD Guidelines for the Testing of Chemicals, Section 4. OECD Publishing, Paris. 9 pp. https://doi.org/10.1787/9789264070769-en

OECD (Organisation for Economic Co-operation and Development), 1983. Test No. 415: One-Generation Reproduction Toxicity Study. In: OECD Guidelines for the Testing of Chemicals, Section 4. OECD Publishing, Paris. 8 pp. https://doi.org/10.1787/9789264070844-en

OECD (Organisation for Economic Co-operation and Development), 1984. Test No. 206: Avian Reproduction Test. In: OECD Guidelines for the Testing of Chemicals, Section 2 OECD Publishing, Paris. 12 pp. https://doi.org/10.1787/9789264070028-en

OECD (Organisation for Economic Co-operation and Development), 1998. Test No. 409: Repeated Dose 90-Day Oral Toxicity Study in Non-Rodents. In: OECD Guidelines for the Testing of Chemicals, Section 4. OECD Publishing, Paris. 9 pp. https://doi.org/10.1787/9789264070721-en

OECD (Organisation for Economic Co-operation and Development), 1999. The Application of the GLP Principles to Field Studies. In: OECD Series on Principles of Good Laboratory Practice and Compliance Monitoring. OECD Publishing, Paris. 16 pp. https://doi.org/10.1787/9789264078635-en

OECD (Organisation for Economic Co-operation and Development), 2000. Guidance Document on the Recognition, Assessment and Use of Clinical Signs as Human Endpoints for Experimental Animals Used in Safety Evaluation. In: OECD series on testing and assessment. OECD Publishing, Paris. 39 pp. https://doi.org/10.1787/20777876

OECD (Organisation for Economic Co-operation and Development), 2001. Test No. 416: Two-Generation Reproduction Toxicity. In: OECD Guidelines for the Testing of Chemicals, Section 4. OECD Publishing, Paris. 13 pp. https://doi.org/10.1787/9789264070868-en

OECD (Organisation for Economic Co-operation and Development), 2002. Report of the SETAC/OECD workshop on avian toxicity testing. In: OECD series on testing and assessment. OECD Publishing, Paris. 198 pp. Available online: http://www.oecd.org/officialdocuments/displaydocument/?cote=ocde/gd(96)166&doclanguage=en

OECD (Organisation for Economic Co-operation and Development), 2004. Detailed Review Paper on Fish Screening Assays for the Detection of Endocrine Active Substances. In: OECD series on testing and assessment. OECD Publishing, Paris. 170 pp. https://doi.org/10.1787/9789264079151-en

OECD (Organisation for Economic Co-operation and Development), 2006a. OECD report of the initial work towards the validation of the rodent uterothophic assay - Phase 1. In: OECD Series on Testing and Assessment. OECD Publishing, Paris. 194 pp. Available online: http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2006)33&doclanguage=en

OECD (Organisation for Economic Co-operation and Development), 2006b. OECD report of the validation of the rodent uterotrophic bioassay - Phase 2: Testing of potent and weak oestrogen agonists by multiple laboratories. In: OECD Series on Testing and Assessment. OECD Publishing, Paris. 142 pp. Available online: http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2006)34&doclanguage=en

OECD (Organisation for Economic Co-operation and Development), 2006c. Summary report of the uterotrophic bioassay peer review panel, including agreement of the working group of national coordinators of the test guidelines programme on the follow-up of this report. In: OECD Series on Testing and Assessment. OECD Publishing, Paris. 16 pp. Available online: http://www.oecd.org/officialdocuments/displaydocument/?cote=env/jm/mono(2006)37&doclanguage=en

OECD (Organisation for Economic Co-operation and Development), 2007a. Detailed review paper for avian two-generation toxicity test. In: OECD Series on Testing and Assessment. OECD, Paris. 167 pp. Available online: http://www.oecd.org/officialdocuments/displaydocument/?cote=env/jm/mono(2007)21&doclanguage=en

OECD (Organisation for Economic Co-operation and Development), 2007b. Guidance Document on the Uterotrophic bioassay – Procedure to test for antioestrogenicity. In: OECD Series on Testing and Assessment. OECD Publishing, Paris. 31 pp. Available online: http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2007)15&doclanguage=en

OECD (Organisation for Economic Co-operation and Development), 2007c. Test No. 426: Developmental Neurotoxicity Study. In: OECD Guidelines for the Testing of Chemicals, Section 4. OECD Publishing, Paris. 26 pp. https://doi.org/10.1787/9789264067394-en

OECD (Organisation for Economic Co-operation and Development), 2007d. Test No. 440: Uterotrophic Bioassay in Rodents. In: OECD Guidelines for the Testing of Chemicals, Section 4. OECD Publishing, Paris. 21 pp. https://doi.org/10.1787/9789264067417-en

OECD (Organisation for Economic Co-operation and Development), 2008. Test No. 407: Repeated Dose 28-day Oral Toxicity Study in Rodents. In: OECD Guidelines for the Testing of Chemicals, Section 4. OECD Publishing, Paris. 13 pp. https://doi.org/10.1787/9789264070684-en

OECD (Organisation for Economic Co-operation and Development), 2009a. Guidance Document on the weanling Hershberger bioassay in rats: a short-term screening assay for (anti)androgenic properties. In: OECD Series on Testing and Assessment. OECD Publishing, Paris. 34 pp. Available online: http://www.oecd.org/officialdocuments/displaydocument/?cote=ENV/JM/MONO(2009)41&doclanguage=en

OECD (Organisation for Economic Co-operation and Development), 2009b. Test No. 230: 21-day Fish Assay. In: OECD Guidelines for the Testing of Chemicals, Section 2 OECD Publishing, Paris. 38 pp. https://doi.org/10.1787/9789264076228-en

OECD (Organisation for Economic Co-operation and Development), 2009c. Test No. 231: Amphibian Metamorphosis Assay. In: OECD Guidelines for the Testing of Chemicals, Section 2 OECD Publishing, Paris. 33 pp. https://doi.org/10.1787/9789264076242-en

OECD (Organisation for Economic Co-operation and Development), 2009d. Test No. 441: Hershberger Bioassay in Rats. In: OECD Guidelines for the Testing of Chemicals, Section 4. OECD Publishing, Paris. 20 pp. https://doi.org/10.1787/9789264076334-en

OECD (Organisation for Economic Co-operation and Development), 2009e. Test No. 451: Carcinogenicity Studies. In: OECD Guidelines for the Testing of Chemicals, Section 4. OECD Publishing, Paris. 15 pp. https://doi.org/10.1787/9789264071186-en

OECD (Organisation for Economic Co-operation and Development), 2009f. Test No. 452: Chronic Toxicity Studies. In: OECD Guidelines for the Testing of Chemicals, Section 4. OECD Publishing, Paris. 16 pp. https://doi.org/10.1787/9789264071209-en

OECD (Organisation for Economic Co-operation and Development), 2009g. Test No. 453: Combined Chronic Toxicity/Carcinogenicity Studies. In: OECD Guidelines for the Testing of Chemicals, Section 4. OECD Publishing, Paris. 20 pp. https://doi.org/10.1787/9789264071223-en

OECD (Organisation for Economic Co-operation and Development), 2010. Guidance document on the diagnosis of endocrine-related histopathology in fish gonads. In: OECD series on testing and assessment. OECD Publishing, Paris. 114 pp. Available online: http://www.oecd.org/officialdocuments/displaydocument/?cote=env/jm/mono(2010)14&doclanguage=en

OECD (Organisation for Economic Co-operation and Development), 2011a. Guidance Document on the Androgenised Female Stickleback screen. In: OECD Series on Testing and Assessment. OECD Publishing, Paris. 47 pp. Available online: http://www.oecd.org/chemicalsafety/testing/48532191.pdf

OECD (Organisation for Economic Co-operation and Development), 2011b. Test No. 234: Fish Sexual Development Test. In: OECD Guidelines for the Testing of Chemicals, Section 2 OECD Publishing, Paris. https://doi.org/10.1787/9789264122369-en

OECD (Organisation for Economic Co-operation and Development), 2011c. Test No. 456: H295R Steroidogenesis Assay. In: OECD Guidelines for the Testing of Chemicals, Section 4. OECD Publishing, Paris. 21 pp. https://doi.org/10.1787/9789264122642-en

OECD (Organisation for Economic Co-operation and Development), 2012a. Test No. 229: Fish Short Term Reproduction Assay. In: OECD Guidelines for the Testing of Chemicals, Section 2. OECD Publishing, Paris. 40 pp. https://doi.org/10.1787/9789264185265-en

OECD (Organisation for Economic Co-operation and Development), 2012b. Test No. 443: Extended One-Generation Reproductive Toxicity Study. In: OECD Guidelines for the Testing of Chemicals, Section 4. OECD Publishing, Paris. 25 pp. https://doi.org/10.1787/9789264185371-en

OECD (Organisation for Economic Co-operation and Development), 2012c. Test No. 455: Performance-Based Test Guideline for Stably Transfected Transactivation In Vitro Assays to Detect Estrogen Receptor Agonists and Antagonists. In: OECD Guidelines for the Testing of Chemicals, Section 4. OECD Publishing, Paris. 39 pp. https://doi.org/10.1787/9789264265295-en

OECD (Organisation for Economic Co-operation and Development), 2013a. Guidance Document supporting OECD Test Guideline 443 on the extended one-generation reproductive toxicity test In: OECD series on testing and assessment. OECD Publishing, Paris. 68 pp. Available online: http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/JM/MONO(2013)10&doclanguage=en

OECD (Organisation for Economic Co-operation and Development), 2013b. Test No. 210: Fish, Early-life Stage Toxicity Test. In: OECD Guidelines for the Testing of Chemicals, Section 2 OECD Publishing, Paris. 24 pp. https://doi.org/10.1787/9789264203785-en

OECD (Organisation for Economic Co-operation and Development), 2014. Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models. In: OECD series on testing and assessment. OECD Publishing, Paris. 154 pp. https://doi.org/10.1787/9789264085442-en

OECD (Organisation for Economic Co-operation and Development), 2015a. Guidance Document on histopathology techniques and evaluation (Part I) for the larval amphibian growth and development assay (LAGDA). In: OECD series on testing and assessment. OECD Publishing, Paris. 28 pp. https://www.oecd.org/env/ehs/testing/37_PART1.pdf

OECD (Organisation for Economic Co-operation and Development), 2015b. Guidance Document on histopathology techniques and evaluation (Part II) for the larval amphibian growth and development assay (LAGDA). In: OECD series on testing and assessment. OECD Publishing, Paris. 23 pp. https://www.oecd.org/env/ehs/testing/37_PART2.pdf

OECD (Organisation for Economic Co-operation and Development), 2015c. Test No. 240: Medaka Extended One Generation Reproduction Test (MEOGRT). In: OECD Guidelines for the Testing of Chemicals, Section 2 OECD Publishing, Paris. 39 pp. https://doi.org/10.1787/9789264242258-en

OECD (Organisation for Economic Co-operation and Development), 2015d. Test No. 241: The Larval Amphibian Growth and Development Assay (LAGDA). In: OECD Guidelines for the Testing of Chemicals, Section 2 OECD Publishing, Paris. 38 pp. https://doi.org/10.1787/9789264242340-en

OECD (Organisation for Economic Co-operation and Development), 2015e. Test No. 493: Performance-Based Test Guideline for Human Recombinant Estrogen Receptor (hrER) In Vitro Assays to Detect Chemicals with ER Binding Affinity. In: OECD Guidelines for the Testing of Chemicals, Section 4. OECD Publishing, Paris. 64 pp. https://doi.org/10.1787/9789264242623-en

OECD (Organisation for Economic Co-operation and Development), 2016a. Test No. 421: Reproduction/Developmental Toxicity Screening Test. In: OECD Guidelines for the Testing of Chemicals, Section 4. OECD Publishing, Paris. 10 pp. https://doi.org/10.1787/9789264264380-en

OECD (Organisation for Economic Co-operation and Development), 2016b. Test No. 422: Combined Repeated Dose Toxicity Study with the Reproduction/Developmental Toxicity Screening Test. In: OECD Guidelines for the Testing of Chemicals, Section 4. OECD Publishing, Paris. 14 pp. https://doi.org/10.1787/9789264264403-en

OECD (Organisation for Economic Co-operation and Development), 2016c. Users' Handbook supplement to the Guidance Document for developing and assessing Adverse Outcome Pathways. In: OECD Series on Adverse Outcome Pathways. OECD Publishing, Paris. 52 pp. https://doi.org/10.1787/5jlv1m9d1g32-en

OECD (Organisation for Economic Co-operation and Development), 2017a. Guidance Document for Describing Non-Guideline In Vitro Test Methods. In: OECD Series on testing and assessment. OECD Publishing, Paris. 17 pp. https://doi.org/10.1787/9789264274730-en

OECD (Organisation for Economic Co-operation and Development), 2017b. Guidance document on Developing and Assessing Adverse Outcome Pathways. In: OECD Series on Testing and Assessment. OECD Publishing, Paris. 52 pp. Available online: http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2013)6&doclanguage=en

OECD (Organisation for Economic Co-operation and Development), 2017c. Test No. 412: Subacute Inhalation Toxicity: 28-Day Study. In: OECD Guidelines for the Testing of Chemicals, Section 4. OECD Publishing, Paris. 23 pp. https://doi.org/10.1787/9789264070783-en

OECD (Organisation for Economic Co-operation and Development), 2017d. Test No. 413: Subchronic Inhalation Toxicity: 90-day Study. In: OECD Guidelines for the Testing of Chemicals, Section 4. OECD Publishing, Paris. 24 pp. https://doi.org/10.1787/9789264070806-en

OECD (Organisation for Economic Co-operation and Development), 2017e. Work plan for the test Guidelines Programme (TGP). Available online: http://www.oecd.org/chemicalsafety/testing/TGP%20work%20plan_August%202017.pdf.

OECD (Organisation for Economic Co-operation and Development). 2018a. Guidance document on Good In Vitro Method Practices (GIVIMP) for the development and Implementation of In Vitro Methods for Regulatory use in Human Safety Assessment In: OECD Series on Testing and Assessment. OECD Publishing, Paris. XX pp. {doi currently not available to be added before publication in EFSA Journal}

OECD (Organisation for Economic Co-operation and Development). 2018b. Guidance Document on standardised test guidelines for evaluating chemicals for endocrine disruption. In: OECD Series on Testing and Assessment. OECD Publishing, Paris. XX pp. {doi currently not available to be added before publication in EFSA Journal}

OECD (Organisation for Economic Co-operation and Development), 2018c. Test No. 408: Repeated Dose 90-Day Oral Toxicity Study in Rodents. In: OECD Guidelines for the Testing of Chemicals, Section 4. OECD Publishing, Paris. XX pp. {doi currently not available to be added before publication in EFSA Journal}

OECD (Organisation for Economic Co-operation and Development), 2018d. Test No. 414: Prenatal Development Toxicity Study. In: OECD Guidelines for the Testing of Chemicals, Section 4. OECD Publishing, Paris. XX pp. {doi currently not available to be added before publication in EFSA Journal}

Östlund-Nilsson S, Mayer I and Huntingford AF, 2006. Biology of the three-spined stickleback. In: *CRC Marine Biology Series*. CRC Press, Boca Raton (FL). 408 pp.

Palace VP, Evans RE, Wautier KG, Mills KH, Blanchfield PJ, Park BJ, Baron CL and Kidd KA, 2009. Interspecies differences in biochemical, histopathological, and population responses in four wild fish species exposed to ethynylestradiol added to a whole lake. Canadian Journal of Fisheries and Aquatic Sciences 66, 1920–1935. https://doi.org/10.1139/f09-125

Panaye A, Doucet JP, Devillers J, Marchand-Geneste N and Porcher JM, 2008. Decision trees versus support vector machine for classification of androgen receptor ligands. Sar and Qsar in Environmental Research 19, 129–151. https://doi.org/10.1080/10629360701843441

Payne AH and Hales DB, 2004. Overview of steroidogenic enzymes in the pathway from cholesterol to active steroid hormones. Endocrine Reviews 25, 947–970. https://doi.org/10.1210/er.2003-0030

Polzonetti-Magni A, Mosconi G, Palermo F, Cottone E, Forno S and Franzoni M, 2004. Amphibians as a model for studying xenoestrogen-dependent neuroendocrine disruption. Ital. J. Zool. 95–100. https://www.tandfonline.com/doi/pdf/10.1080/11250000409356615

Renjith P and Jegatheesan K, 2015. Machine learning classifier algorithms to predict endocrine toxicity of chemicals. International Journal of Toxicological and Pharmacological Research, 7, 303–308.

Ribas L, Valdivieso A, Diaz N and Piferrer F, 2017. Appropriate rearing density in domesticated zebrafish to avoid masculinization: links with the stress response. Journal of Experimental Biology 220(Pt 6), 1056–1064. https://doi.org/10.1242/jeb.144980

Ribay K, Kim M, Wang W, Pinolini D and Zhu H, 2016. Predictive modelling of estrogen receptor binding agents using advanced cheminformatics tools and massive public data. Frontiers in Environmental Science 4, https://doi.org/10.3389/fenvs.2016.00012

Schapaugh AW, McFadden LG, Zorrilla LM, Geter DR, Stuchal LD, Sunger N and Borgert CJ, 2015. Analysis of EPA's endocrine screening battery and recommendations for further review. Regulatory Toxicology and Pharmacology 2015;72, 552–561. https://doi.org/10.1016/j.yrtph.2015.05.028

Scholz S and Kluver N, 2009. Effects of endocrine disrupters on sexual, gonadal development in fish. Sex Dev 3, 136–151. https://doi.org/10.1159/000223078

Scott HM, Mason JI and Sharpe RM, 2009. Steroidogenesis in the fetal testis and its susceptibility to disruption by exogenous compounds. Endocrine Reviews 30, 883–925. https://doi.org/10.1210/er.2009-0016

Sebillot A, Damdimopoulou P, Ogino Y, Spirhanzlova P, Miyagawa S, Du Pasquier D, Mouatassim N, Iguchi T, Lemkine GF, Demeneix BA and Tindall AJ, 2014. Rapid fluorescent detection of (anti)androgens with spiggin-gfp medaka. Environmental Science and Technology 48, 10919–10928. https://doi.org/10.1021/es5030977

Slater EP, Redeuilh G and Beato M, 1991. Hormonal regulation of vitellogenin genes: an estrogen-responsive element in the *Xenopus* A2 gene and a multihormonal regulatory region in the chicken II gene. Molecular Endocrinology 5, 386–396. https://doi.org/10.1210/mend-5-3-386

Stinckens E, Vergauwen L, Schroeder AL, Maho W, Blackwell BR, Witters H, Blust R, Ankley GT, Covaci A, Villeneuve DL and Knapen D, 2016. Impaired anterior swim bladder inflation following exposure to the thyroid peroxidase inhibitor 2-mercaptobenzothiazole part II: Zebrafish. Aquatic Toxicology, 173, 204–217. https://doi.org/10.1016/j.aquatox.2015.12.023

Susser M, 1991. What is a cause and how do we know one? A grammar for pragmatic epidemiology. American Journal of Epidemiology 133, 635–648. https://www.ncbi.nlm.nih.gov/pubmed/2018019

Topping CJ and Luttik R, 2017. Simulation to aid in interpreting biological relevance and setting of population-level protection goals for risk assessment of pesticides. Regulatory Toxicology and Pharmacology, 89, 40–49. https://doi.org/10.1016/j.yrtph.2017.07.011

Trigunaite A, Dimo J and Jorgensen TN, 2015. Suppressive effects of androgens on the immune system. Cellular Immunology 294, 87–94. https://doi.org/10.1016/j.cellimm.2015.02.004

US EPA (United States Environmental Protection Agency), 1996. OPPTS 850.1500 Fish life cycle toxicity [EPA 712–C–96–122]. In. US EPA, Cincinnati (OH). 4 pp. Available online: https://www.epa.gov/sites/production/files/2015-07/documents/850-1500.pdf

US EPA (United States Environmental Protection Agency), 2009a. OCSPP 890.2100: Avian Two-Generation Toxicity Test in the Japanese Quail [EPA No. 740-C-15-003]. In: Endocrine Disruptor Screening Program Test Guidelines. United States Environmental Protection Agency's Office of Chemical Safety and Pollution Prevention (OCSPP), US EPA, Washington (DC)pp. Available online: https://www.regulations.gov/contentStreamer?documentId=EPA-HQ-OPPT-2009-0576-0017&contentType=msw12

US EPA (United States Environmental Protection Agency), 2009b. OPPTS 890.1200: Aromatase (Human Recombinant) [EPA 740-C-09-004]. In: Endocrine Disruptor Screening Program Test Guidelines. Office of Prevention, Pesticides and Toxic Substances (OPPTS), US EPA, Washington (DC). 24 pp. Available online: https://www.regulations.gov/contentStreamer?documentId=EPA-HQ-OPPT-2009-0576-0004&contentType=pdf

US EPA (United States Environmental Protection Agency), 2009c. OPPTS 890.1450: Pubertal Development and Thyroid Function in Intact Juvenile/Peripubertal Female Rats [EPA 740-C-09-009]. In: Endocrine Disruptor Screening Program Test Guidelines. Office of Prevention, Pesticides and Toxic Substances (OPPTS), US EPA, Washington (DC). 31 pp. Available online: https://www.regulations.gov/contentStreamer?documentId=EPA-HQ-OPPT-2009-0576-0009&contentType=pdf

US EPA (United States Environmental Protection Agency), 2009d. OPPTS 890.1500: Pubertal Development and Thyroid Function in Intact Juvenile/Peripubertal Male Rats [EPA 740-C-09-012]. In: Endocrine Disruptor Screening Program Test Guidelines. Office of Prevention, Pesticides and Toxic Substances (OPPTS), US EPA, Washington (DC). 28 pp. Available online: https://www.regulations.gov/contentStreamer?documentId=EPA-HQ-OPPT-2009-0576-0010&contentType=pdf

Vedani A, Dobler M and Smiesko M, 2012. VirtualToxLab - a platform for estimating the toxic potential of drugs, chemicals and natural products. Toxicology and Applied Pharmacology 261, 142–153. https://doi.org/10.1016/j.taap.2012.03.018

Vermeire T, Aldenberg T, Buist H, Escher S, Mangelsdorf I, Paune E, Rorije E and Kroese D, 2013. OSIRIS, a quest for proof of principle for integrated testing strategies of chemicals for four human health endpoints. Regulatory Toxicology and Pharmacology 67, 136–145. https://doi.org/10.1016/j.yrtph.2013.01.007

Wang M, Chai L, Zhao H, Wu M and Wang H, 2015. Effects of nitrate on metamorphosis, thyroid and iodothyronine deiodinases expression in Bufo gargarizans larvae. Chemosphere, 139, 402–409. https://doi.org/10.1016/j.chemosphere.2015.07.037

Warren GL, Andrews CW, Capelli AM, Clarke B, LaLonde J, Lambert MH, Lindvall M, Nevins N, Semus SF, Senger S, Tedesco G, Wall ID, Woolven JM, Peishoff CE and Head MS, 2006. A critical assessment of docking programs and scoring functions. Journal of Medicinal Chemistry 49, 5912–5931. https://doi.org/10.1021/jm050362n

Wheeler J, Panter G, Weltje L and Thorpe KL, 2013. Test concentration setting for fish in vivo endocrine screening assays. Chemosphere, 92, 1067–1076.

White ER, Nagy JD and Gruber SH, 2014. Modeling the population dynamics of lemon sharks. Biol Direct 9, 23. https://doi.org/10.1186/1745-6150-9-23

WHO/IPCS (World Health Organization/International Programme on Chemical Safety), 2002. Global Assessment of the State-of-the-Science of Endocrine Disruptors. Journal 2002. Available online: http://www.who.int/ipcs/publications/new_issues/endocrine_disruptors/en/

WHO/IPCS (World Health Organization/International Programme on Chemical Safety), 2009. Principles and Methods for the Risk Assessment of Chemicals in Food. Environmental Health Criteria 240, 689 pp. Available online: http://www.who.int/foodsafety/chem/principles/en/index1.html

Wu S, Fisher J, Naciff J, Laufersweiler M, Lester C, Daston G and Blackburn K, 2013. Framework for identifying chemicals with structural features associated with the potential to act as developmental or reproductive toxicants. Chemical Research in Toxicology 26, 1840–1861. https://doi.org/10.1021/tx400226u

Yang C, Tarkhov A, Marusczyk J, Bienfait B, Gasteiger J, Kleinoeder T, Magdziarz T, Sacher O, Schwab CH, Schwoebel J, Terfloth L, Arvidson K, Richard A, Worth A and Rathman J, 2015. New publicly available chemical query language, CSRML, to support chemotype representations for application to data mining and modeling. Journal of Chemical Information and Modeling 55, 510–528. https://doi.org/10.1021/ci500667v

Zhang L, Sedykh A, Tripathi A, Zhu H, Afantitis A, Mouchlis VD, Melagraki G, Rusyn I and Tropsha A, 2013. Identification of putative estrogen receptor-mediated endocrine disrupting chemicals using QSAR- and structure-based virtual screening approaches. Toxicology and Applied Pharmacology 272, 67–76. https://doi.org/10.1016/j.taap.2013.04.032

Zhao CY, Zhang RS, Zhang HX, Xue CX, Liu HX, Liu MC, Hu ZD and Fan BT, 2005. QSAR study of natural, synthetic and environmental endocrine disrupting compounds for binding to the androgen receptor. SAR and QSAR in Environmental Research 16, 349–367. https://doi.org/10.1080/10659360500204368

Abbreviations

ADME	adsorption, distribution, metabolism, excretion
AFSS	androgenised female stickleback screen
AGD	anogenital distance
AhR	aryl hydrocarbon receptor
AMA	amphibian metamorphosis assay
AOP	adverse outcome pathway
AR	androgen receptor
BP	biocidal product
CAR	Competent authority report
CAS	Chemical Abstracts Service
CoMFA	comparative molecular field analysis
CF	Conceptual framework
CV	coefficient of variation
DAR	Draft Assessment Report
DIT	developmental immunotoxicity
DNT	developmental neurotoxicity
EASZY	Detection of endocrine active substances, acting through estrogen receptors using transgenic cyp 19a1b-GFP zebrafish embryos
EATS	Estrogen, androgen, thyroid, steroidogenic
ECHA	European Chemicals Agency
ED	endocrine disruptor
EFSA	European Food Safety Authority
EOGRTS	extended one-generation reproductive toxicity study (OECD TG 443)
ER	estrogen receptor
FDA	Food and Drug Administration (United States)
FLCTT	fish life cycle toxicity test (OPPTS 850.1500)
FSDT	fish sexual development test (OECD TG 234)
FSH	follicle-stimulating hormone
FSTRA	fish short-term reproduction assay (OECD TG 229)
FXR	farnesoid X receptor
GD	Guidance document
GFP	green fluorescent protein
GIVIMP	Good In Vitro Method Practices
GR	glucocorticoid receptor
GSI	gonadosomatic index
HDL	high-density lipoprotein
HPG	hypothalamic–pituitary–gonadal
HPT	hypothalamic–pituitary–thyroid
IPCS	International Programme on Chemical Safety
IUPAC	International Union of Pure and Applied Chemistry
JMASA	juvenile medaka anti-androgen screening assay
JRC	Joint Research Centre
KE	key event
KER	key event relationship
LABC	levator ani/bulbocavernosus muscle complex
LAGDA	larval amphibian growth and development assay (OECD TG 241)
LDL	low-density lipoprotein
LH	luteinising hormone
LLoQ	lower limit of quantification
LOAEL	lowest-observed-adverse-effect level
LOD	limit of detection
LXR	liver X receptor
MEOGRT	Medaka extended one-generation reproduction test (OECD TG 240)
MIE	molecular initiating event
MoA	mode of action
MTC	maximum tolerated concentration

MTD	maximum tolerated dose
NIS	sodium–iodide symporter
NMDR	non-monotonic dose response
NR	nuclear receptor
OECD	Organisation for Economic Co-operation and Development
OPPTS	Office of Prevention, Pesticides and Toxic Substances (United States)
PND	postnatal day
PPAR	peroxisome proliferator-activated receptor
PPP	Plant protection product
PR	progesterone receptor
PXR	pregnane X receptor
RADAR	rapid androgen disruption adverse outcome reporter assay
RAR	Renewal Assessment Report
RXR	retinoic acid receptor
(Q)SAR	(quantitative) structure–activity relationship
SAR	structure–activity relationship
SSC	secondary sex characteristics
T3	triiodothyronine
T4	thyroxine
TG	Test guideline
TH	thyroid hormone
TPO	thyroid peroxidase
TR	thyroid hormone receptor
TRH	thyrotropin-releasing hormone
TSH	thyroid-stimulating hormone
UDP	uridine diphosphate
ULoQ	upper limit of quantification
US EPA	United States Environmental Protection Agency
VTG	vitellogenin
WHO	World Health Organization
WoE	weight of evidence
XETA	*Xenopus* embryonic thyroid signalling assay

Glossary

Adverse effect	A change in the morphology, physiology, growth, development, reproduction or life span of an organism, system or (sub)population that results in an impairment of functional capacity, an impairment of the capacity to compensate for additional stress, or an increase in susceptibility to other influences (WHO/IPCS, 2009)
Adverse outcome pathway (AOP)	An AOP is an analytical construct that describes a sequential chain of causally linked events at different levels of biological organisation that lead to an adverse health or ecotoxicological effect
Analogy	A consistent observation across (related) substances having a well-defined MoA
Apical endpoint	An observable outcome in a whole organism, such as a clinical sign or pathological state that is indicative of a disease state that can result from exposure to a toxicant. As such, the apical endpoint is representing a measurable outcome responding to multiple different toxicity pathways/MoAs and can potentially be indicative of adverse effects
Biological plausibility	The biological plausibility relies on an understanding of the fundamental biological processes involved and whether they are consistent with the causal relationship being proposed. In the context of this guidance, the biological plausibility is considered to be the level of support for the link between the adverse effect and the endocrine activity. In addition, in the context of the MoA/AOP frameworks, biological plausibility is one of the elements to be considered in the weight of evidence analysis based on the evolved Bradford Hill considerations, where reference is made to the biological plausibility of the key event relationships

Biomarker	A biological parameter that is objectively measured and evaluated as an indicator of normal biological state or pathological processes
Coherence	Extent to which a hypothesised causal association is compatible with pre-existing theory and knowledge. Coherence analysis is part of the weight of evidence and is used to strengthen the predictive performance of adverse effects by considering: theoretical coherence (compatible with pre-existing theory), factual coherence (compatible with pre-existing knowledge), biological coherence (compatible with current biological knowledge or other levels of biological organisation) and statistical coherence (compatible with a reasonable statistical model, e.g. dose response)
Consistency	In this guidance, consistency is the pattern of effects across species/strains/organs/test systems that are expected based on the postulated MoA/AOP. In developing a MoA, consistency also refers to the repeatability of the KEs in the postulated MoA in different studies. Consistent observation of the same KE(s) in a number of studies with different study designs increases the support
Dose and incidence concordance	Dose concordance and incidence concordance are elements necessary for the evaluation of the empirical support. In a MoA/AOP context, dose and incidence concordance are verified when the key events are observed at doses or incidences below or similar to those associated with the adverse effect (or key events downstream)
Dose-response relationship	The dose–response relationship describes the change (in nature, incidence, magnitude and/or severity) in an effect on an organism caused by different levels of exposure (or doses) to a stressor (usually a chemical) after certain exposure duration. This definition includes the following assumptions: the response observed is due to the chemical administered, the magnitude of the response is in fact related to the dose and the observed effect is quantifiable
'EATS-mediated' (parameters)	Parameters measured *in vivo* that may contribute to the evaluation of adversity, while at the same time (due to the nature of the effect and the existing knowledge as described in OECD GD 150) they are also considered indicative of an EATS MoA and thus (in the absence of other explanations) also imply underlying *in vivo* mechanistic information. This group includes the parameters mainly from OECD CF level 4 and 5 tests labelled in OECD GD 150 as 'endpoints for estrogen-mediated activity', 'endpoints for androgen-mediated activity', 'endpoints for thyroid-related activity' and/or 'endpoints for steroidogenesis-related activity'
EATS-related adversity	Adversity identified on the basis of 'EATS-mediated' and/or 'sensitive to, but not diagnostic of, EATS' parameters
ED criteria	The criteria are legally defined in Commission Delegated Regulation (EU) No 2017/2100 and Commission Regulation (EU) No 2018/605 for biocidal products and plant protection products, respectively. They are based on the 2002 WHO/IPCS definition of an endocrine disruptor. They ask for consideration, in a weight of evidence approach, of all relevant scientific information including human and/or animal evidence, therefore allowing for the identification of both known and presumed endocrine-disrupting substances. The present guidance is written in accordance with these criteria
Empirical evidence	The information that can be acquired by observation or experimentation
Empirical support	Beside biological plausibility and essentiality, empirical support constitutes a third aspect of considerations for systematic assessment of confidence in a given MoA/AOP and involves dose, temporal, and incidence concordance
Endocrine activity	Interaction with the endocrine system that can potentially result in a response of the endocrine system, target organs and tissues. A substance that has an endocrine activity it has the potential to alter the function(s) of the endocrine system
Endocrine disruptor	An exogenous substance or mixture that alters function(s) of the endocrine system and consequently causes adverse health effects in an intact organism, or its progeny, or (sub)populations (WHO/IPCS, 2002)
Endocrine modality	A modality is an axis, pathway, signalling process, in this case within the endocrine system

Endocrine system	The endocrine system is a highly integrated and widely distributed group of organs that orchestrates a state of metabolic equilibrium or homoeostasis, among the various organs of the body. In endocrine signalling, molecules, i.e. hormones, act on target cells that are separate from their site of synthesis
Essentiality	Essentiality is one of the elements to be considered when performing the weight of evidence analysis using the evolved Bradford Hill considerations. In the context of the MoA/AOP frameworks, essentiality refers to key events. For determining essentiality it should be demonstrated whether or not downstream KEs and/or the adverse effect is prevented/decreased if an upstream event is experimentally blocked. It is generally assessed, on the basis of direct experimental evidence of the absence/reduction of downstream KEs when an upstream KE is blocked or diminished (e.g., in null animal models or reversibility studies)
Hormone	Substances which are produced by endocrine glands and secreted into the circulation, and which exert a regulatory effect elsewhere in the body
Human relevance	The extent to which certain results can be applied to humans for a given purpose (here: the identification of an endocrine-disrupting property)
Key event	A change in biological or physiological state that is both measurable and essential to the progression of a defined biological perturbation leading to a specific adverse outcome
Key event relationship	A scientifically based relationship that connects two key events, defines a directed relationship between the two (i.e., identifies one as upstream and the other as downstream), and facilitates inference or extrapolation of the state of the downstream key event from the known, measured, or predicted state of the upstream key event
Line(s) of evidence	A set of relevant information of similar type grouped to assess a hypothesis. There is no fixed rule on how much similarity of the information is required within the same line of evidence. This is for the assessor(s) to decide, and depends on what they find useful for the purpose of the scientific assessment
Mechanism of action	A detailed molecular description of the mechanistic interaction through which a substance/molecule produces its effect
Mode of action (MoA)	A biologically plausible sequence of key events at different levels of biological organisation, starting with the exposure to a chemical and leading to an observed (adverse) effect
Molecular initiating event (MIE)	A specialised type of key event that represents the initial point of chemical interaction on molecular level within the organism that results in a perturbation that starts the adverse outcome pathway
Population relevance	The extent to which an effect (e.g. elicited by a substance) can alter the sustainable performance and development of populations of non-target organisms
Postulated MoA	A postulated MoA is conceptualised as a single sequence of events proceeding from exposure to a given chemical, postulated MIE to the observed adverse effect via a series of postulated intermediate KEs which are not yet qualitative or quantitatively characterised in terms of biological plausibility and empirical support for the KER and essentiality of the KEs
Relevance	Refers to the appropriateness of the data for the intended purpose of the assessment
Reliability	Evaluates the inherent quality of a test report or publication relating to preferably standardised methodology and the way the experimental procedure and results are described to give evidence of the clarity and plausibility of the findings
'Sensitive to, but not diagnostic of, EATS' (parameters)	Parameters measured *in vivo* that may contribute to the evaluation of adversity, however, due to the nature of the effect and the existing knowledge as described in OECD GD 150, these effects cannot be considered diagnostic on their own of any one of the EATS modalities. Nevertheless, in the absence of more diagnostic parameters, these effects might provide indications of an endocrine MoA that might warrant further investigation. This includes parameters from OECD CF level 3, 4 and 5 *in vivo* assays and labelled in OECD GD 150 as endpoints potentially 'sensitive to, but not diagnostic of, EATS modalities'

Specificity	In this guidance specificity should be understood as the extent to which the MoA for the adverse effect is likely to be endocrine-related, i.e. whether an adverse effect is a consequence of the hypothesised endocrine MoA, and not a result of other non-endocrine mediated toxicity, including excessive systemic toxicity
Substance	In this guidance 'substance' is defined scientifically and refers to any chemical substance. For the respective regulatory context, refer to Regulation (EC) No 1107/2009 and Commission Regulation (EU) No 2018/605 for plant protection products, and to Regulation (EU) No 528/2012 and Commission Delegated Regulation (EU) No 2017/2100 for biocidal products
Systematic review	A systematic review is an overview of existing evidence pertinent to a clearly formulated question, which uses pre-specified and standardised methods to identify and critically appraise relevant research, and to collect report and analyse data from the studies that are included in the review
Temporal concordance/ temporality	Temporality is one of the elements necessary for the evaluation of the empirical observations. Are key events, within the MoA, observed in the hypothesised order?
Uncertainty	Uncertainty refers to all types of limitations in the knowledge available to assessors at the time an assessment is conducted and within the time and resources agreed for the assessment
Weight of evidence (WoE)	Weight of Evidence can be generally described as a stepwise process/ approach of collecting and weighing evidence to reach a conclusion on a particular problem formulation with (pre)defined degree of confidence

Appendix A – Additional considerations on how to assess the potential for thyroid disruption for human health

Abbreviations

Triiodothyronine (T3); thyroxine (T4); thyroid hormone (TH); thyroid-stimulating hormone (TSH); thyrotropin-releasing hormone (TRH); hypothalamic–pituitary–thyroid axis (HPT axis); developmental neurotoxicity (DNT); uridine diphosphate (UDP).

Background

The thyroid gland and its associated hormones are involved in metabolism, growth and development in all taxonomic groups. Because of the highly conserved nature of TH physiology, environmental factors affecting thyroid function or TH signalling in one species may well similarly affect others, including humans. The primary function of the thyroid is production of the iodine-containing hormones triiodothyronine (T3) and thyroxine (T4). The production of thyroid hormones (THs) is primarily regulated by thyroid-stimulating hormone (TSH) released from the anterior pituitary gland. TSH release is in turn stimulated by the thyrotropin-releasing hormone (TRH) from the hypothalamus. The THs provide negative feedback to TSH and TRH: when the THs are high, TSH production is suppressed. Feedback mechanisms are also in place for the regulation of TRH production (Joseph-Bravo et al., 2016).

The hypothalamic–pituitary–thyroid axis (HPT axis) is highly conserved across evolution in vertebrates. The regulation of serum TH levels and of TH action in various tissues involves a complex interplay of physiological processes. The thyroid function depends on iodine uptake, TH synthesis and storage in the thyroid gland, stimulated release of hormone into and transport through the circulation, hypothalamic and pituitary control of TH synthesis, cellular TH transport, tissue-specific TH de-iodination and degradation of THs by catabolic hepatic enzymes. All these processes can be affected by environmental factors that can adversely affect the thyroid function.

There are notable differences in the systemic regulation of TH levels between commonly used experimental animal models and humans. Although the HPT axis and the basic physiological processes regulating TH synthesis and release are qualitatively similar across species, there are, however, quantitative species-specific differences (Janssen and Janssen, 2017). All these aspects are making the relationship between changes in circulating THs, including the ones mediated by differences in metabolism, and downstream adverse effects very complex and additional elements, such as for example: species specific metabolic capacity and age specific differences in sensitivity, have to be taken into consideration. Therefore, species differences in the sensitivity of specific developmental outcomes as a result of substance-induced changes of circulating levels of THs cannot be ruled out at this time. Similarly, the assumption that thyroid effects observed in rat are not in many cases human relevant can be substantiated using, for instance, evidence of species specific differences in metabolic capacity, and based on weight of evidence.

Therefore, this appendix is intended to provide additional guidance on which data could be provided and considered in the weight of evidence to substantiate that some specific thyroid effects are not human relevant and how to address specific thyroid related DNT concerns. This appendix is not intended to be exhaustive and covers all MoAs associated to thyroid effects for which the principles detailed in this guidance should be applied.

Using the current understanding of thyroid physiology and toxicology (European Commission, 2017), it is proposed that the following be applied when interpreting data from experimental animals:

1) Substances inducing histopathological changes (i.e. follicular cell hypertrophy and/or hyperplasia and/or neoplasia) in the thyroid, with or without changes in the circulating levels of THs, would pose a hazard for human thyroid hormone insufficiency in adults as well as pre- and post-natal neurological development of offspring.
2) Substances that alter the circulating levels of T3 and/or T4 without histopathological findings would still present a potential concern for neurodevelopment.
3) In the absence of substance-specific data which provide proof of the contrary, humans and rodents are considered to be equally sensitive to thyroid-disruption (including cases where liver enzyme induction is responsible for increased TH clearance).

In case an applicant considers generating additional data in order to investigate human relevance of the effect observed in rat, the following paragraphs can give more specific guidance on the mode of action of the thyroid-disruption and on the weight of evidence for human relevance.

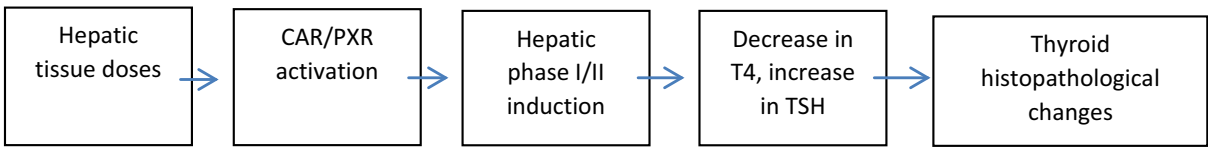
Investigation of increase in thyroid hormone metabolism in the liver

In cases where changes in thyroid follicular cell histopathology, with or without changes in THs, are observed in tested animal species, human relevance of such effects could be further investigated (Boobis et al., 2006). One possible explanation for the changes in TH levels or thyroid histopathology is that the substance causes induction of certain metabolic enzymes in the liver resulting in increased clearance of T4. The induction of T4-uridine diphosphate [UDP]-glucuronyl transferase is suggestive of increased clearance of THs with concomitant reduction in circulating T4, this will result in an increase of TSH that, in turn, would stimulate thyroid growth manifested by follicular cell hypertrophy/hyperplasia/neoplasia (Curran and DeGroot, 1991; Capen, 1997; Ennulat et al., 2010).

To investigate whether liver enzyme induction is responsible for the effects seen on TH levels and/or thyroid histopathology and weight, as well as whether the effect is or not likely to be human relevant, the following three pieces of information are needed:

1) Results of analysis of serum/plasma samples (if available) for TSH, T3 and T4 in the existing repeated dose toxicity studies. If unavailable, a specifically designed *in vivo* toxicity study should be considered. In this study, TSH, T3 and T4 should be measured and, where possible, additional data on liver enzyme induction (e.g. measurement of UDPGT) should be included.
2) Comparative studies of enzyme activity induced by the test substance in liver *in vitro* systems should be measured in both the relevant test species (e.g. rat, mouse and dog) and humans. The metabolism of the specific substance (ADME properties) in both test species and humans, and the activity of possible metabolites must be considered when this comparison is conducted.
3) The presence of other possible thyroid-disrupting modes of action such as interference with TH synthesis should also be excluded, e.g. by evaluating *in vitro* the potential for inhibition of the sodium–iodide symporter (NIS) (Cianchetta et al., 2010; Hallinger et al., 2017; Kogai and Brent, 2012) and thyroid peroxidase (TPO) (Kambe and Seo, 1997; Paul et al., 2014; Paul Friedman et al., 2016; Wu et al., 2016). It must, however, be acknowledged that substances may interfere with the thyroid hormone system through many different mechanisms of action, and that currently validated/standardised *in vitro* assays do not exist to investigate all these different pathways and a reasonable effort is anticipated, based on available tools and current understanding of thyroid physiology.

An example of a postulated mode of action is reported below:



The assessment of qualitative/quantitative differences in hepatic induction can therefore be part of the WoE and used to provide evidence of non-human relevance.

Investigations of perturbations of circulating thyroid hormone in the absence of histological changes in adults

A decrease in T4 (total or free) in the absence of adverse histological changes should act as a trigger for further studies. It is known from the broad knowledge of biology (e.g. human clinical experience and epidemiological data) that a drop in T4 results in impaired pre- and postnatal-neurological development (Alshehri et al., 2015). Therefore, the hazard assessment of a substance should consider the most sensitive population and reductions in T4 levels should act as a trigger for further studies of F1 generation (e.g. as part of most updated OECD test guidelines 421/422, 426, 416, 443) (OECD, 2001, 2007, 2012, 2016a,b) depending on the other information available. However, since in this case, disruption of thyroid homeostasis is the critical effect that may lead to adverse

effects on the developing nervous system, a special study developed by the US EPA to investigate critical periods of development (i.e. in pregnant females, the fetus and newborn) could be conducted in place of the rat DNT study to generate mechanistic data to confirm or refute the observed change in circulating TH (US EPA, 2005). This study is intended to generate specific data on the thyroid to establish the ability of a chemical to disrupt thyroid function in pregnant females and in the fetus and newborn. This special study is therefore expected to be conducted based on the results of a study(ies) in adult animals that provide evidence that a substance produces effects on thyroid function.

Further investigations of thyroid disruption

An in-depth understanding of the fundamental principles that regulate TH homoeostasis is critical for hazard identification of substances which alter thyroid homeostasis. The hazard identification is currently hampered by a lack of internationally validated test methods. To appropriately investigate thyroid concerns, existing test protocols need to be modified. When considering such modifications the recommendations on how to investigate thyroid effects in rodent models from the American Thyroid Association should be considered (Bianco et al., 2014).

References

Alshehri B, D'Souza DG, Lee JY, Petratos S and Richardson SJ, 2015. The diversity of mechanisms influenced by transthyretin in neurobiology: development, disease and endocrine disruption. Journal of Neuroendocrinology 27, 303–323. https://doi.org/10.1111/jne.12271

Bianco AC, Anderson G, Forrest D, Galton VA, Gereben B, Kim BW, Kopp PA, Liao XH, Obregon MJ, Peeters RP, Refetoff S, Sharlin DS, Simonides WS, Weiss RE, Williams GR and American Thyroid Association Task Force on Strategies to Investigate Thyroid Hormone Economy and Action, 2014. American Thyroid Association Guide to investigating thyroid hormone economy and action in rodent and cell models. Thyroid, 24, 88–168. https://doi.org/10.1089/thy.2013.0109

Boobis AR, Cohen SM, Dellarco V, McGregor D, Meek ME, Vickers C, Willcocks D and Farland W, 2006. IPCS framework for analyzing the relevance of a cancer mode of action for humans. Critical Reviews in Toxicology, 36, 781–792. https://doi.org/10.1080/10408440600977677

Capen CC, 1997. Mechanistic data and risk assessment of selected toxic end points of the thyroid gland. Toxicologic Pathology, 25, 39–48. https://doi.org/10.1177/019262339702500109

Cianchetta S, di Bernardo J, Romeo G and Rhoden, KJ, 2010. Perchlorate transport and inhibition of the sodium iodide symporter measured with the yellow fluorescent protein variant YFP-H148Q/I152L. Toxicology and Applied Pharmacology, 243, 372–380. https://doi.org/10.1016/j.taap.2009.12.004 7

Curran PG and DeGroot LJ, 1991. The effect of hepatic enzyme-inducing drugs on thyroid hormones and the thyroid gland. Endocrine Reviews, 12, 135–150. https://doi.org/10.1210/edrv-12-2-135

European Commission (Directorate-General for the Environment), DTU National Food Institute Denmark; Brunel University London, 2017. Supporting the organisation of a workshop on thyroid disruption – Final Report (Framework Contract ENV.A.3/FRA/2014/0029 on implementation of the Community strategy on Endocrine Disrupters). In. Publications Office of the European Union, Luxembourgpp. Available online: https://doi.org/10.2779/921523

Ennulat D, Walker D, Clemo F, Magid-Slav M, Ledieu D, Graham M, Botts S and Boone L, 2010. Effects of hepatic drug-metabolizing enzyme induction on clinical pathology parameters in animals and man. Toxicologic Pathology, 38, 810–828. https://doi.org/10.1177/0192623310374332

Hallinger DR, Murr AS, Buckalew AR, Simmons SO, Stoker TE and Laws SC, 2017. Development of a screening approach to detect thyroid disrupting chemicals that inhibit the human sodium iodide symporter (NIS). Toxicology *In Vitro,* 40, 66–78. https://doi.org/10.1016/j.tiv.2016.12.006

Janssen ST and Janssen OE, 2017. Directional thyroid hormone distribution via the blood stream to target sites. Molecular and Cellular Endocrinology, 458, 16–21. https://doi.org/10.1016/j.mce.2017.02.037

Joseph-Bravo P, Jaimes-Hoy L and Charli JL, 2016. Advances in TRH signaling. Reviews in Endocrine and Metabolic Disorders, 17, 545–558. https://doi.org/10.1007/s11154-016-9375-y

Kambe F and Seo H, 1997. Thyroid-specific transcription factors. Endocrine Journal, 44, 775–784. https://www.ncbi.nlm.nih.gov/pubmed/9622292

Kogai T and Brent GA, 2012. The sodium iodide symporter (NIS): regulation and approaches to targeting for cancer therapeutics. Pharmacology and Therapeutics, 135, 355–370. https://doi.org/10.1016/j.pharmthera.2012.06.007

OECD (Organisation for Economic Co-operation and Development), 2001. Test No. 416: Two-Generation Reproduction Toxicity. In: OECD Guidelines for the Testing of Chemicals, Section 4. OECD Publishing, Paris. 13 pp. https://doi.org/10.1787/9789264070868-en

OECD (Organisation for Economic Co-operation and Development), 2007. Test No. 426: Developmental Neurotoxicity Study. In: OECD Guidelines for the Testing of Chemicals, Section 4. OECD Publishing, Paris. 26 pp. https://doi.org/10.1787/9789264067394-en

OECD (Organisation for Economic Co-operation and Development), 2012. Test No. 443: Extended One-Generation Reproductive Toxicity Study. In: OECD Guidelines for the Testing of Chemicals, Section 4. OECD Publishing, Paris. 25 pp. https://doi.org/10.1787/9789264185371-en

OECD (Organisation for Economic Co-operation and Development), 2016a. Test No. 421: Reproduction/ Developmental Toxicity Screening Test. In: OECD Guidelines for the Testing of Chemicals, Section 4. OECD Publishing, Paris. 10 pp. https://doi.org/10.1787/9789264264380-en

OECD (Organisation for Economic Co-operation and Development), 2016b. Test No. 422: Combined Repeated Dose Toxicity Study with the Reproduction/Developmental Toxicity Screening Test. In: OECD Guidelines for the Testing of Chemicals, Section 4. OECD Publishing, Paris. 14 pp. https://doi.org/10.1787/9789264264403-en

Paul KB, Hedge JM, Rotroff DM, Hornung MW, Crofton KM and Simmons SO, 2014. Development of a thyroperoxidase inhibition assay for high-throughput screening. Chemical Reserach in Toxicology, 27, 387–399. https://doi.org/10.1021/tx400310w

Paul Friedman K, Papineni S, Marty MS, Yi KD, Goetz AK, Rasoulpour RJ, Kwiatkowski P, Wolf DC, Blacker AM and Peffer RC, 2016. A predictive data-driven framework for endocrine prioritization: a triazole fungicide case study. Critical Reviews in Toxicology, 46, 785–833. https://doi.org/10.1080/10408444.2016.1193722

US EPA (United States Environmental Protection Agency), 2005. Guidance for Thyroid Assays in Pregnant Animals, Fetuses and Postnatal Animals, and Adult Animals. In. US EPA, Office of Pesticide Programs, Health Effects Division, Washington (DC). 12 pp. Available online: https://www.epa.gov/sites/production/files/2015-06/documents/thyroid_guidance_assay.pdf

Wu Y, Beland FA and Fang JL, 2016. Effect of triclosan, triclocarban, 2,2′,4,4′-tetrabromodiphenyl ether, and bisphenol A on the iodide uptake, thyroid peroxidase activity, and expression of genes involved in thyroid hormone synthesis. Toxicological In Vitro, 32, 310–319. https://doi.org/10.1016/j.tiv.2016.01.014

Appendix B – Recommendations for design, conduction and technical evaluation of hormonal studies

Abbreviations

Follicle-stimulating hormone (FSH); luteinising hormone (LH); triiodothyronine (T3); thyroxine (T4); thyroid-stimulating hormone (TSH); post-natal day (PND); radioimmunoassay (RIA), enzyme-linked immunosorbent assay (ELISA)

Background

Hormonal studies are generally initiated to investigate the endocrine functions following administration of a substance. They can be incorporated in the planned toxicological studies or evaluated in separate investigative studies. The purpose is to compare base-line conditions (e.g. hormonal level in the control group) with changes after stimulation or inhibition of the hormonal pathway as a consequence of the administration of the test substance.

The hormonal investigation is generally applied for the detection of effects related to previous indication from animal studies performed with the substance. Reasons for concern are in most instances related to the reproductive system, the adrenal system or the thyroid gland. Concern may be caused by histopathological changes (e.g. in gonads, adrenals, and thyroid), organ weight changes or findings in clinical chemistry. If a concern is identified before the initiation of a toxicological study, a targeted investigation can be included in the standard toxicology protocol, (adding a satellite group if necessary) or specific mechanistic studies may be initiated.

Repeated administration (at least 7 days) is generally required to reach a steady state for the response and adaptation of hormone dependent organs (Sandow, 2006). At least two doses are necessary for a sufficient effect size and to achieve a biologically relevant (and statistically significant), difference between treated groups and control group. Although the inclusion of a vehicle treated group is mandatory, the additional inclusion of a positive control is not necessary for routine studies because enough information exist about the effect size of established chemicals that affect the endocrine system.

It is anticipated that circulating levels of hormones will be frequently determined as part of the toxicological evaluation for active substances in plant protection and biocidal products to support the evaluation of endocrine activities. There is guidance available in the medical field to support, e.g. the conduct and interpretation of thyroid hormone measurements. However, for toxicological purposes, specific recommendations are needed (Bianco et al., 2014). A number of factors (e.g. stress, circadian rhythm, and oestrous cycle) may have an impact on hormone concentrations and on study results and, as such, they are very important factors to be considered during the investigation and during the assessment of the results. The intention of this Appendix is to formulate a list of practical recommendations for applicants and assessors concerning methods for measuring hormones to evaluate the potential for endocrine activity (Chapin and Creasy, 2012; Stanislaus et al., 2012; Andersson et al., 2013; FDA, 2015).

Material below is subdivided into recommendations for thyroid hormones and reproductive hormones. Non-EATS pathways are outside the scope of this Annex. It should also be mentioned that the current recommendations represent current best practice and are not prescriptive. However, the recommendations were prepared with the intention of standardising the conditions under which hormonal assays are conducted, addressing the issues of high biological and potential analytical variability. Bearing in mind that a variety of the methodologies have been developed and have often been validated in the test laboratories, the recommendations are not prescriptive and are formulated mainly to indicate which methods should be avoided as these may have a significant effect on the measurements.

1) Recommendations for thyroid hormone analysis

Thyroid hormones are routinely measured in laboratories conducting toxicological studies, thus ensuring a significant body of expertise and knowledge. Consequently, a detailed list of recommendations on methodologies for the measurement of thyroid hormones was formulated and is presented below.

Hormones. All three thyroid hormones, i.e. T3, T4 and TSH should be measured. Measurement of a single hormone on its own, e.g. T4, without complementary parameters such as TSH, thyroid weight,

histopathology of thyroid and pituitary, should not be used to draw conclusion regarding changes in the hypothalamus-pituitary-thyroid axis, but raises a concern for effects on the thyroid hormone system, which needs to be clarified.

Free or bound fraction to be measured. A high volume of serum (~ 200 μL) is required for measurement of the free fraction, possibly compromising the feasibility of this assay in routine studies or studies in pups. Free hormone can be measured however in specifically designed mechanistic studies on a case-by-case basis. To measure accurately free hormone levels, the sample should be pretreated, e.g. ultracentrifugation or dialysis. Chromatography or equally sensitive techniques (e.g. radioimmunoassay (RIA) or enzyme-linked immunosorbent assay (ELISA)) should be applied for detection of free hormone.

Species. The current recommendations are applicable for measurements in rats. Other species (e.g. dog) can be used as well, but the assay needs to be adjusted to the specific conditions for the species in question.

Age. T4 and T3 can be measured starting from post-natal day (PND) 4, at weaning age and in post-pubertal animals. The measurement of the thyroid hormones in fetuses are not required currently in the EU, however, should this become necessary, the addition of a satellite group should be considered to avoid interference of the hormonal assay with other examinations of the fetuses. Pooling of blood for thyroid hormone analysis could be necessary for fetal samples within a litter in order to obtain enough material to run the assay.

Sex. Both sexes can be used for measurement of thyroid hormones. Synchronisation of females is not a pre-requisite for thyroid hormonal assay. No sex difference regarding the serum thyroid hormone levels exists in fetuses, and pups between PND 4 and PND 21.

Number of animals. Eight to ten animals per group are in general enough to ensure sufficient statistical power of the study. As a lower number of animals is recommended under certain circumstances (e.g. OECD TG 407 (OECD, 2008), n = 5 per sex), power analysis can be used to calculate the minimum effect size that is likely to be identified in this study type. The following is an example showing the percentage of thyroid hormone change differences which are assumed to be detected (Wilcoxon test, two-sided, power 75%, $p < 0.05$) dependent on the group sample sizes per sex (see Table A.1).

Table A.1: Thyroid hormone changes presumed to be detected considering variation and animal number

Rays per group and sex	5	6	8	10	15	20	25
% Decrease at a CV of 25%	−73.4	−54.7	−41.6	−35.2	−27.1	−22.8	−20.1
% Increase at a CV of 35%	102.7	76.5	58.2	49.2	37.9	31.9	28.1

Wilcoxon test, two-sided (power 75%; $p < 0.05$). CV: coefficient of variation.

Animal care. Animal care and housing should fulfil the requirements according to current EU legislation (Directive 2010/63/EU on the protection of animals used for scientific purposes[15]). Recommended practise of group housing of animals, when 2-5 rats are kept in one cage of suitable size has no impact on thyroid hormone measurements.

Consideration on hormonal physiology and circadian rhythm. Samples assigned for thyroid hormonal assay should be collected between 8 a.m. and noon (when considering a standard/regular 12:12 h light/dark cycle). All of the samples of one study should be taken in the shortest possible time (not more than 2 h). Animals' stratification and randomisation is mandatory for sampling. For practical reasons and considering the restriction in time, staggering of animals for terminal sampling might be necessary (e.g. by parturition staggering). However, the same number of animals from the control and the treated groups should be sampled on one day and all groups should be represented to the extent possible (stratification).

Anaesthesia. For adult rats, the use of isoflurane is recommended as a suitable and relatively fast method of anaesthesia, while CO_2 should be avoided for animal welfare reasons and due to interference with the concentrations of the thyroid hormones in exposed animals.

[15] Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the protection of animals used for scientific purposes. OJ L 276, 20.10.2010, p. 33–79. Available online: http://data.europa.eu/eli/dir/2010/63/oj

Blood sampling. The maximum amount of collected blood should be in accordance with the EU and national animal welfare regulations. To reduce the level of stress associated with the technical procedure, blood sampling should be executed by a trained technician and should not exceed the time of 3 min per animal under anaesthesia and 1 min per animal if not under anaesthesia. For in-life sampling, a separate room may be used where possible. If animals are moved to a new location, animals should be given at least 30 min to acclimatise. Extended acclimatisation for up to 24 h is not necessary. Specific considerations should be made for the acclimatisation time when dealing with hormonal investigations for the HPA axis (Balcombe et al., 2004).

- ***In adults,*** restraint during tail vein sampling might stress the animal and should thus be avoided. For animal welfare reasons, cardiac puncture for in-life sampling in adult animals should be avoided. If the method requires preparatory procedures (e.g. shaving for jugular vein sampling), these should be performed one day prior to sampling.
- ***In pups,*** decapitation followed by trunk blood collection or cardiac puncture are the methods of choice.
- ***For fetuses,*** decapitation or sampling from umbilical cord blood are the methods of choice.

Euthanasia. Usage of ether should be avoided.

- ***For adults,*** irreversible isoflurane anaesthesia followed by exsanguination is recommended, while the use of Isoflurane alone should be avoided. Decapitation or exsanguination without prior anaesthesia contradicts the EU legislation.
- ***For pups,*** the same recommendations as for adults apply.

Sample collection. Whole blood can be collected in serum separation tubes and left to clot for at least 30 min at room temperature. When plasma is used for further sample processing, sodium-citrate-treated tubes should be avoided, while heparin- and EDTA-treated tubes can be used, following validation of sample stability.

Sample storage. Upon collection of blood and separation from the matrix (e.g. plasma or serum), samples can be divided in different aliquots and stored until further processing and analysis. However, sample storage conditions (e.g. temperature, length, freeze-thaw stability) must be validated.

Quantitation methods. All methods might be suitable, but quality criteria need to be defined. If free hormone is measured, pretreatment of samples should be performed (e.g. ultracentrifugation or dialysis) and the measurements should be performed using chromatography or an equally sensitive technique. Validation of quantitation methods should be performed for each species.

Assay validation. Considering that different assays have already been established by laboratories and that restricting detection methods to a certain range might hinder future development of the technologies, for the scope of this guidance document it is necessary to ensure that certain quality criteria are met, specifically:

a) The lower limit of quantification (LLoQ) and the upper limit of quantification (ULoQ) should be established.
b) Reproducibility of the assay should be assessed and the coefficients of the inter- and intra-assay variation should be calculated and they should be in line with the limits established for the particular commercial kit.
c) In untreated control animals, the criteria for coefficient of variation (CV) for T3 and T4 measurements (< 25%), as stated in OECD TG 407 (OECD, 2008), should be met. If %CV exceeds the recommended level (in isolated cases), an explanation of the events should be provided otherwise the study validity might be questioned.
d) Repeatability of the assay within a day or across several days should be proven.
e) The type of applied quality control samples (e.g. spiked samples, biological control samples, reference range, etc.) should be recorded. A serum dilution curve should also be run to show that the assay is valid for the serum samples under investigation.
f) The performance of the assay with a particular matrix (serum or plasma) should be assessed.
g) A validation study, conducted with a positive control (reference compound) should be available to establish the laboratory's proficiency in performing the assay. Different dose levels should be used for the positive control, and it is crucial to choose an ad hoc positive control.
h) Stability of the sample under selected storage conditions should be validated.
i) Validation of the assay should be carried out for each species separately.

j) If the measurements of the free fraction of T3 and T4 are conducted in mechanistic studies, pretreatment of samples is required, followed by chromatographic/immunoassay detection of the non-bound fractions of the hormones.
k) Cross-reactivity of antibodies used in the assay should be established at least at the level of the kit manufacturer.
l) If possible, lot-to-lot variation of reagents (e.g. antibodies) should be assessed.

All of the above-mentioned criteria should be included in the method validation report and should be accessible to the assessors.

Use of historical control data. Under normal circumstances, historical control data are not required for the evaluation of the results and the effect should be detected by comparing to values in the concomitant control group. Historical controls should be consulted only as a qualitative measure of the assay reliability. If the historical control data are consulted, it should be demonstrated that the same assay methodology (including sampling time) was used; that the assay was conducted for animals of the same strain and age groups and kept under standardised housing/dietary/environmental conditions. Furthermore, the period between the historical control sampling and the evaluated study must be considered carefully since over time, parameters may change in a given population of animals.

Statistical analysis of data. No specific statistical analysis methodology is recommended when data on circulating thyroid hormones concentrations are analysed. High variability should trigger outlier statistics and justification for each excluded data point should be provided.

2) Recommendations for reproductive hormones analysis

Hormones. Measurement of oestradiol, testosterone and other hormones (e.g. luteinising hormone (LH), follicle-stimulating hormone (FSH), progesterone) may provide an important contribution to the identification of endocrine activities; however, assessment of a panel of hormones (e.g. FSH, LH and Prolactin) is preferable to the measurement of a single hormone. Where possible, selection of the hormones to be measured in a study should be based on information gathered in previous toxicological tests. Recommendations described below are equally applicable to oestradiol, testosterone, LH, FSH, progesterone. The same general considerations applied for the thyroid hormones are applicable for the sex hormones and will be not repeated here. Recommendations listed below should be considered as additional considerations for sex hormones.

Sex. Study design should address differences between males and females. Information from both sexes may be useful for assessing reproductive hormones, depending on the indications gathered in previous studies. When hormones are measured in female animals, synchronisation is not a necessity, however, stage of the oestrous cycle at the time of blood collection should be considered.

Number of animals. Statistical power analysis should be performed to establish either group size, or if the group size is defined by the test guidelines, to establish the effect size that can be determined using given number of animals. A higher number of females might be needed due to differences in the oestrous cycle.

Consideration of effects of circadian rhythm. Blood sampling should be accomplished in a 3-h time window in the morning if samples are to be processed for the sex hormone measurement. Stratification of animals from treated and control groups is necessary to control for differences in timing of blood collection. Considering the restrictions imposed by a relatively short time-window, sampling (e.g. terminal sampling) can be done on different days; however the groups should be stratified, so that all groups are represented to the extent possible. For stratification and randomisation of females, the stage of oestrous cycle should be taken into consideration.

Blood sampling. To reduce stress, blood sampling should be performed by a trained technician and should not exceed 3 min. Any method of blood sampling that is approved in the laboratory and that would guarantee the lowest possible stress level can be used. The maximum amount of collected blood should be in accordance with the EU and national animal welfare regulations. Thus, if several hormones are intended to be analysed and the amount of blood/serum is not sufficient, pooling of samples collected from one group/sex can be considered.

Sample collection. Whole blood can be processed to serum or plasma, depending on the protocol established in the laboratory.

Sample storage. Upon blood collection and separation of matrix (e.g. plasma or serum), samples can be aliquoted and stored frozen until further processing. Care should be taken, to reduce the time a sample is kept at room temperature to a minimum. Chosen storage conditions should guarantee sample stability.

References

Andersson H, Rehm S, Stanislaus D and Wood CE, 2013. Scientific and regulatory policy committee (SRPC) paper: assessment of circulating hormones in nonclinical toxicity studies III. Female reproductive hormones. Toxicologic Pathology, 41, 921–934. https://doi.org/10.1177/0192623312466959

Balcombe JP, Barnard ND and Sandusky C, 2004. Laboratory routines cause animal stress. Contemporary Topics in Laboratory Animal Science, 43, 42–51.

Bianco AC, Anderson G, Forrest D, Galton VA, Gereben B, Kim BW, Kopp PA, Liao XH, Obregon MJ, Peeters RP, Refetoff S, Sharlin DS, Simonides WS, Weiss RE, Williams GR, American Thyroid Association Task Force on Strategies to Investigate Thyroid Hormone Economy and Action, 2014. American Thyroid Association Guide to investigating thyroid hormone economy and action in rodent and cell models. Thyroid, 24, 88–168. https://doi.org/10.1089/thy.2013.0109

Chapin RE and Creasy DM, 2012. Assessment of circulating hormones in regulatory toxicity studies II. Male reproductive hormones. Toxicologic Pathology, 40, 1063–1078. https://doi.org/10.1177/0192623312443321

FDA (Food and Drug Administration), 2015. Nonclinical Evaluation of Endocrine-Related Drug Toxicity In: Guidance for Industry. Office of Communications, Division of Drug Information, Center for Drug Evaluation and Research, FDA, Silver Spring (MD). 10 pp. Available online: https://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm369043.pdf

OECD (Organisation for Economic Co-operation and Development), 2008. Test No. 407: Repeated Dose 28-day Oral Toxicity Study in Rodents. In: OECD Guidelines for the Testing of Chemicals, Section 4. OECD Publishing, Paris. 13 pp. https://doi.org/10.1787/9789264070684-en

Sandow J, 2006. Assays in Endocrine Safety Pharmacology. In: Edition S (ed.) Drug Discovery and Evaluation: Safety and Pharmacokinetic Assays. Heidelberg, New York: Springer-Verlag Berlin.

Stanislaus D, Andersson H, Chapin R, Creasy D, Ferguson D, Gilbert M, Rosol TJ, Boyce RW and Wood, CE, 2012. Society of toxicologic pathology position paper: review series: assessment of circulating hormones in nonclinical toxicity studies: general concepts and considerations. Toxicologic Pathology, 40, 943–50. https://doi.org/10.1177/0192623312444622

Appendix C – Information requirements for active substances under the Biocidal Products[1] and Plant Protection Products Regulations[2,8] which could potentially provide information on endocrine-disrupting properties

There are specific rules for adaptation from standard information requirements concerning some of the studies that may require recourse to testing vertebrates. These adaptations mostly refer to risk management related considerations, such as the absence of uses in which human exposure may occur, or certain substance properties, that from a risk management perspective would make the conduct of a study unnecessary (e.g. *'reproductive toxicity studies do not need to be carried out if a substance is known to have an adverse effect on fertility, meeting the criteria for classification as reproductive toxicity Cat. 1A or 1B […]'*). Assessment of whether a substance meets the ED criteria is, however, a hazard assessment, specifically of the ED hazardous properties of the substance. Therefore, where there is an option to waive a study pertaining to the mandatory information requirements (core data set) based on risk assessment or risk management considerations, it needs to be considered whether the study would still be necessary for ED hazard assessment, in order to establish a complete and adequate database for the ED assessment strategy set out in this guidance.

C.1. Toxicological data

	PPP	BP[(a)]
Toxicokinetics and metabolism studies in mammals (OECD TG 417)	Information requirement	Information requirement
Repeated dose toxicity		
Short-term repeated dose toxicity study (28 days; OECD TG 407), in rodents. Preferred species is rat	Available studies shall be reported	Available studies shall be reported
Subchronic repeated dose toxicity study (90 days; OECD TG 408), in rodents. Preferred species is rat	Information requirement	Information requirement
Subchronic repeated dose toxicity study (90 days; OECD TG 409), in a non-rodent species. Preferred species is dog	Information requirement	Further repeat dose studies are triggered
Long-term repeated dose toxicity (≥ 12 months; included in OECD TG 453; OECD TG 452), in a rodent species. Preferred species is rat	Information requirement[(b)]	Information requirement[(b)]
Further repeat dose studies	Triggered	Triggered
Reproductive toxicity		
Pre-natal developmental toxicity study (OECD TG 414) in a first species, rabbit is preferred	Information requirement	Information requirement
Pre-natal developmental toxicity study (OECD TG 414) in a second species, rat is preferred	Information requirement[(c)]	Triggered
Developmental neurotoxicity (OECD TG 426)	Triggered	Triggered
Two-generation reproductive toxicity study (OECD TG 416), in rats	Information requirement[(d)]	Information requirement[(d)]
Extended one-generation reproduction toxicity (OECD TG 443) including the second generation and neurotoxicity and immunotoxicity cohorts	See notes[(d),(e)]	See notes[(d),(e)]
Carcinogenicity		
Carcinogenicity testing in a first species (OECD TG 451), rat is the preferred species	Information requirement[(f)]	Information requirement[(f)]
Carcinogenicity testing in a second species (OECD TG 451), mouse is the preferred species	Information requirement[(f)]	Information requirement[(f)]
Endocrine-disrupting properties[(g)]		
H295R Steroidogenesis assay (OECD TG 456)	Triggered	Triggered
Stably transfected human estrogen receptor alpha transcriptional activation assay for detection of estrogenic agonist-activity of chemicals (OECD TG 455)	Triggered	Triggered
Uterotrophic assay (mechanistic *in vivo* tests) (OECD TG 440)	Triggered	Triggered

	PPP	BP[a]
Hershberger assay (mechanistic *in vivo* test) (OECD TG 441)	Triggered	Triggered
Peripubertal male and female assays (OPPTS 890.1500 and 890.1450)	Triggered	Triggered
15-day intact adult male rat assay (US EPA 2007)	Triggered	Triggered
Relevant human health data	Information requirement	Information requirement
Epidemiological studies on the general population	Information requirement	Information requirement
Literature data[h]	Information requirement	Information requirement in the ED criteria

(a): Note that in the information requirements of the Biocidal Products Regulation the terms 'core data set' and 'additional data set' are used for the studies that in the tables below (column BP) are referred to as, respectively, 'information requirement' and 'triggered'.
(b): A long-term repeated dose toxicity study (≥ 12 months) must not be undertaken if a combined long-term repeated dose/ carcinogenicity study (OECD TG 453) is submitted.
(c): The study should not be conducted if developmental toxicity has been adequately assessed as part of an extended one-generation reproductive toxicity study (OECD TG 443).
(d): An extended one-generation reproduction toxicity study (OECD TG 443) may be provided as an alternative to the two-generation reproductive toxicity study (OECD TG 416).
(e): The need to conduct further studies with regard to developmental immunotoxicity and neurotoxicity should be considered along with the extended one-generation reproduction toxicity study (OECD TG 443 and with the developmental neurotoxicity study (OECD TG 426).
(f): For a new active substance, the information requirements for carcinogenicity study and long-term repeated dose toxicity are combined with a combined chronic toxicity/carcinogenicity study (OECD TG 453).
(g): If there is any evidence from *in vitro*, repeat-dose or reproduction toxicity studies that the active substance may have endocrine-disrupting properties then additional information or specific studies will be required to:

- elucidate the mode/mechanism of action
- provide sufficient evidence for relevant adverse effects.

(h): A summary of all relevant data from the scientific peer-reviewed open literature on the active substance, metabolites and breakdown or reaction products and plant protection products containing the active substance should be submitted according to (EFSA, 2011).

C.2. Ecotoxicological data

	PPP	BP[a]
Effects on birds and other terrestrial vertebrates		
Sub-chronic and reproductive toxicity to birds (OECD TG 206)	Information requirement unless exposure of adults or exposure of nest sites during the breeding season is unlikely to occur	Triggered
Long-term and reproductive toxicity to mammals	Information requirement under the mammalian section	TriggeredIf needed, information is derived from mammalian data
Effects on terrestrial vertebrate wildlife (birds, mammals, reptiles and amphibians)	Available and relevant data, including data from the open literature regarding the potential effects on birds, mammals, reptiles and amphibians shall be presented and taken into account in the risk assessment	Effects on other non-target, non-aquatic organismsTriggered
Endocrine-disrupting properties	Consideration shall be given to whether the active substance is a potential endocrine disrupter according to European Union or internationally agreed guidelines. This may be done by consulting the mammalian toxicology section. In addition, other available information on toxicity profile and mode of action shall be taken into account. If, as a result of this assessment, the active substance is identified as a potential endocrine disruptor, the type and conditions of the study to be performed	Indication of endocrine activityTriggered

	PPP	BP[a]
	shall be discussed with the national competent authorities	
Effects on fish		
Long-term and chronic toxicity to fish		
Fish early life stage test (OECD TG 210)	Information required when exposure of surface water is likely and the substance is deemed to be stable in water (less than 90% loss of the original substance over 24 h via hydrolysis)	Triggered
Fish full life cycle test (OPPTS 850.1500)	Triggered if there is concern regarding ED properties identified in the screening testing battery or for which there are other indications of endocrine disruption (see point 8.2.3); for this purpose appropriate additional endpoints shall be included	Triggered
Endocrine-disrupting properties for aquatic organisms[b]		
Fish short-term reproduction assay (OECD TG 229)[c]	Screening test battery always required unless ED properties can be excluded based on information on toxicity profile and mode of action	Not an information requirement
21-day fish assay: a short-term screening for estrogenic and androgenic activity, and aromatase inhibition (OECD TG 230)	Screening test battery always required unless ED properties can be excluded based on information on toxicity profile and mode of action	Not an information requirement
Fish sexual development test (OECD TG 234)	Screening test battery always required unless ED properties can be excluded based on information on toxicity profile and mode of action	Not an information requirement
Amphibian metamorphosis assay (OECD TG 231)	Screening test battery always required unless ED properties can be excluded based on information on toxicity profile and mode of action	Not an information requirement
Literature data[d]	Information requirement	Information requirement in the ED criteria

(a): Note that in the information requirements of the Biocidal Products Regulation the terms 'core data set' and 'additional data set' are used for the studies that in the tables below (column BP) are referred to as, respectively 'information requirement' and 'triggered'.

(b): Consideration should be given to whether the active substance is a potential endocrine disruptor in aquatic non-target organisms according to European Union or internationally agreed guidelines. In addition, other available information on toxicity profile and mode of action should be taken into account. If, as a result of this assessment, the active substance is identified as a potential endocrine disruptor, the type and conditions of the studies to be performed should be discussed with the national competent authorities.

(c): The OECD TG 229 and 230 have a similar study design and include similar endpoints except for fecundity, gonad histology/ histopathology which are only measured in the OECD TG 229.

(d): A summary of all relevant data from the scientific peer-reviewed open literature on the active substance, metabolites and breakdown or reaction products and plant protection products containing the active substance should be submitted according to (EFSA, 2011).

Reference

EFSA (European Food Safety Authority), 2011. Submission of scientific peer-reviewed open literature for the approval of pesticide active substances under Regulation (EC) No 1107/2009 (OJ L 309, 24.11.2009, p. 1–50). EFSA Journal 2011;9(2):2092, 49 pp. https://doi.org/10.2903/j.efsa.2011.2092

Appendix D – Databases, software tools and literature-derived (Q)SARs

D.1. Databases with information relevant to ED identification

Database	Link	Availability	Description
Endocrine Disruptor Knowledge Base (EDKB) database (FDA)	http://www.fda.gov/ScienceResearch/BioinformaticsTools/EndocrineDisruptorKnowledgebase/default.htm	Freely available	Biological activity database (Ding et al., 2010) including *in vitro* and *in vivo* experimental data with over 3,000 records for more than 1,800 chemicals, as well as chemical structure search capabilities. Among the data are an ER binding data set (containing 131 ER binders and 101 non-ER binders), and an AR binding data set (containing 146 AR binders and 56 non-AR binders). Searchable by assay type and by structure; provides a search ranking based on a structure similarity index
Estrogenic Activity Database (EADB) (FDA)	http://www.fda.gov/ScienceResearch/BioinformaticsTools/EstrogenicActivityDatabaseEADB/default.htm	Freely available	EADB (Shen et al., 2013) contains a comprehensive set of estrogenic activity data and is a component of the enhanced EDKB. It contains 18,114 estrogenic activity data points for 8,212 chemicals tested in 1,284 binding assays, reporter gene assays, cell proliferation assays, and *in vivo* assays in 11 different species. Software that allows for the generation of Decision Forest models that can be used to predict ED or other endpoints is also available on the same website
Endocrine Disruption Screening Program for the 21st Century (EDSP21) Dashboard (US EPA)	https://actor.epa.gov/edsp21/	Freely available	Provides access to new chemical data on over 1,800 chemicals of interest, to help the Endocrine Disruptor Screening Program evaluate chemicals for endocrine-related activity. Data sources: ToxCast/Tox21 HTS data, ExpoCastDB, DSSTox, PhysChemDB
Endocrine Active Substances Information System (EASIS) (European Commission)	https://easis.jrc.ec.europa.eu/	Freely available	Searchable database giving information on chemical identity (e.g. CAS number), chemical structure, toxicity (both to humans and wildlife), mode of action, for about 520 chemicals, including those on the EU priority list of substances
NURSA (Nuclear Receptor Signalling Atlas)	http://www.nursa.org/	Freely available	Information on chemical structure, crystal structure, SMILES, physical descriptors, nuclear receptors and mechanism of endocrine action
OECD (Q)SAR Toolbox (OECD, ECHA)	https://www.qsartoolbox.org/	Freely available	Although primarily a tool for chemical categories and read-across, it also includes several databases, including: 166,072 ER binding data from Danish EPA (pregenerated predictions, not experimental values) as well as 1,606 experimental ER binding affinity values from the OASIS commercial database, with Relative ER Binding Affinity data, where the data generated is all relative to the positive control 17-beta-oestradiol

Database	Link	Availability	Description
Toxicology Data Network (Toxnet) Developmental and Reproductive Toxicology Database (DART)	https://toxnet.nlm.nih.gov/newtoxnet/dart.htm	Freely available	Bibliographic database containing over 200,000 references to literature published since 1965. It covers teratology and other aspects of developmental and reproductive toxicology. Users can search by subject terms (e.g. endocrine disruptor), title words, chemical name, Chemical Abstracts Service Registry Number, and author
ToxRefDB (US EPA)	https://www.epa.gov/sites/production/files/2015-08/documents/readme_toxrefdb_20141106.pdf	Freely available (as MS Excel files - ftp://newftp.epa.gov/comptox/High_Throughput_Screening_Data/Animal_Tox_Data)	Contains mammalian toxicity information for over 400 pesticides reviewed by the US EPA Office of Pesticide Programs
Toxicity ForeCaster (ToxCast™) Data (US EPA)	https://www.epa.gov/chemical-research/toxicity-forecaster-toxcasttm-data https://actor.epa.gov/dashboard/	Freely available	The ToxCast webpage includes links to downloads of data sets such as • ToxCast & Tox21 data spreadsheet • Data and supplemental files from the CERAPP project • HTS data used for the estrogen receptor model (ToxCast ER prediction model (Judson et al., 2015)) The iCSS ToxCast (AcToR) Dashboard can be searched for HTS data on over 9,000 chemicals and information on approximately 1,000 assay endpoints
eChem Portal (OECD)	https://www.echemportal.org/echemportal/index.action	Freely available	Webportal that allows searches in 37 data sets with a total of 824,153 chemicals across 822,671 endpoints including developmental toxicity and reprotox. Some of the data sets present are ECHA Chem, ACToR, EFSA's Chemical Hazards Database, and JECDB
AOP Knowledge Base in e.AOP.Portal (OECD)	https://aopkb.oecd.org/index.html	Freely available	The OECD e.AOP.Portal is the main entry point for the AOP Knowledge Base (AOP-KB), a web-based platform which aims to bring together all knowledge on how chemicals can induce adverse effects
COSMOS DB	http://cosmosdb.eu/	Freely available	COSMOS DB is a database compiled within the EU FP7 COSMOS project and contains over 12,500 toxicity studies for 1,660 compounds across 27 endpoints, including developmental and reproductive toxicity. COSMOS DB Version 2 is supported by the COSMOS DataShare Point initiative
Danish (Q)SAR Database	http://qsar.food.dtu.dk/	Freely available	The Danish (Q)SAR database is a repository of estimates from over 200 (Q)SAR models from free and commercial platforms for over 600,000 chemicals. The (Q)SAR models include endpoints for physicochemical properties, environmental fate, ecotoxicity, absorption, metabolism and toxicity. The human health endpoints include ER, TR, PXR binding, ER activation, AR antagonism and teratogenic potential
(Q)SAR Data Bank	https://qsardb.org/	Freely available	(Q)SARDB is a repository for (Q)SAR and QSPR models and data sets. It includes (Q)SAR prediction results for ER binding and developmental toxicity

D.2. Software tools for predicting endocrine activity

Software	Link	Availability	Effect addressed	Description
Endocrine Disruptor Knowledge Base (EDKB) database (FDA)	http://www.fda.gov/ScienceResearch/BioinformaticsTools/EndocrineDisruptorKnowledgebase/default.htm	Freely available	A, E	Quantitative models to predict the binding affinity of compounds to the estrogen and androgen nuclear receptor proteins
ADMET Predictor (Simulations Plus Inc.)	https://www.simulations-plus.com/software/admetpredictor/	Commercial	E	Qualitative and quantitative prediction of estrogen receptor toxicity in rats. Based on two models: a qualitative model and, if toxic, the quantitative ratio of IC_{50} oestradiol/IC_{50} compound
ACD/Labs Percepta Predictors - Toxicity Module	http://www.acdlabs.com/products/percepta/predictors.php	Commercial	E	ER binding affinity prediction. Identify and visualise specific structural toxicophores. Identify analogues from its training set. Algorithms and data sets not disclosed. Predictions associated with confidence intervals and probabilities, providing prediction reliability
Derek Nexus (Lhasa Ltd)	http://www.lhasalimited.org	Commercial	E	Classification models (different levels of likelihood) based on four alerts for estrogenicity
MolCode Toolbox (Molcode Ltd)	http://molcode.com	Commercial	E, Other	Quantitative prediction of rat ER binding affinity and AhR binding affinity
CASE Ultra (MultiCASE Inc.)	www.multicase.com	Commercial	E, A	Quantitative models predicting the likelihood of estrogen and androgen receptor binding potential in terms of RBA. Binary models classify a chemical to be an ER or AR binder or not. Both types of models identify structural alerts that may contribute to activity
TIMES (Laboratory of Mathematical Chemistry, Bourgas University)	http://oasis-lmc.org	Commercial	E, A, Other	Classification models for the prediction of estrogen, androgen and aryl hydrocarbon binding. The chemical is predicted to fall in one of several activity bins (ranges of binding affinity)
VirtualToxLab (Vedani et al., 2009a,b)	http://www.biograf.ch	Commercial	E, A, T, S, Other	Classification model for endocrine-disrupting potential based on simulations of the interactions towards aryl hydrocarbon, estrogen α/β, androgen, thyroid α/β, glucocorticoid, liver X, mineralocorticoid, peroxisome proliferator-activated receptor γ, as well as the enzymes CYP450 3A4 and 2A13. Based on a fully automated protocol. The interactions with the macromolecular targets are simulated and quantified in terms of individual binding affinities, combining the flexible docking routine with multidimensional (Q)SAR

Software	Link	Availability	Effect addressed	Description
OECD (Q)SAR Toolbox (OECD, ECHA)	https://www.qsartoolbox.org	Freely available	E	The OECD (Q)SAR Toolbox (Dimitrov et al., 2016; OECD, 2014a,b) is a standalone software application for assessing the hazards of chemicals by grouping substances into categories and filling data gaps. It includes several databases that can be searched as well as (Q)SAR models, such as the MultiCASE ERBA (Q)SAR, which is based on a hierarchical statistical analysis of a training set composed of structures and ER binding data of 313 chemicals, the OASIS ERBA, the Danish EPA's Relative ERBA (Q)SAR and an expert system from US EPA based upon binding to the rainbow trout ER (rtER)
Endocrine Disruptome (Faculty of Pharmacy, University of Ljubljana, National Institute of Chemistry, Slovenia)	http://endocrinedisruptome.ki.si/	Freely available	E, A, T, S, Other	Web service for predicting endocrine disruption potential of molecules, entering structure/SMILES information (Kolsek et al., 2014). Includes docking to 18 crystal structures of 14 different nuclear receptors (e.g. AR, ER, GR, LXR, PPAR, RXR, TR)
EU project COSMOS KNIME workflow	https://knimewebportal.cosmostox.eu; model executable in the browser of the WebPortal	Freely available	E, A, T, S, Other	Prediction of potential NR binding (PPAR, AR, AhR, ER, GR, PR, farnesoid X receptor (FXR), LXR, PXR, TR, VDR, RXR). Developed by studying the physicochemical features of known nuclear receptor binders and elucidating the structural features needed for binding to the ligand binding pocket using the Protein Data Bank and ChEMBL. Evaluation of potential receptor binding based on the structural fragments and physicochemical features that were identified as essential to bind to the NR and induce a response
Chemotyper (Altamira, LLC)	https://chemotyper.org	Freely available		Software tool that allows the screening of data sets against a predefined set of 686 chemotypes that can be related to a range of molecular initiating events and adverse outcomes (Yang et al., 2015)
Danish (Q)SAR Database	http://qsar.food.dtu.dk	Freely available	E, A, T, Other	The Danish (Q)SAR database is a repository of pregenerated estimates from over 200 (Q)SAR models from free and commercial platforms for over 600,000 chemicals. The (Q)SAR for human health endpoints include ER, TR, PXR binding, ER activation, AR antagonism
(Q)SAR Data Bank ((Q)SARDB)	https://qsardb.org/	Freely available	E	(Q)SARDB (Ruusmann et al., 2015) is a repository for (Q)SAR and QSPR models and data sets. Some models can be downloaded or executed directly from the website. They can be referred to via unique and persistent identifiers (HDL and DOI). It includes (Q)SAR models for predicting ER binding
Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS) (US EPA)	https://www.epa.gov/chemical-research/sequence-alignment-predict-across-species-susceptibility	Freely available	Extrapolation of toxicity information across species	SeqAPASS is an online screening tool that allows to extrapolate toxicity information across species. Using the National Center for Biotechnology Information (NCBI) protein database SeqAPASS evaluates the similarities of amino acid sequences and protein structure to identify whether a protein target is present for a chemical interaction in other non-target species

D.3. Literature-derived (Q)SAR models for predicting nuclear receptor binding

The table lists examples of (Q)SAR models predicting nuclear receptor binding from the scientific literature. It is not exhaustive. It does not imply endorsement of the listed models or non-endorsement of not listed models. The applicability, e.g. applicability domain of the models and relevance for the specific assessment, should be derived on a case-by-case basis.

Model reference	Effect addressed	Method/type of model	Data set size and applicability
AR binding			
Hong et al. (2003)	Rat AR binding	3D (Q)SAR (CoMFA)	Training set consisting of 146 compounds with relative binding assay data determined with a competitive binding assay using a recombinant rat AR ligand binding domain protein commercially available. Predictive power was determined by leave-one-out
Soderholm et al. (2008)	AR binding	3D (Q)SAR and docking	219,680 compounds from Asinex commercial library (http://www.asinex.com)
Tamura et al. (2006)	AR binding	3D (Q)SAR (CoMFA)	35 chemicals for antagonists model and 13 chemicals for agonist and antagonist activity models
Todorov et al. (2011)	AR binding	COmmon REactivity PAttern (COREPA) modelling approach	202 structurally diverse chemicals with relative binding data obtained from a competitive radiometric binding assay, using radiolabeled [3*H*]–R1881 as the tracer and AR recombinant rat protein expressed in *Escherichia coli*
Vinggaard et al. (2008)	Human AR binding	MultiCASE analysis to identify the most representative chemical fragments responsible for the AR antagonism	Training consisting of 523 chemicals covering a wide range of chemical structures (e.g. organochlorines and polycyclic aromatic hydrocarbons) and various functions (e.g. natural hormones, pesticides, plasticisers, plastic additives, brominated flame retardants and roast mutagens)
Zhao et al. (2005)	AR binding	(Q)SARs based on multiple linear regression, radical basis function neural network and support vector machine (SVM)	146 structurally diverse natural, synthetic and environmental chemicals
ER binding			
Akahori et al. (2005)	Human ERα binding	A two-step (Q)SAR using discriminant and multilinear regression (MLR) analyses	Alkylphenols, phthalates, diphenylethanes and benzophenones
Asikainen et al. (2004)	ERα and ERβ binding	Consensus kNN (Q)SAR	Calf (53), mouse (68), rat (130), human ERα (61), human ERβ (61)
Browne et al. (2015), Judson et al. (2015)	ER bioactivity	ToxCast ER predictive model: Computational network model integrating 18 *in vitro* HTS assays measuring ER binding, dimerisation, chromatin binding, transcriptional activation and ER-dependent cell proliferation	The data set comprises concentration-response data on 1,812 chemicals with full data on ER pathway *in vitro* assaysActivity patterns across the *in vitro* assays are used to predict ER agonist or antagonist bioactivity and discriminate from assay-specific interference and cytotoxicity

Model reference	Effect addressed	Method/type of model	Data set size and applicability
Demyttenaere-Kovatcheva et al. (2005)	ER α and β	CoMFA	Diphenolic Azoles: 72 in training and 32 in test set
Fang et al. (2001)	Rat ER binding	Pharmacophore by CATALYST	232 chemicals from NCTR data set
Ghafourian et al. (2005)	Rat ER binding	TSAR 3D and 2D descriptors, partial least-squares (PLS) analysis by SIMCA-P, cluster analysis in MINITAB	131 chemicals from NCTR data set
Hong et al. (2005)	ER binding	Decision forest	232 structurally diverse compounds, validated using a test set of 463 compounds
Islam et al. (2008)	ER binding	Pharmacophore by Catalyst	35 compounds in the training set plus 102 compounds in the test set
Kramer and Giesy (1999)	Bovine calf uterine ER binding	Quantitative structure-binding relationship (QSBR)	25 hydroxy PCBs
Kurunczi et al. (2005)	Rat ER binding	PLS model	45
Lill et al. (2004)	ER binding	Multidimensional (Q)SAR (Raptor)	116 chemicals from NCTR data set
Marini et al. (2005)	ER binding	Various multivariate methods e.g. a back-propagation neural network	132 heterogeneous compounds
Mansouri et al. (2016), Marini et al. (2005) (CERAPP project: Collaborative Estrogen Receptor Activity Prediction Project)	*In vitro* and *in vivo* ER activity	(Q)SAR modelling by hierarchical clustering: classification models to predict *in vitro* and *in vivo* ER activity (binding, agonist, antagonist *in vitro* ER activity, and mouse *in vivo* uterotrophic ER binding)	*In vitro* ER activity data from different sources including the Tox21 (~ 8,000 chemicals in four assays), EADB (~ 8,000 chemicals), METI (~ 2,000 chemicals), ChEMBL (~ 2,000 chemicals)*In vitro* ER activity data from EADB(Q)SAR and docking approaches were used with a common training set of 1,677 chemical structures from the US EPA, resulting in a total of 40 categorical and 8 continuous models developed for binding, agonist and antagonist ER activity
Mekenyan et al. (2010)	ER binding	COREPA modelling approach combined with metabolic simulation	645 chemicals, including 497 steroid and environmental chemicals and 148 chemicals synthesised for medicinal purposes
Mukherjee et al. (2005)	ER binding	(Q)SAR based on multiple linear regression	25 triphenylacrylonitriles
Netzeva et al. (2006)	Estrogen-responsive gene expression *in vitro* reporter gene assay	Classification tree	117 aromatic compounds published including bisphenols, benzophenones, flavonoids, biphenyls, phenols and other aromatic chemicals
Ng et al. (2014)	ER binding	Competitive docking approach for performing ligand-docking in ERs. Ability to distinguish agonists from antagonists	Three sets of ligands: 66 compounds (47 agonists and 19 antagonists) extracted from PDB ERα complexes; 106 ER binders from the DUD (67 agonists, 39 antagonists); 4,018 ER decoys (2,570 agonist decoys, 1,448 antagonist decoys) from the DUD

Model reference	Effect addressed	Method/type of model	Data set size and applicability
Ribay et al. (2016)	ERα binding	Enhanced predictive model developed by using advanced cheminformatics tools integrating publicly available bioassay data; hybrid model performance showed significant improvement over the original (Q)SAR models	Training set: 259 binders and 259 non-binders. 264 external compounds
Saliner et al. (2006)	Human ERα binding	Models developed using quantum similarity methods	117 aromatic chemicals
Salum Lde et al. (2007))	ERα modulators	3D (Q)SAR (CoMFA) and 2D Hologram (Q)SAR	Two training sets containing either 127 or 69 compounds
Salum et al. (2008)	Binding affinity values for both ERα and ERβ	3D (Q)SAR: CoMFA and GRID	81 hER modulators
Taha et al. (2010)	ERβ binding	Pharmacophore modelling by CATALYST	Training set: 119 compounds; Test set: 23 compounds
Tong et al. (2004)	ER binding	Decision Forest classifier	Data set 1 : 232 chemicals tested in-house (131 active, 101 inactive)
			Data set 2:, literature compilation of 1,092 chemicals (350 active, 736 inactive)
Vedani et al. (2005)	Rat ER binding	Protein Modelling and 6D-(Q)SAR	106 compounds
Zhang et al. (2013)	ER binding	Quantitative prediction of binding affinity to both ER subtypes. Concurrent use of structure-based docking as complement to (Q)SARs for binding affinity in a consensus prediction approach	Database of relative binding affinity of a large number of ERα and/or ERβ ligands (546 for ERα and 137 for ERβ)
Other nuclear receptor binding			
Dybdahl et al. (2012)	Pregnane X receptor	(Q)SAR model for human pregnane X receptor (PXR) binding	631 molecules (299 positives and 332 negatives) with human PXR LBD binding assay. Cross-validation of the model showed a sensitivity of 82%, a specificity of 85%, and a concordance of 84%
Hong et al. (2016)	Rat α-fetoprotein binding activity	Model developed using a novel pattern recognition method (Decision Forest), the molecular descriptors were calculated from two-dimensional structures by Mold2 software	125 training chemicals (average balanced accuracy of 69%), external validation with 22 chemicals (balanced accuracy of 71%)
Huang et al. (2016)	NR	Cluster-based approach	Based on the structural information and activity data from the Tox21 10k library for nuclear receptor and stress response pathway assays (over 50 million data points), predictive models for 72 *in vivo* toxicity end points were built
Lagarde et al. (2016)	NR binding	3D agonist and antagonist selective pharmacophores; structure-based and ligand -based pharmacophore modelling	7,853 actives, 458,981 decoys, and 339 structures divided into 54 data sets form the NRLiSt BDB (http://nrlist.drugdesign.fr)

Model reference	Effect addressed	Method/type of model	Data set size and applicability
Lill et al. (2005)	AhR, ER, AR binding affinity	Multidimensional-dimensional (Q)SAR: Quasar and Raptor	Database containing 121 Aryl hydrocarbon compounds (91 training and 30 external test), 116 ER (93/23) and 72 AR (56/16)
Mellor et al. (2016), Steinmetz et al. (2015)	NR binding: PPAR, AR, AhR, ER, GR, PR, FXR, LXR, PXR, TR, VDR, RXR	Prediction of potential NR binding; freely available at https://knimewebportal.cosmostox.eu	Developed by studying the physicochemical-chemical features of known nuclear receptor binders and elucidating the structural features needed for binding to the ligand- binding pocket using the Protein Data Bank and ChEMBL
Al Sharif et al. (2017), Tsakovska et al. (2014)	Potential for full PPARγ agonism	PPARγ virtual screening. PPARγ active full agonists share at least four common pharmacophoric features; the most active ones have additional interactions	Developed taking into consideration structural elements (e.g. hydrogen bonds, hydrophobic and aromatic) of the ligands essential for their interactions with the receptor. The key protein interaction of the most active agonists include hydrogen binding to 4/5 amino acids in the receptor pocket; the most active agonists interact directly with H12 residues

AhR: aryl hydrocarbon receptor; AR: androgen receptor; ER: estrogen receptor; ERα: estrogen receptor alpha; ERβ: estrogen receptor beta; FXR: farnesoid X receptor; GR: glucocorticoid receptor; LXR: liver X receptor; NR: nuclear receptor; PPAR: peroxisome proliferator-activated receptor; PR: progesterone receptor; PXR: pregnane X receptor; RXR: retinoic acid receptor; THR: thyroid hormone receptor; VDR: vitamin D receptor.

References

Akahori Y, Nakai M, Yakabe Y, Takatsuki M, Mizutani M, Matsuo M and Shimohigashi Y, 2005. Two-step models to predict binding affinity of chemicals to the human estrogen receptor alpha by three-dimensional quantitative structure-activity relationships (3D-QSARs) using receptor-ligand docking simulation. SAR and QSAR in Environmental Research, 16, 323–337. https://doi.org/10.1080/10659360500204442

Al Sharif M, Tsakovska I, Pajeva I, Alov P, Fioravanzo E, Bassan A, Kovarich S, Yang C, Mostrag-Szlichtyng A, Vitcheva V, Worth AP, Richarz AN and Cronin MT, 2017. The application of molecular modelling in the safety assessment of chemicals: a case study on ligand-dependent PPARgamma dysregulation. Toxicology, 392, 140–154. https://doi.org/10.1016/j.tox.2016.01.009

Asikainen AH, Ruuskanen J and Tuppurainen KA, 2004. Consensus kNN QSAR: a versatile method for predicting the estrogenic activity of organic compounds *in silico*. A comparative study with five estrogen receptors and a large, diverse set of ligands. Environmental Science and Technology, 38, 6724–6729. https://www.ncbi.nlm.nih.gov/pubmed/15669333

Browne P, Judson RS, Casey WM, Kleinstreuer NC and Thomas RS, 2015. Screening chemicals for estrogen receptor bioactivity using a computational model. Environmental Science and Technology, 49, 8804–8814. https://doi.org/10.1021/acs.est.5b02641

Demyttenaere-Kovatcheva A, Cronin MT, Benfenati E, Roncaglioni A and Lopiparo E, 2005. Identification of the structural requirements of the receptor-binding affinity of diphenolic azoles to estrogen receptors alpha and beta by three-dimensional quantitative structure-activity relationship and structure-activity relationship analysis. Journal of Medicinal Chemistry, 48, 7628–7636. https://doi.org/10.1021/jm050604j

Dimitrov SD, Diderich R, Sobanski T, Pavlov TS, Chankov GV, Chapkanov AS, Karakolev YH, Temelkov SG, Vasilev RA, Gerova KD, Kuseva CD, Todorova ND, Mehmed AM, Rasenberg M and Mekenyan OG, 2016. QSAR Toolbox - workflow and major functionalities. SAR and QSAR in Environmental Reserach, 2016, 1–17. https://doi.org/10.1080/1062936x.2015.1136680

Ding D, Xu L, Fang H, Hong H, Perkins R, Harris S, Bearden ED, Shi L and Tong W, 2010. The EDKB: an established knowledge base for endocrine disrupting chemicals. BMC Bioinformatics, 11 (Suppl 6), S5. https://doi.org/10.1186/1471-2105-11-s6-s5

Dybdahl M, Nikolov NG, Wedebye EB, Jonsdottir SO and Niemela JR, 2012. QSAR model for human pregnane X receptor (PXR) binding: screening of environmental chemicals and correlations with genotoxicity, endocrine disruption and teratogenicity. Toxicology and Applied Pharmacology, 262, 301–309. https://doi.org/10.1016/j.taap.2012.05.008

ECHA (European Chemicals Agency) and EFSA (European Food Safety Authority), 2018. Outcome of the different consultations with relevant parties on the Guidance for the identification of endocrine disruptors in the context of Regulations (EU) No. 528/2012 and (EC) No. 1107/2009. EFSA supporting publication 2018:EN-1447. 9 pp. doi:10.2903/sp.efsa.2018.EN-1447; ECHA-18-G-01-EN

Fang H, Tong W, Shi LM, Blair R, Perkins R, Branham W, Hass BS, Xie Q, Dial SL, Moland CL and Sheehan DM, 2001. Structure-activity relationships for a large diverse set of natural, synthetic, and environmental estrogens. Chemical Research in Toxicology, 14, 280–294. https://www.ncbi.nlm.nih.gov/pubmed/11258977

Ghafourian T and Cronin MT, 2005. The impact of variable selection on the modelling of oestrogenicity. SAR and QSAR in Environmental Reserach, 16(1–2), 171–190. https://doi.org/10.1080/10629360412331319808

Hong H, Fang H, Xie Q, Perkins R, Sheehan DM and Tong W, 2003. Comparative molecular field analysis (CoMFA) model using a large diverse set of natural, synthetic and environmental chemicals for binding to the androgen receptor. SAR and QSAR in Environmental Reserach, 14, 373–388. https://doi.org/10.1080/10629360310001623962

Hong H, Shen J, Ng HW, Sakkiah S, Ye H, Ge W, Gong P, Xiao W and Tong W, 2016. A rat alpha-fetoprotein binding activity prediction model to facilitate assessment of the endocrine disruption potential of environmental chemicals. International Journal of Environmental Reserach and Public Health, 13, 372. https://doi.org/10.3390/ijerph13040372

Hong H, Tong W, Xie Q, Fang H and Perkins R, 2005. An *in silico* ensemble method for lead discovery: decision forest. SAR and QSAR in Environmental Reserach, 16, 339–347. https://doi.org/10.1080/10659360500203022

Huang R, Xia M, Sakamuru S, Zhao J, Shahane SA, Attene-Ramos M, Zhao T, Austin CP and Simeonov A, 2016. Modelling the Tox21 10 K chemical profiles for *in vivo* toxicity prediction and mechanism characterization. Nature Communications, 7, 10425. https://doi.org/10.1038/ncomms10425

Islam MA, Nagar S, Das S, Mukherjee A and Saha A, 2008. Molecular design based on receptor-independent pharmacophore: application to estrogen receptor ligands. Biological & Pharmaceutical Bulletin, 31, 1453–1460. https://doi.org/10.1248/bpb.31.1453

Judson RS, Magpantay FM, Chickarmane V, Haskell C, Tania N, Taylor J, Xia MH, Huang RL, Rotroff DM, Filer DL, Houck KA, Martin MT, Sipes N, Richard AM, Mansouri K, Setzer RW, Knudsen TB, Crofton KM and Thomas RS, 2015. Integrated model of chemical perturbations of a biological pathway using 18 in vitro high-throughput screening assays for the estrogen receptor. Toxicological Sciences, 148, 137–154. https://doi.org/10.1093/toxsci/kfv168

Kolsek K, Mavri J, Sollner Dolenc M, Gobec S and Turk S, 2014. Endocrine disruptome–an open source prediction tool for assessing endocrine disruption potential through nuclear receptor binding. Journal of Chemical Information and Modeling, 54, 1254–1267. https://doi.org/10.1021/ci400649p

Kramer VJ and Giesy JP, 1999. Specific binding of hydroxylated polychlorinated biphenyl metabolites and other substances to bovine calf uterine estrogen receptor: structure-binding relationships. Science of the Total Environment, 233, 141–161. https://doi.org/10.1016/s0048-9697(99)00221-1

Kurunczi L, Seclaman E, Oprea TI, Crisan L and Simon Z, 2005. MTD-PLS: A PLS variant of the minimal topologic difference method. III. Mapping interactions between estradiol derivatives and the alpha estrogenic receptor. Journal of Chemical Information and Modeling, 45, 1275–1281. https://doi.org/10.1021/ci050077c

Lagarde N, Delahaye S, Zagury JF and Montes M, 2016. Discriminating agonist and antagonist ligands of the nuclear receptors using 3D-pharmacophores. Journal of Cheminformatics, 8. https://doi.org/10.1186/s13321-016-0154-2

Lill MA, Dobler M and Vedani A, 2005. In silico prediction of receptor-mediated environmental toxic phenomena - application to endocrine disruption. Sar and Qsar in Environmental Research, 16, 149–169. https://doi.org/10.1080/10629360412331319826

Lill MA, Vedani A and Dobler M, 2004. Raptor: combining dual-shell representation, induced-fit simulation, and hydrophobicity scoring in receptor modeling: Application toward the simulation of structurally diverse ligand sets. Journal of Medicinal Chemistry, 47, 6174–6186. https://doi.org/10.1021/jm049687e

Mansouri K, Abdelaziz A, Rybacka A, Roncaglioni A, Tropsha A, Varnek A, Zakharov A, Worth A, Richard AM, Grulke CM, Trisciuzzi D, Fourches D, Horvath D, Benfenati E, Muratov E, Wedebye EB, Grisoni F, Mangiatordi GF, Incisivo GM, Hong H, Ng HW, Tetko IV, Balabin I, Kancherla J, Shen J, Burton J, Nicklaus M, Cassotti M, Nikolov NG, Nicolotti O, Andersson PL, Zang Q, Politi R, Beger RD, Todeschini R, Huang R, Farag S, Rosenberg SA, Slavov S, Hu X and Judson RS, 2016. CERAPP: collaborative estrogen receptor activity prediction project. Environmental Health Perspectives, 124, 1023–1033. https://doi.org/10.1289/ehp.1510267

Marini F, Roncaglioni A and Novic M, 2005. Variable selection and interpretation in structure-affinity correlation modeling of estrogen receptor binders. Journal of Chemical Information and Modeling, 45, 1507–1519. https://doi.org/10.1021/ci0501645

Mekenyan O and Serafimova R 2010. Mechanism-based modeling of estrogen receptor binding affinity. A Common Reactivity Pattern (COREPA) Implementation. In: Devillers J (ed.) Endocrine Disruption Modeling. CRC Press, Boca Raton, FL. pp. 259–293.

Mellor CL, Steinmetz FP and Cronin MT, 2016. The identification of nuclear receptors associated with hepatic steatosis to develop and extend adverse outcome pathways. Critical Reviews in Toxicology, 46, 138–152. https://doi.org/10.3109/10408444.2015.1089471

Mukherjee S, Saha A and Roy K, 2005. QSAR of estrogen receptor modulators: exploring selectivity requirements for ER(alpha) versus ER(beta) binding of tetrahydroisoquinoline derivatives using E-state and physicochemical parameters. Bioorganic Medicnal and Chemical Letters, 15, 957–961. https://doi.org/10.1016/j.bmcl.2004.12.048

Netzeva TI, Saliner AG and Worth AP, 2006. Comparison of the applicability domain of a quantitative structure-activity relationship for estrogenicity with a large chemical inventory. Environmental Toxicology and Chemistry, 25, 1223–1230. https://doi.org/10.1897/05-367r.1

Ng HW, Zhang WQ, Shu M, Luo H, Ge WG, Perkins R, Tong WD and Hong HX, 2014. Competitive molecular docking approach for predicting estrogen receptor subtype alpha agonists and antagonists. Bmc Bioinformatics, 15. https://doi.org/10.1186/1471-2105-15-s11-s4

OECD (Organisation for Economic Cooperation and Development) 2014a. The Guidance Document for Using the OECD (Q)SAR Application Toolbox to Develop Chemical Categories According to the OECD Guidance on Grouping Chemicals. In: OECD series on testing and assessment. OECD Publishing, Paris. 118 pp. https://doi.org/10.1787/9789264221482-en

OECD (Organisation for Economic Cooperation and Development), 2014b. Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models. In: OECD series on testing and assessment. OECD Publishing, Paris. 154 pp. https://doi.org/10.1787/9789264085442-en

Ribay K, Kim M, Wang W, Pinolini D and Zhu H, 2016. Predictive modelling of estrogen receptor binding agents using advanced cheminformatics tools and massive public data. Frontiers in Environmental Science, 4. https://doi.org/10.3389/fenvs.2016.00012

Ruusmann V, Sild S and Maran U, 2015. QSAR DataBank repository: open and linked qualitative and quantitative structure-activity relationship models. Journal of Cheminformatics, 7, 32. https://doi.org/10.1186/s13321-015-0082-6

Saliner AG, Netzeva TI and Worth AP, 2006. Prediction of estrogenicity: validation of a classification model. SAR and QSAR in Environmental Reserach, 17, 195–223. https://doi.org/10.1080/10659360600636022

Salum LB, Polikarpov I and Andricopulo AD, 2008. Structure-based approach for the study of estrogen receptor binding affinity and subtype selectivity. Journal of Chemical Information Modeling, 48, 2243–2253. https://doi.org/10.1021/ci8002182

Salum Lde B, Polikarpov I and Andricopulo AD, 2007. Structural and chemical basis for enhanced affinity and potency for a large series of estrogen receptor ligands: 2D and 3D QSAR studies. Journal of Molecular Graphics and Modeling, 26, 434–442. https://doi.org/10.1016/j.jmgm.2007.02.001

Shen J, Xu L, Fang H, Richard AM, Bray JD, Judson RS, Zhou G, Colatsky TJ, Aungst JL, Teng C, Harris SC, Ge W, Dai SY, Su Z, Jacobs AC, Harrouk W, Perkins R, Tong W and Hong H, 2013. EADB: an estrogenic activity database for assessing potential endocrine activity. Toxicological Sciences, 135, 277–291. https://doi.org/10.1093/toxsci/kft164

Soderholm AA, Viiliainen J, Lehtovuori PT, Eskelinen H, Roell D, Baniahmad A and Nyronen TH, 2008. Computationally identified novel diphenyl- and phenylpyridine androgen receptor antagonist structures. Journal of Chemical Information and Modeling, 48, 1882–1890. https://doi.org/10.1021/ci800149w

Steinmetz FP, Mellor CL, Meinl T and Cronin MT, 2015. Screening chemicals for receptor-mediated toxicological and pharmacological endpoints: using public data to build screening tools within a KNIME workflow. Molecular Information, 34, 171–178. https://doi.org/10.1002/minf.201400188

Taha MO, Tarairah M, Zalloum H and Abu-Sheikha G, 2010. Pharmacophore and QSAR modeling of estrogen receptor beta ligands and subsequent validation and in silico search for new hits. Journal of Molecular Graphics & Modelling, 28, 383–400. https://doi.org/10.1016/j.jmgm.2009.09.005

Tamura H, Ishimoto Y, Fujikawa T, Aoyama H, Yoshikawa H and Akamatsu M, 2006. Structural basis for androgen receptor agonists and antagonists: Interaction of SPEED 98-listed chemicals and related compounds with the androgen receptor based on an in vitro reporter gene assay and 3D-QSAR. Bioorganic and Medicinal Chemistry, 14, 7160–7174. https://doi.org/10.1016/j.bmc.2006.06.064

Todorov M, Mombelli E, Ait-Aissa S and Mekenyan O, 2011. Androgen receptor binding affinity: a QSAR evaluation. Sar and Qsar in Environmental Research, 22, 265–291. https://doi.org/10.1080/1062936x.2011.569508

Tong WD, Xie W, Hong HX, Shi LM, Fang H and Perkins R, 2004. Assessment of prediction confidence and domain extrapolation of two structure-activity relationship models for predicting estrogen receptor binding activity. Environmental Health Perspectives, 112, 1249–1254. https://doi.org/10.1289/txg.7125

Tsakovska I, Al Sharif M, Alov P, Diukendjieva A, Fioravanzo E, Cronin MTD and Pajeva I, 2014. Molecular modelling study of the PPAR gamma receptor in relation to the mode of action/adverse outcome pathway framework for liver steatosis. International Journal of Molecular Sciences, 15, 7651–7666. https://doi.org/10.3390/ijms15057651

Vedani A, Dobler M and Lill MA, 2005. Combining protein modeling and 6D-QSAR. Simulating the binding of structurally diverse ligands to the estrogen receptor. Journal of Medical Chemistry, 48, 3700–3703. https://doi.org/10.1021/jm050185q

Vedani A and Smiesko M, 2009a. In silico toxicology in drug discovery - concepts based on three-dimensional models. Alternatives to Laboratory Animals, 37, 477–496. https://www.ncbi.nlm.nih.gov/pubmed/20017578

Vedani A, Smiesko M, Spreafico M, Peristera O and Dobler M, 2009b. VirtualToxLab - in silico prediction of the toxic (endocrine-disrupting) potential of drugs, chemicals and natural products. Two years and 2,000 compounds of experience: a progress report. ALTEX, 26, 167–176. https://www.ncbi.nlm.nih.gov/pubmed/19907904

Vinggaard AM, Niemela J, Wedebye EB and Jensen GE, 2008. Screening of 397 chemicals and development of a quantitative structure–activity relationship model for androgen receptor antagonism. Chemical Reserach Toxicology, 21, 813–823. https://doi.org/10.1021/tx7002382

WHO/IPCS (World Health Organization/International Programme on Chemical Safety), 2009. Principles and Methods for the Risk Assessment of Chemicals in Food. Environmental Health Criteria 240. 689 pp. Available from: http://www.who.int/foodsafety/chem/principles/en/index1.html

Yang C, Tarkhov A, Marusczyk J, Bienfait B, Gasteiger J, Kleinoeder T, Magdziarz T, Sacher O, Schwab CH, Schwoebel J, Terfloth L, Arvidson K, Richard A, Worth A and Rathman J, 2015. New publicly available chemical query language, CSRML, to support chemotype representations for application to data mining and modeling. Journal of Chemical Information and Modeling, 55, 510–528. https://doi.org/10.1021/ci500667v

Zhang L, Sedykh A, Tripathi A, Zhu H, Afantitis A, Mouchlis VD, Melagraki G, Rusyn I and Tropsha A, 2013. Identification of putative estrogen receptor-mediated endocrine disrupting chemicals using QSAR- and structure-based virtual screening approaches. Toxicology and Applied Pharmacology, 272, 67–76. https://doi.org/10.1016/j.taap.2013.04.032

Zhao CY, Zhang RS, Zhang HX, Xue CX, Liu HX, Liu MC, Hu ZD and Fan, BT, 2005. QSAR study of natural, synthetic and environmental endocrine disrupting compounds for binding to the androgen receptor. SAR and QSAR in Environmental Reserach, 16, 349–367. https://doi.org/10.1080/10659360500204368

Appendix E – Excel template for reporting the available information relevant for ED assessment

See zip file 'EDGD_Appendix-E.zip'

E.1. Excel template for reporting effects

E.2. Guidance to fill in the 'Data' sheet template

The excel template mentioned in Section 3.2.2 suggested to gather the information in a tabular format, is published as a separate file along with the instructions on how to fill it in. This has to be considered as an independent document. Further revisions after the publication of the guidance could be performed by ECHA and EFSA.

Appendix F – Example on how to develop the search strategy protocol

This Appendix aims at giving some more guidance on how a systematic literature review can be conducted, focusing in particular on the choice of the bibliographic databases and on how to build a search string. It reports an example given by Berger et al. (2013) where the EFSA Guidance on systematic review (EFSA, 2010) was applied on real cases. The example was slightly adapted, where needed, in order to make it more fit for purpose in the context of this Guidance.

However, this appendix does not report the full process, as described in the **Figure F.1** and does not repeat the principles of the EFSA Guidance document (EFSA, 2011); as a consequence it is recommended not to read it in isolation but to always consult together with section 3.2 of this guidance and the EFSA Guidance documents (EFSA, 2010, 2011).

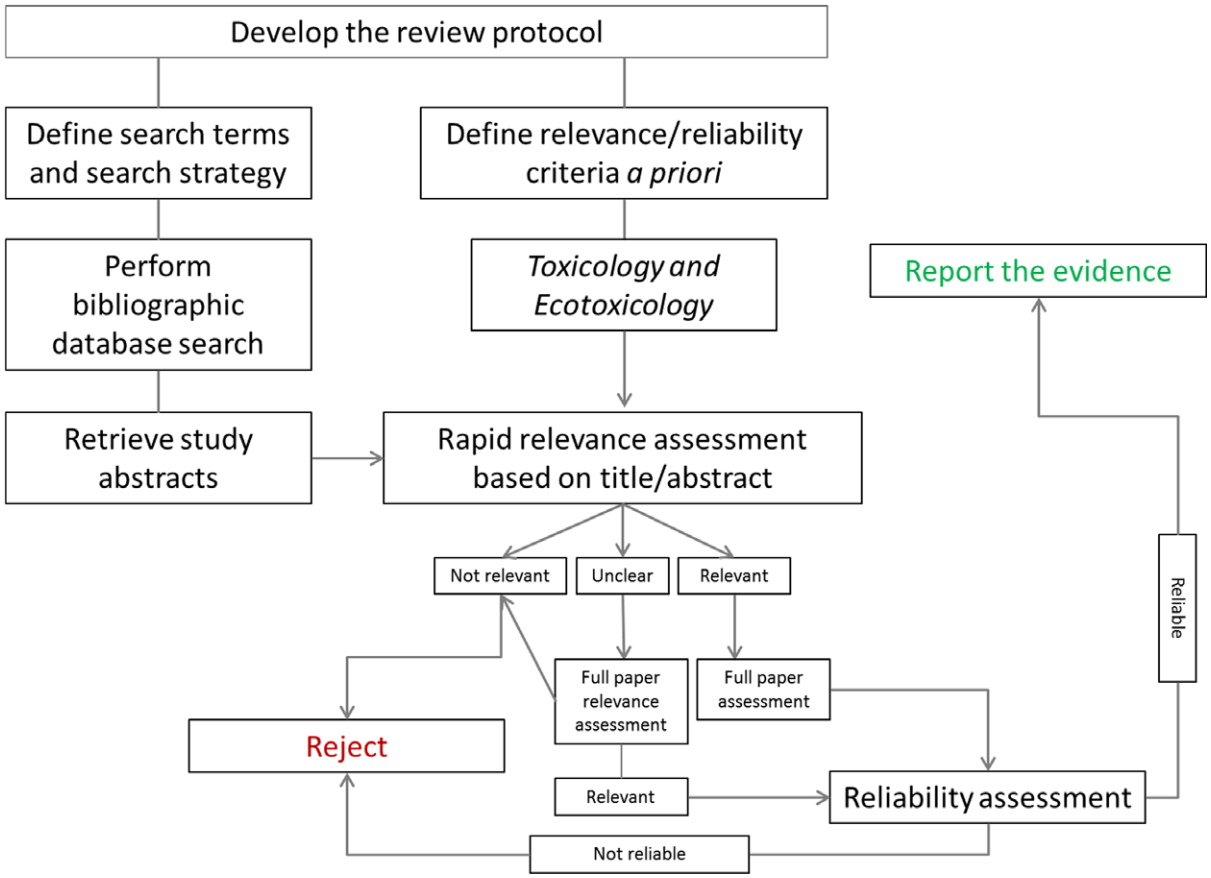


Figure F.1: Overall process of systematic literature review adapted from Berger et al. (2013)

The EFSA Guidance (2011) provides specific instructions with respect to article 8(5) of the Regulation (EC) No 1107/2009[2] stating the following: '*Scientific peer-reviewed open literature, as determined by the Authority, on the active substance and its relevant metabolites dealing with side-effects on health, the environment and non-target species and published within the last ten years before the date of submission of the dossier shall be added by the applicant to the dossier*'. The systematic literature review done in the context of this Guidance should, therefore, also be developed within this legal framework an in line with article 8(5) of the Regulation (EC) 1107/2009[2] for PPPs.

It has to be noted that in the Regulation (EU) No 528/2012[1] for biocides, there is no similar legal provision. Therefore, the time span needs to be decided and agreed on a case-by-case basis.

Although the scope of the guidance is restricted to EATS modalities and vertebrates, from a systematic literature review it is expected that, in general data on (i) ED modalities other than EATS and (ii) invertebrates are retrieved and reported (see Sections 2 and 3.1).

For the purpose of this guidance, only the example of metalaxyl-M is reported as an illustration out of the three used in Berger et al. (2013). Some information has been updated and/or adapted e.g. information on the bibliographic databases and search terms to include the last version of the data requirements.

Choice of the bibliographic databases

For the choice of the bibliographic database, the EFSA Guidance (EFSA, 2011) neither recommends specific database nor lists specific requirements guiding in the selection of bibliographic databases (e.g. subject areas covered, updating frequency). The applicant is, however, asked to justify the chosen sources and to demonstrate that reasonable efforts have been made to locate all sources of relevant literature. In the context of this Guidance, in addition to the bibliographic databases for peer review literature, it is recommended to always search other databases which are particularly relevant for endocrine disruptor properties, like ToxCast (see also Appendix D.1).

For metalaxyl-M, four different bibliographic databases were selected according to their subject areas and accessibility: Web of Knowledge (now Web of Science), Scopus, Agricola and Pubmed. It was showed that Web of Knowledge and Scopus appeared to yield comparable results since all relevant papers were found in both databases. In Table F.1, some features of the four selected databases are reported.

Table F.1: Features of the four selected databases

	Web of knowledge (now Web of Science)	Scopus	Agricola	Pubmed
Accessibility	License fee	License fee	Free	License fee
Areas covered	Biomedical sciences, natural sciences, engineering, social sciences, arts and humanities Strongest coverage of natural sciences & engineering, computer science, materials sciences, patents, data sets	Science, technology, medicine, social sciences, and arts and humanities	Animal and veterinary sciences Entomology Plant sciences Forestry Aquaculture and fisheries Farming and farming systems Agricultural economics Extension and education	Biomedical and life science
Type of publications	Articles, conference papers, monographs and reports	Articles and monographs	Articles, monographs, proceedings, theses, patents, translations, audio-visual materials, computer software, and technical reports	Articles, leaflets, monographs and news
Wild cards/ truncation	The asterisk (*) represents any group of characters, including no character The question mark (?) represents any single character The dollar sign ($) represents zero or one character At least three characters must precede the wildcard in Title and Topic searches. For example, zeo* is acceptable but ze* is not Wildcards may be used inside a word. For example, odo$r finds odor and odour You may use different wildcards in one term: l?chee$ matches lichee, lichees, lychee, lychees You cannot use wildcards after special characters (/@ #) and punctuation	? represents any single character * represents any number of characters, even zero Punctuation: Commas, hyphens, ?, ! etc., are ignored Stop words: Words like "the," "it," and "of" are excluded from search (Refer to the list found in Scopus help)Override with Exact phrase: {} will find only an exact match for a word, phrase or character (including stop words)	? = replaces a single character # = looks for alternate spellings * = serves as a truncation symbol to search for different forms of a word	* = serves as a truncation symbol

	Web of knowledge (now Web of Science)	Scopus	Agricola	Pubmed
Boolean operators	AND, OR, NOT, SAME, NEAR	AND, OR, NOT	OR, NOT, FREQ, AND, ADJn	AND, OR, NOT

For each of the searched database, specific consideration is needed when there is the need to search for phrase like fish common names, e.g. Fathead minnow. For example, in Web of Science, those need to be enclosed in quotation marks. In Scopus, phrases should be enclosed in double quote marks or curly brackets. • Double quotes ""will search for indistinct phrases. For example, "heart-attack" will search for heart-attack, heart attack, heart attacks, and so on • Curly brackets {} will search for a specific phrase. It limits the search to only the specified character string, and symbols can be used; {heart-attack} will only search for heart-attack.

Search strategy

In the EFSA Guidance (EFSA, 2011), two search strategies are proposed, the single concept and the targeted search strategy (based on information requirements) (see also Section 3.2). The applicant in consultation with Member States can decide which approach to follow. The most important aspect to consider when deciding the approach is to avoid bias and ensure extensiveness.

Single concept search strategy

For metalaxyl-M, the literature search also included information on the active substance Metalaxyl since metalaxyl-M is the biologically active isomer in Metalaxyl, which is a racemate of *R*-(metalaxyl-M) and *S*-isomers. In addition to the literature search on Metalaxylucodep>/ucodep> (both isomers), all known synonyms and chemical names were included:

- CAS Number: 57837-19-1 (metalaxyl)

  OR

- IUPAC name: Methyl *N*-(methoxyacetyl)-*N*-(2,6-xylyl)-DL-alaninate (metalaxyl)

  OR

- CAS Number: 70630-17-0 (metalaxyl-M)

  OR

- IUPAC name: Methyl(*R*)-2-{[(2,6-dimethylphenyl)methoxyacetyl]amino}propionate (metalaxyl-M)

  OR

- Mefenoxam (metalaxyl-M)

For the purpose of this guidance, all the search terms for the metabolites and product names are not used since metabolites and formulations are considered out of the scope.

Targeted search strategy

Toxicology

For toxicology general terms were used as well as specific terms linked to data requirements (only the relevant ones for the ED assessment are reported here):

- tox* OR hazard OR adverse OR health OR effect*
- NOAEL OR NOEL OR LOAEL OR LOEL OR BMD
- "in vivo" OR "in vitro"
- acute OR subacute OR subchronic OR chronic
- oral OR dermal OR gavage OR diet* OR inhal*
- rat* OR dog* OR rabbit* OR guinea pig* OR mouse OR mice OR hamster
- metabolism OR metabolite* OR metabolic OR distribution OR adsorption OR
- excretion OR elimination OR kinetic OR PBPK
- CYP OR cytochrome OR enzym*
- gen* OR muta* OR chromos* OR clastogen* OR DNA
- carcino* OR cancer*
- immun*
- neur* OR behav*
- endocrin* OR hormon*
- reproduct* OR development* OR malformation* OR anomal* OR fertil* OR foet*
- OR fet* OR matern* OR pregnan* OR embryo*
- epidem* OR medical* OR poison*

Ecotoxicology

For ecotoxicology the terms listed below were used:

- tox* OR hazard OR adverse OR poison OR effect*
- in vivo OR *in vitro*
- bird* OR mallard OR duck OR quail OR bobwhite OR Anas* OR Colinus*
- vertebrat* OR mammal* OR rat OR mouse OR mice OR rabbit OR hare
- invertebrat* OR aquatic OR fish OR fathead minnow OR Medaka OR zebrafish OR stickleback OR sheephead minnow OR daphni* OR chiron* OR sediment dwell* OR marin* OR estuarine OR crusta* OR gastropod* OR mollusc OR reptile OR amphib*
- endocrin*
- bee* OR api* OR bumble*
- arthropod* OR typhlodromus OR aphidius OR insect*
- worm* OR *worm OR eisenia
- collembol* OR macro organism OR folsomia OR springtail OR mite* OR Hypoaspis
- reproduct* OR development* OR malformation* OR anomal* OR fertil* OR fecund*

As explained above, this appendix just reports an example of the search terms targeted on the information requirements for PPPs. For biocides, therefore, the search terms need to be adapted in line with the biocidal information requirements and it may be that some search terms are not relevant.

In the context of this Guidance, it is suggested to perform as a starting point, the literature search by using the single concept approach since it is considered to be highly sensitive, and less time consuming than the targeted search strategy. If a large number of hits is retrieved by using the single concept approach, this can be further refined by running a search targeted on the information requirements.

For the definition of relevance and reliability criteria, please refer to Section 3.2 above in this guidance and to the EFSA Guidance on systematic review (EFSA, 2010, 2011).

Results of the search

The results of the search in Web of Science are provided both for toxicology (Table F.2) and ecotoxicology (Table F.3) as an example. In particular, it is illustrated how the search terms can be combined in a search string and how the results should be reported. The number does not match with the one reported by Berger et al. (2013) as the search was slightly adapted to the purpose of this Guidance, e.g. only the terms considered relevant for the ED assessment were used; the search was conducted without any temporal limits.

Table F.2: Results of the search for the database Web of Science, in terms of number of hits, for the toxicology

	Database: web of science (topic)
Search terms and combination	Number of hits at each step
metalaxyl* OR "57837-19-1" OR "Methyl N-(methoxyacetyl)-N-(2,6- xylyl)-DL-alaninate" OR 70630-17-0 OR "Methyl(R)-2-{[(2,6- dimethylphenyl) methoxyacetyl]amino} propionate" OR mefenoxam	2,119
tox* OR hazard OR adverse OR health OR effect*	12,624,692
NOAEL OR NOEL OR LOAEL OR LOEL OR BMD	31,009
acute OR subacute OR subchronic OR chronic	2,016,955
oral OR dermal OR gavage OR diet* OR inhal*	1,368,594
rat OR rats OR dog* OR rabbit* OR guinea pig* OR hamster OR mouse OR mice	3,380,980
metabolism OR metabolite* OR metabolic OR distribution OR adsorption OR excretion OR elimination OR kinetic OR PBPK	4,458,284
CYP OR cytochrome OR enzym*	1,279,215
gen* OR muta* OR chromos* OR clastogen* OR DNA	10,498,284
carcino* OR cancer*	2,502,916
"in vivo" OR "in vitro"	1,874,227

	Database: web of science (topic)
Mechanis*	3,357,312
Neur* OR behav*	5,361,618
reproduct* OR development* OR malformation* OR anomal* OR fertil* OR foet* OR fet* OR matern* OR pregnan* OR embryo*	5,469,463
Epidm* OR medicl* OR poison*	1,521,130
OR/2-15	31,268,400
1 AND 16	**1,593**

Table F.3: Results of the search for the database Web of Science, in terms of number of hits, for the ecotoxicology

	Database: web of science (topic)
Search terms and combination	Number of hits at each step
metalaxyl* OR "57837-19-1" OR "Methyl N-(methoxyacetyl)-N-(2,6- xylyl)-DL-alaninate" OR 70630-17-0 OR "Methyl(R)-2-{[(2,6- dimethylphenyl)methoxyacetyl] amino} propionate" OR mefenoxam	2,119
tox* OR hazard OR adverse OR health OR effect* OR chronic*	12,948,708
"in vivo" OR "in vitro"	1,874,227
Mechanis*	3,357,312
bird* OR mallard OR duck OR quail OR bobwhite OR Anas* OR Colinus*)	259,620
vertebrat* OR mammal* OR rat OR mouse OR mice OR rabbit OR hare	3,401,382
invertebrat* OR aquatic OR fish OR fathead minnow OR Medaka OR zebrafish OR stickleback OR sheephead minnow OR daphni* OR chiron* OR sediment dwell* ORmarin* OR estuarine OR crusta* OR gastropod* OR mollusc OR reptile OR amphib*	757,262
endocrin*	151,328
bee* OR api* OR bumble*	7,716,086
arthropod* OR typhlodromus OR aphidius OR insect*	260,781
worm* OR *worm OR eisenia	96,168
collembol* OR macro organism OR folsomia OR springtail OR mite* OR Hypoaspis	44,651
reproduct* OR development* OR malformation* OR anomal* OR fertil* OR fecund*	4,725,825
OR/2-13	23,928,276
14 AND 1	1,399

As reported in the table, the number of hits which would undergo the rapid relevance assessment, based on title and abstract, is **1,593** and **1,399,** for toxicology and ecotoxicology, respectively.

References

Berger E, Čoja T, Dellantonio A, Hrdina-Zödl B, Hutzenlaub N, Jölli D, Müller M and Prohaska C, 2013. Case studies for the application of the Guidance of EFSA on Submission of scientific peer-reviewed open literature for the approval of pesticide active substances under Regulation (EC) No 1107/2009, using substances for which dossiers are submitted under Regulation (EU) No 1141/2010. EFSA supporting publication 2013:EN-51. Available online: https://efsa.onlinelibrary.wiley.com/doi/pdf/10.2903/sp.efsa.2013.EN-511

EFSA (European Food Safety Authority), 2010. Application of systematic review methodology to food and feed safety assessments to support decision making. EFSA Journal 2010;8(6):1637, 90 pp. https://doi.org/10.2903/j.efsa.2010.1637

EFSA (European Food Safety Authority), 2011. Submission of scientific peer-reviewed open literature for the approval of pesticide active substances under Regulation (EC) No 1107/2009 (OJ L 309, 24.11.2009, p. 1–50). EFSA Journal 2011;9(2):2092, 49 pp. https://doi.org/10.2903/j.efsa.2011.2092

Appendix G – Example of MoA for non-target organisms (fish)

Based on the available information and lines of evidence for adversity and endocrine activity (see Table 3 in Section 3.3.3), a MoA for aromatase inhibition leading to reproductive dysfunction in fish can be postulated as shown in the figure below (see also AOP 25).

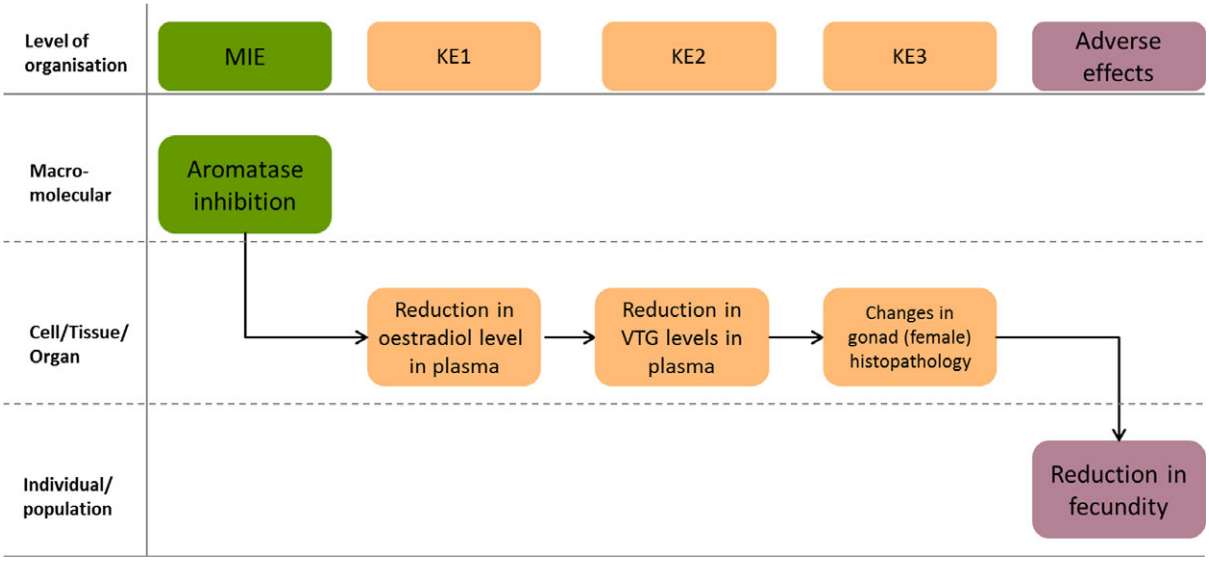


Figure G.1: Postulated MoA for aromatase inhibition leading to reproductive dysfunction in fish

The molecular initiating event is the inhibition of CYP 19 activity (AOP 25; Villeneuve 2016). This resulted in reduction in oestradiol and VTG levels in female fish leading to changes in female gonad histopathology (decreased yolk formation and decreased post-ovulatory follicles and mean ovarian stages scores). Those effects ultimately resulted in reduced fecundity.

Table G.1 supports the analysis of the dose–response and temporal concordance for the KEs, identified above, for the postulated MoA. This analysis should allow answering the questions reported in Section 3.5.1. It has to be noted that the design of ecotoxicological standard studies do not always allow for the assessment of the temporal concordance since the relevant parameters are measured at the end of each study only. Even the combination of different studies with different study design in terms of length could not properly address the temporal concordance since the majority of the available standard studies focus on sexually mature fish.

Table G.1: Example of a table which allows analysis of both dose–response and temporal concordance between the key events (KEs) for non-target organisms

***[Species: Pimephales promelas]* dose–response and temporal concordance between the key events**				
	KE 1 decreased oestradiol level	**KE 2 decreased VTG level**	**KE 3 gonad histopathology**	**Adverse effect fecundity**
Dose (µg/L)				
0.5	++ (3 weeks)	++ (3 weeks)		++ (3 weeks)
0.558		+ (36 weeks)	+ (36 weeks)	+ (36 weeks)
1		+ (3 weeks)		+ (3 weeks)

Only key events with available data for dose-response and temporal concordance are included.
+ indicates effects only observed at the highest tested dose, ++ indicates effects observed in a dose related manner.

The MoA is demonstrated and documented in Table G2 where the conclusion on the biological plausibility of the link between the adverse effect and the endocrine activity for the postulated MoA is reported together with the list of identified uncertainties.

Table G.2: Conclusion on the biological plausibility of the link between the adverse effect and the endocrine activity for the postulated MoA

	Key event relationships (KERs)			
	MIE to KE 1	**KE 1 to KE 2**	**KE 2 to KE 3**	**KE 3 to AE**
Biological plausibility for the KERs	STRONG – The link between aromatase inhibition and decrease in oestradiol level (E2) is supported by the available knowledge (AOP 25, Villeneuve 2016)	MODERATE – The role of E2 as major regulator of VTG production is well known. Therefore, it can be assumed that a decrease in oestradiol level will also lead to a decrease in VTG in plasma	MODERATE – Based on the available knowledge, it is not clear whether a decrease in VTG can lead to the observed histopathology changes in ovary. However, specific gonad histopathology is categorised as 'EAS-mediated' by the OECD GD 150. In addition, the link between VTG level and yolk formation is also supported by the biological knowledge	STRONG – the link between changes in female gonad histopathology and decreased fecundity is supported by the biological knowledge
Empirical support for the KERs	MODERATE – There is little direct support for dose-response concordance of these key events in vivo. However, using in vitro systems concentrations that reduce aromatase activity tend to elicit reductions in oestradiol production	STRONG – Although the decrease in oestradiol and VTG levels were observed at the same concentrations, this can be scientifically explained by a number of factors (e.g. dose spacing in the test system; higher variation in VTG concentration in plasma than in circulating steroids)	MODERATE – histopathology changes were measured only in longer term study and only observed at the highest tested concentration. The VTG decrease was observed at the same concentration. However, this can be due to the dose spacing and tested concentrations	STRONG – fecundity was observed at the same concentration as histopathology changes and above
Essentiality of KEs	MODERATE – No data are available to support the assessment of essentiality. However, the available knowledge and validated AOP (25) supports the essentiality of key events			
Consistency	The KEs have been observed consistently in three different studies with different duration. The pattern of effects is consistent between the studies; there are no conflicting observations. Consistency across species cannot be assessed because there are only studies on one species			
Analogy	Aromatase inhibition is well established for compounds belonging to the same chemical class			
Specificity	Liver histopathology changes observed in one study at the highest tested concentration where other effects were also observed. However, the positive indication of endocrine activity from various studies and cell lines allowed to exclude a non-ED MOA			

Identified uncertainties	**Comment**
Uncertainty 1 *[Measurements of aromatase inhibition]*	Direct measurements in vivo are difficult. In vitro measurements provide support; however, the extrapolation in vitro–*in vivo* is uncertain
Uncertainty 2 *[Empirical support for KER]*	A clear dose and temporal concordance cannot be established due to drawbacks in the available studies (i.e. dose spacing and tested concentrations)
Uncertainty 3 *[Oestradiol and histopathology assessment]*	Oestradiol level and gonad histopathology only measured in one study

Overall conclusion on the postulated MoA
The overall biological plausibility is strong and is substantiated by a strong/moderate empirical support. Therefore, the substance meets the ED criteria for non-target organisms The available knowledge, including the AOP 25, is specific for fish. However, extrapolation of this specific MoA to other oviparous vertebrates is biologically supported since the key events are conserved among oviparous (e.g. VTG is an egg yolk precursor protein synthesised in the liver of oviparous vertebrates)