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

Draft for public consultation

Drafted by EFSA and ECHA staff, with support from JRC

7 December 2017

Disclaimer

Applicability and public consultation on this draft guidance document

On 15 June 2016, the European Commission endorsed and published two draft legal acts setting scientific criteria to identify endocrine disruptors under Regulations (EC) No 1107/2009 for plant protection products (PPPs)[1] and (EU) No 528/2012 for biocidal products (BPs)[2].

On 17 October 2016, with a view to ensure a harmonised implementation of the criteria once they become applicable, the Commission mandated the European Food Safety Authority (EFSA) and the European Chemicals Agency (ECHA) to jointly develop - with the support of the Joint Research Centre (JRC) - a guidance document for the implementation of the criteria PPPs and BPs[3]. The original mandate has been complemented on 30/11/2017[4].

The present draft '**Guidance for the identification of endocrine disruptors in the context of Regulations (EU) No 528/2012 and (EC) No 1107/2009**' has been developed for implementing the scientific criteria for the determination of endocrine disrupting properties as included in the draft legal acts endorsed and published by the European Commission on 15 June 2016 and subsequently modified during the negotiations with Member States at the relevant committee or expert group. The draft criteria for PPPs as voted on 4 July 2017 and those adopted for BPs the 4 of September 2017 were equivalent in content.

The criteria to identify endocrine disruptors adopted by the Commission in the context of Regulation (EU) No 528/2012 were published in the Official Journal[5] on 17 November 2017 following no objection by the co-legislators. They enter into force on the 7 of December 2017 and will be applicable from the 7 of June 2018, date when this guidance needs to be available.

The criteria to identify endocrine disruptors in the context of Regulation (EC) No 1107/2009 have been objected by the European Parliament on 4 October 2017 on legal grounds[6] and discussions with Member States on the criteria will be resumed. The Commission considers that the criteria for PPPs should not differ substantially from those adopted for BPs and will prepare a new proposal accordingly following the foreseen procedures[7].

Further, like the criteria to identify endocrine disruptors, the draft guidance document is largely based on the 2002 World Health Organization/International Programme for Chemical Safety (WHO/IPCS) definition of an endocrine disruptor[8], which is generally applicable to all chemical substances. As a consequence, the principles outlined in this draft guidance document may be useful and applicable for the determination of endocrine disrupting properties of any substance, provided that the criteria set for the determination of endocrine disrupting properties under the respective framework applicable to the substance, do not differ substantially from those set in the Commission Delegated Regulation (EU) 2017/2100.

After the public consultation on this draft guidance document, competent scientific bodies consisting of representatives of Member States' competent authorities for biocidal products and, if applicable, the Standing Committee for Plants, Animals, Food and Feed, will be consulted on a revised version of the guidance document, which will address the views expressed during the public consultation and which may also take into account any regulatory developments as regards the criteria to identify endocrine disruptors in the context of Regulation (EC) No 1107/2009.

1 https://ec.europa.eu/health/sites/health/files/endocrine_disruptors/docs/2016_pppcriteria_en.pdf

2 https://ec.europa.eu/health/sites/health/files/endocrine_disruptors/docs/2016_bpcriteria_en.pdf

3 https://ec.europa.eu/health/sites/health/files/endocrine_disruptors/docs/hazardbasedcriteria_mandate_en.pdf

4 https://ec.europa.eu/health/sites/health/files/endocrine_disruptors/docs/hazardbasedcriteria_mandateletter_en.pdf

5 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 301/1.

6 http://www.europarl.europa.eu/sides/getDoc.do?type=TA&reference=P8-TA-2017-0376&format=XML&language=EN

7 https://ec.europa.eu/health/endocrine_disruptors/next_steps_en

8 WHO/IPCS (World Health Organization/International Programme on Chemical Safety), 2002. Global Assessment of the State-of-the-science of Endocrine Disruptors. WHO/PCS/EDC/02.2, 180 pp.

Contents

Abbreviations IV

Glossary of Terms VI

1. Introduction 1

2. Scope of the guidance document 2

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

3.1. General overview of the assessment strategy 5

3.2. Gather all relevant information 10

3.2.1. Sources of the information in the dossier 10

3.2.2. Evaluate the data quality (relevance and reliability) 11

3.2.2.1. Data from standard studies 11

3.2.2.2. Other scientific data 12

3.2.3. Extracting and reporting the information 14

3.3. Assemble and assess lines of evidence for endocrine activity and adversity 15

3.3.1. Assembling the line(s) of evidence for adverse effects 17

3.3.2. Assembling the line(s) of evidence for endocrine activity 17

3.3.3. Assessment of the lines of evidence for adverse effects and endocrine activity. 18

3.3.4. Reporting the lines of evidence 18

3.4. Initial analysis of the evidence 26

3.4.1. Scenarios based on 'EATS-mediated' parameters sufficiently investigated 28

3.4.2. Scenarios based on 'EATS-mediated' parameters not sufficiently investigated 28

3.5. MoA analysis 30

3.5.1. Postulate MoA(s) considering the adversity and/or endocrine activity 31

3.5.2. Establish the biological plausibility for the link between the adverse effect (s) and endocrine activity for the postulated MoA(s) 34

3.5.2.1. Biological plausibility for the key event relationships 35

3.5.2.2. Empirical support for dose–response/incidence and temporal concordance for the key event relationship 35

3.5.2.3. Essentiality, consistency, analogy and specificity of the evidence for the association of the KEs with the adverse effect 37

3.5.2.4. Human relevance 38

3.5.2.5. Relevance at population level for non-target organisms (vertebrates) 38

3.5.2.6. Extent of support for the overall assessment of the biologically plausible link 38
3.5.3. Conclusion on the MoA analysis ... 41
3.6. Overall conclusion on the ED criteria ... 41

4. Information sources for endocrine disruptor identification ... **44**
4.1. Non-test methods ... 49
4.2. In vitro test methods ... 51
4.3. In vivo test methods ... 54
4.3.1. Mammalian ... 54
4.3.1.1. OECD CF level 3 tests ... 54
4.3.1.2. OECD CF level 4 and 5 tests ... 58
4.3.2. Non-mammalian ... 67
4.3.2.1 Parameters ... 67
4.3.2.2 Fish ... 69
4.3.2.2.1 OECD CF level 3 tests ... 70
4.3.2.2.2 OECD CF level 4 and 5 tests ... 71
4.3.2.3 Amphibians ... 76
4.3.2.3.1 OECD CF level 3 tests ... 76
4.3.2.3.2 OECD CF level 4 and 5 tests ... 77
4.3.2.4 Birds ... 79
4.4. Epidemiological data, field studies and population models ... 83
4.4.1. Epidemiological data ... 83
4.4.2. Field studies and monitoring data ... 83
4.4.3. Population models ... 84

5. Recommendations ... **85**
5.1. Recommendations for applicants and assessors ... 85
5.2. Recommendations for future research ... 86

6. References ... **87**

Appendix A – Additional considerations on how to assess the potential for thyroid disruption ... **95**

Appendix B –.Recommendations for design, conduction and technical evaluation of hormonal studies ... **99**

Appendix C – Information requirements for active substances under the Biocidal Products and Plant Protection Products Regulations which could potentially provide information on endocrine-disrupting properties ... **104**

Appendix D – Databases, software tools and literature-derived (Q)SARs 109

Appendix E – . Excel template for reporting the available information relevant for ED assessment 128

Abbreviations

Abbreviation	Explanation
AMA	Amphibian metamorphosis assay
AOP	Adverse outcome pathway
AR	Androgen receptor
BP	Biocidal product
CF	Conceptual framework
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
EC	European Commission
ECHA	European Chemicals Agency
ED	Endocrine disruptor
EFSA	European Food Safety Authority
ER	Estrogen receptor
FLCTT	Fish life cycle toxicity tests (EPA OPPTS 850.1500)
GD	Guidance document
GSI	Gonadal somatic index
HPG	Hypothalamic–pituitary–gonadal
HPT	Hypothalamic–pituitary–thyroid
ICPS	International Programme on Chemical Safety
JMASA	Juvenile Medaka Anti-Androgen Screening Assay
JRC	Joint Research Centre
LABC	Levator ani/bulbocavernosus muscle complex
LAGDA	Larval amphibian growth and development assay
LH	Luteinising hormone
MEOGRT	Medaka extended one-generation reproduction test
MIE	Molecular initiating event
MoA	Mode of action
NR	Nuclear receptor
OECD	Organisation for Economic Co-operation and Development
OPPTS	Office of Prevention, Pesticides and Toxic Substances
PND	Postnatal day
PPAR	Peroxisome proliferator-activated receptor
PPP	Plant protection product

Abbreviation	Explanation
(Q)SAR	(Quantitative) structure–activity relationship
SSC	Secondary sex characteristics
T4	Thyroxine
TG	Test guideline
TH	Thyroid hormone
TSH	Thyroid-stimulating hormone
US EPA	United States Environmental Protection Agency
US FDA	United States Food and Drug Administration
VTG	Vitellogenin
WHO	World Health Organization
WoE	Weight of evidence
XETA	Xenopus embrionic thyroid signalling assay

143

144 # Glossary of Terms

Term	Explanation / Definition
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	Analogy should be interpreted in the context of the MoA framework. Therefore, it should be substantiated by a consistent observation across (related) substances having a well-defined MoA.
Biological plausibility	In the context of this guidance, the biological plausibility focuses on both providing credible support for the link between the adverse effect and the endocrine activity as well biological plausibility for the key event relationships.
Biomarker	A biological characteristic that is objectively measured and evaluated as an indicator of normal biological state or pathological processes
Coherence	Extent to which a hypothesized causal association is compatible with pre-existing theory and knowledge.
Consistency	In this guidance, consistency considers the pattern of effects across species/strains/organs/test systems that would be expected based on the postulated MoA/AOP. In developing a MOA, consistency should also refer to the repeatability of the KEs in the putative MoA in different studies. Consistent observation of the same KE(s) in a number of studies with different study design would increase the support.
Dose concordance	In a MoA/AOP context, dose concordance is verified when the key events are observed at doses 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 an effect on an organism caused by different levels of exposure (or doses) to a stressor (usually a chemical) after a certain exposure duration.
"EATS-mediated" (parameters)	Parameters measured in OECD CF Level 4 and 5 *in vivo* assays and 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' (OECD 2012b, 2012a). These effects are considered potentially adverse effects, while at the same time (due to the nature of the effect and the existing knowledge) they are also considered indicative of an EATS MoA and thus (in the absence of other explanations) imply an underlying *in vivo* mechanistic explanation (e.g. anogenital distance).

Term	Explanation / Definition
Empirical evidence	The information that can be acquired by observation or experimentation by scientists which record and analyse data/information.
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 which can potentially result in an effect on the endocrine system, target organs and tissues.
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 a pathway, signalling process or hormonal mechanism 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 homeostasis, among the various organs of the body. In endocrine signalling, the molecules, i.e. hormones, act on target cells that are distant from their site of synthesis. An endocrine hormone is frequently carried by the blood from its site of release to its target.
Essentiality	Essentiality refers to key events. For determining essentiality it should be demonstrated whether or not downstream KEs and/or the adverse effect is prevented if an upstream event is experimentally blocked. It is assessed, generally, then, 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).
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 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.
Incidence concordance	The incidence concordance is the measure of the frequency of appearance of KE downstream compared to KE upstream. A positive incidence concordance is demonstrated when KE downstream is less frequent than KE upstream.
Line(s) of evidence	A set of relevant information of similar type grouped to assess a hypothesis.

Term	Explanation / Definition
Mechanism of action	A detailed molecular description of the mechanistic interaction through which a substance/molecule produces its effect.
Mode of action (MoA)	Biologically plausible sequence of substance-specific key events, starting with exposure and proceeding through the interaction of the substance or its metabolites with a cell leading to an observed effect supported by robust experimental observations. A mode of action describes a functional or anatomical change at the cellular or biochemical level resulting from the exposure of a living organism to a substance.
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.
Putative MoA	A putative 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 characterized in terms of biological plausibility and empirical support for the KER and essentiality of the KEs.
Relevance	Covers the extent to which data and tests are appropriate for a particular hazard identification or risk characterisation (Klimisch et al., 1997).
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. Reliability of data is closely linked to the reliability of the test method used to generate the data (Klimisch, Andreae, and Tillmann 1997).
'Sensitive to, but not diagnostic of, EATS' (parameters)	Adverse effects which due to the nature of the effect cannot be exclusively attributed to one or more of the EATS modalities. Mechanistic information is required to elucidate whether the effect is mediated by an EATS activity and therefore is a consequence of endocrine disruption. The individual endpoints / parameters may not in themselves be diagnostic of an endocrine disruption modality. Such diagnosis often relies on a combination of endpoints or assays in a weight of evidence assessment.
Specificity	In this guidance specificity should be understood as the extent to which the MoA for the adverse effect is 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 systemic toxicity.
Substance	"Substance" indicates active substances as well as safeners and synergists (for PPPs) and co-formulants (for BPs).
Temporal concordance	The key events are observed in the hypothesized order.
Uncertainty	Uncertainty refers to all types of limitations in the knowledge available to assessors at the time an assessment is conducted

Term	Explanation / Definition
	and within the time and resources agreed for the assessment (EFSA Guidance on Uncertainty in Scientific Assessments).
Weight of evidence (WoE)	Weight of Evidence can be generally described as a stepwise process/approach of collecting evidence and weighing them to reach a conclusion on a particular problem formulation with (pre)defined degree of confidence (EFSA 2017).

1. Introduction

The European Commission (EC) asked the European Food Safety Authority (EFSA) and the European Chemicals Agency (ECHA) to develop a common guidance document for the implementation of the scientific criteria for the determination of endocrine-disrupting properties pursuant to Biocidal products (EU) No 528/2012 (EU 2012) and the Plant Protection Products (EC) No 1107/2009 (EU 2009). The requested technical and scientific assistance is provided for under Article 31 of Regulation (EC) No 178/2002 (EU 2002) laying down the general principles and requirements of food law, establishing the European Food Safety Authority and laying down procedures in matters of food safety.

According to the scientific criteria for the determination of endocrine-disrupting properties (ED criteria) for both BPs (EU 2017a) and PPPs (EU 2017b) there is an obligation to assess active substances as well as safeners and synergists (for PPPs) and co-formulants (for BPs) for their potential ED properties. In this document the term 'substance' is used to address any of these substance categories.

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, i.e. 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. Chapter **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 (EU 2017a). An approach is proposed for analysing the information provided in a dossier submitted for approval of a substance in the context of the PPP or BP Regulations.

Chapter **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, Chapter **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 Chapter **3**). The rationale for grouping effects is 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 2012a) for their interpretation with regard to estrogen, androgen, thyroid and steroidogenic (EATS) modalities and following the Joint Research Centre's (JRC) screening methodology to identify potential endocrine disruptors (JRC 2016).

Chapter **5** gives recommendations for applicants and assessors from evaluating authorities and for future research and Chapter **6** provides the references. The guidance is complemented with a list of 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, conduction and technical evaluation of hormonal studies; **Appendix C** –- Information requirements under the Biocidal Products and Plant Protection Products Regulations; **Appendix D** – Databases, software tools and literature-derived (Q)SARs; **Appendix E** – Excel template for reporting the available information relevant for ED assessment).

2. Scope of the guidance document

This document is intended to provide guidance for applicants and 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 and (EC) 1107/2009 (EU 2017a).

Like the criteria to identify endocrine disruptors, this guidance document is largely based on the WHO/IPCS definition of an endocrine disruptor (WHO/IPCS 2002), which is generically applicable to all chemical substances. As a consequence, the principles outlined in this draft guidance document may be useful and applicable for the determination of endocrine disrupting properties of any substance, provided that the criteria set for the determination of endocrine disrupting properties under the respective framework applicable to the substance, do not differ substantially from those set in the Commission Delegated Regulation (EU) 2017/2100 (EU 2017a).

It should however be noted that the guidance given in this document is limited to the steps necessary to identify a substance as 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 for deciding whether a biocidal active substance identified as endocrine disruptor could be exempted in line with Article 5 (2) (a) from the exclusion from approval in accordance with Article 5 (1) (d) of Regulation (EU) No 528/2012 (EU 2012). Applicants should consider this when determining the needs for generation of further information through experimental testing of animals.

Although the ED criteria cover all endocrine disrupting modes of action, i.e. adverse effects which may be caused by any endocrine modality, this guidance document only addresses the effects caused by estrogen, androgen, thyroid and steroidogenic (EATS) modalities. This is because the EATS modalities are currently the best characterised 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 2012a), which is supported by the 'OECD Conceptual Framework for Testing and Assessment of Endocrine Disrupters' providing a grouping of the studies into five levels according to the kind of information provided (OECD CF; (OECD 2012b, 2012a). OECD GD 150 including the OECD CF is currently undergoing revision and the references made in this guidance to the OECD GD 150 are based on the draft of this document of July 2017 (OECD 2017b). Therefore, when the revised version of the OECD GD 150 is released, additional test guidelines, endpoints and associated guidance given on their interpretation should also be used to support the ED assessment as outlined in this document. However, even though the revised version of the OECD GD 150 includes additional assays related to retinoid, juvenile hormones and ecdysterone modalities, no clear guidance on their interpretation is provided. Consequently, these additional assays currently do not allow any firm conclusions regarding endocrine MoAs.

Nonetheless, with progress of science it is anticipated that the knowledge of how other endocrine modalities, beyond EATS, may lead to adverse effects will become available and should be used to support ED identification. If available, information on non-EATS modalities needs to be considered for the ED assessment.

For similar reasons as for the EATS-modalities, the focus of this guidance is on vertebrate (non-target) organisms, i.e. mammals, fish, amphibians, birds and reptiles as for the vertebrates our current understanding of the endocrine system and availability of test methods is most advanced.

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.

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 light of the criteria applicable for the BP and PPP Regulations (EU 2009, 2012). 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 determining endocrine-disrupting properties for humans are separated from those applicable to non-target organisms; both sets of criteria are further sub-divided into two sections; one section on the identification of an ED and one section on the information to be considered for determination the ED properties.

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

According to the ED criteria (EU 2017a) a substance shall be considered as having endocrine disrupting properties if it meets all of the following criteria:

a) *it shows an adverse effect in an intact organism or its progeny, 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*;
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' since the term 'endocrine mode of action' in point (c) includes both the endocrine activity and a biologically plausible link to an adverse effect.

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

it has an *endocrine activity*, i.e. it *has the capacity to* alter the function(s) of the endocrine system; and

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

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

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:

- Are there endocrine activity and adverse effect(s) relevant for humans which can be biologically plausible linked in an endocrine MoA?
- Are there endocrine activity and adverse effect(s) relevant for non-target organisms which can be biologically plausible linked in an endocrine MoA?

It should be noted that for non-target organisms a substance is considered as having endocrine disrupting properties if the conditions (a), (b) and (c) above are fulfilled, unless there is evidence demonstrating that the adverse effects identified are not relevant at the (sub)population level (for further details on the relevance at the (sub)population level see Section **3.5.2.5**).

From a regulatory point of view, a firm conclusion on whether a substance does or does not meet the ED criteria is always required for substances under the PPP and BP Regulations for both humans and non-target organisms. Therefore, both questions must be answered.

It is recognised that the information needed to conclude on ED properties for humans and non-target organisms may overlap and that there may be information available on non-target vertebrates that can be considered relevant for the ED assessment in relation to humans and *vice versa*.

The second section in the criteria specifies 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.

With regard to weight of evidence, a reference is given to the approach provided in the 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 state that in the weight of evidence assessment the factors listed in **Table 1** shall be considered.

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

- Firstly, weight of evidence 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 weight of evidence 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
- Secondly, weight of evidence is used for the mode of action analysis, to establish the link between the adverse effect(s) and the endocrine activity (see Section **3.5**).

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

Table 1. Factors which must be considered in the weight of evidence assessment

The ED criteria 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 is conducted for endocrine disrupting properties with respect to humans or non-target organisms. Therefore, the factors to be considered are listed separately.	
Factors for humans	***Factors for 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[9]	*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*[9]
	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 biological plausibility of the link between the adverse effects and the endocrine mode of action[9]	*the biological plausibility of the link between the adverse effects and the endocrine mode of action*[9]
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*

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 needs to be collected and assessed. The OECD GD 150 lists tests (test guidelines) and endpoints 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 endocrine disrupting properties measured in the standardised test guidelines.

[9] Should be read as 'endocrine activity' see above

Grouping of parameters relevant for identification of endocrine disrupting properties

Based on OECD GD 150, the JRC screening methodology to identify potential endocrine disruptors (JRC 2016) grouped the parameters into four groups considering that they can provide different types of information towards EATS modalities. In the context of this guidance, this grouping is considered very helpful for guiding the assessors in the evaluation of the scientific evidence. In particular, it gives the key elements for the interpretation of the adverse effects and of the endocrine activity when identifying substances with endocrine disrupting properties. The four groups are:

- ***In vitro* mechanistic** – parameters measured in OECD CF Level 2 *in vitro* assays (i.e. *in vitro* mechanistic information, e.g. estrogenic activity in a transactivation assay). These parameters provide information on the mechanism through which a substance potentially could cause endocrine activity and/or adversity (e.g. by binding to and activating a receptor or interfering with hormone production).
- ***In vivo* mechanistic** – parameters measured in OECD CF Level 3 *in vivo* assays plus hormone levels (also when hormones are measured in OECD CF Level 4 and 5 assays) (e.g. serum hormone levels measured in repeated dose toxicity studies which can provide valuable information on potential interference at the cellular level and, thus, evidence for a potentially adverse effect). These parameters provide information on endocrine activity at a higher biological level (organ, tissue).
- **EATS-mediated** – parameters measured in OECD CF Level 4 and 5 *in vivo* assays and 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' (e.g. anogenital distance). These effects are considered potentially adverse effects, while at the same time (due to the nature of the effect and the existing knowledge) they are also considered indicative of an EATS MoA and thus (in the absence of other explanations) imply an underlying *in vivo* mechanistic explanation.
- **Sensitive to, but not diagnostic of, EATS** – parameters measured in OECD CF Level 4 and 5 *in vivo* assays and labelled in OECD GD 150 as endpoints potentially 'sensitive to, but not diagnostic of, EATS modalities' (e.g. fertility). These effects are considered potentially adverse. However, due to the nature of the effect and the existing knowledge, these effects cannot be considered (exclusively) diagnostic 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.

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, mediation by 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 potentially causes adversity (e.g. by binding to and activating a receptor). In contrast, *in vivo* effects provide information regarding adversity and/or endocrine activity.

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

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 fact that 'EATS-mediated' parameters provide evidence for both endocrine activity and the resulting adverse effects. 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 study (e.g. an extended one-generation reproductive toxicity study; OECD TG 443 (OECD 2012d)). If there is no adversity seen in the 'EATS-mediated' parameters, but adversity is observed 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. Therefore, in the context of this guidance, the 'EATS-mediated' parameters listed in the

OECD GD 150 are considered diagnostic of an endocrine MoA and will therefore drive the assessment strategy. The assessment strategy is applicable both for humans and non-target organisms.

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 as not-target organisms than for other taxonomic groups. Therefore, it is recommended to strive for a conclusion on the ED properties with regard to humans and in parallel, using the same database, strive for a conclusion on mammals as non-target organisms. With regard to non-target organisms, the assessment for mammals should be performed first. If based on this assessment the criteria are not met for mammals as non-target organisms, only then the assessment should proceed to consider the other taxonomic groups, which may require the generation of additional data.

According to the ED criteria all relevant scientific data should be included in the dossier and considered in the assessment. In this context, it should be highlighted that there may be data available on non-target organisms relevant for ED properties with regard to humans and *vice versa.*

For the assessment of ED properties with regard to humans, all relevant data must be considered. The same evidence can be used to conclude for mammals as non-target organisms. However, there may be cases where different conclusions as to whether the ED criteria are met may be reached for humans versus mammals as non-target organisms. For example an adverse effect may be dismissed as not relevant for humans while the same effect is relevant for mammals at the (sub)population level or *vice versa.*

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 test methods and knowledge on how to interpret the results is 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. Although extrapolation of the conclusion based on fish and/or amphibian data to other oviparous species may be, in many cases, scientifically justified, uncertainties may still remain. However the suggested approach is considered sufficient for ED hazard identification with regard to non-target organisms.

Figure 1 illustrates the steps of the assessment. Each of the steps outlined in the figure are described in the following sections. 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, literature data retrieved with systematic literature methodology, and other scientific data. All types of data described in Chapter **4** could be considered, and where relevant, included in the dossier for enabling the assessment of the ED properties. The information is then evaluated for its quality, extracted and reported in the dossier/RAR/DAR. Guidance on how to perform this step is given in Section **3.2.**

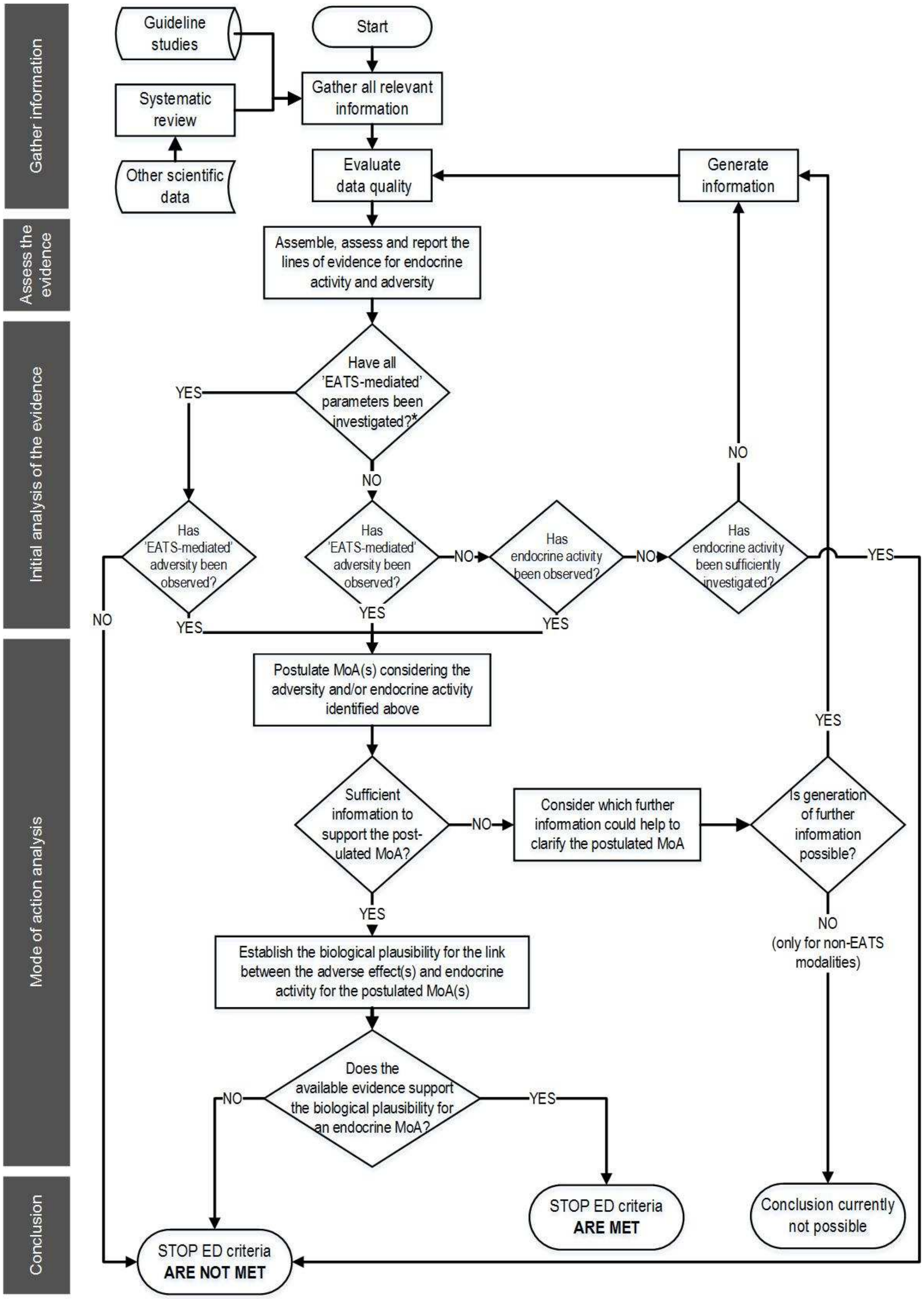
Assess the evidence. In this step the information is assembled into lines of evidence for both adversity and endocrine activity. The lines of evidence are assessed and reported in the dossier/RAR/CAR. Guidance on how to perform this step is given in Section **3.3.**

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 the biologically plausible link between observed adverse effects and endocrine activity. Depending on the available evidence, the assessor and the applicant need to identify the information that must 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 the biologically plausible link between observed adverse effects and endocrine activity is given in Section **3.5.**

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. Different situations are outlined, depending on the outcome of the MoA analysis, see Section **3.6**.

Figure 1. Flowchart illustrating the ED assessment strategy



* For adversity, to have been sufficiently investigated, the ''EATS-mediated'' parameters foreseen to be measured in an Extended one-generation reproductive toxicity study (OECD TG 443; with cohort 1a/1b including the mating of cohort 1b to produce the F2 generation) must be covered. For non-target organisms the corresponding 'EATS-mediated'' parameters are those foreseen to be measured in the Medaka extended one generation test (MEOGRT; OECD TG 240) and the Larval amphibian growth and development assay (LAGDA; OECD TG 241).

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):

(i) scientific data generated in accordance with internationally agreed study protocols […];

(ii) other scientific data selected applying a systematic review methodology […].'

3.2.1. Sources of the information in the dossier

The applicant should consider all relevant scientific data, which provides information on (potential) ED properties, when preparing 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.

Indications of what type of information is to be considered relevant are provided in Chapter **4**.

The standard information requirements for PPPs and BPs include a number of studies that are useful for the ED assessment as requested by the ED criteria. These are listed in Tables **C.1** and **C.2** in Appendix **C –** - according to the current legal frameworks.

According to the data requirements for PPPs and BPs, additional information or specific studies may be required if there is indication that the substance may have ED properties in order to:

- elucidate the mode of action
- provide sufficient evidence for relevant adverse effects.

It should be highlighted that the information requirements of the BP and PPP Regulations may not always provide the information necessary to perform the assessment of the ED properties with regard to humans and/or non-target organisms. Therefore, applicants may need to generate additional information to enable a conclusion. Any suitable source of information reported in Chapter **4** could be considered to provide the additional information necessary. Further details on what types of potential additional data is needed is given in Sections **3.4** and **3.5**.

The literature data should be retrieved in line with the principles of systematic review of literature and reported in the dossier in a transparent manner. Systematic review is a method that aims to systematically identify, evaluate and synthesise evidence for a specific question with the goal of providing an objective and transparent scientific basis for decision making. Systematic reviews promote a more integrated use of the entire body of evidence that is available and relevant for answering a specific question. A crucial and fundamental principle of systematic review is that it is a structured and clearly documented process that promotes objectivity and transparency. There may also be specific mechanistic (non-guideline) investigations conducted by the applicant to support the registration. Although not conducted following "internationally agreed study protocols", such investigations were carried out under GLP and they shall be considered as part of the information extracted from the dossier, after an assessment of their quality according to Section **3.2.2**.

The process of the systematic review reduces bias in the selection of the studies by the extensiveness and reproducibility of the search strategy and the transparent reporting of how studies have been selected and included in the review. The transparent reporting of the search strategy allows an independent judgement to be made on how much of the relevant information has been taken into account.

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 shall be followed (EFSA 2011). These guidances provide instructions on how to identify and select scientific peer-reviewed open literature according to the principles of the systematic literature review, i.e. methodological rigour, transparency and reproducibility. To ensure those fundamental features of the systematic literature search, an *a priori* definition of the review question and the criteria for relevance and reliability should be carried out.

The starting point when conducting a systematic literature search is the design of an appropriate search strategy. Two general search approaches are recommended by (EFSA 2011):

- 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.), including pertinent metabolites and representative formulations.
- A targeted search strategy for individual endpoints. For endocrine disruption, if this option is used, particular attention should be given when designing a proper search strategy in order to avoid bias and capture as much relevant scientific peer-reviewed open literature as possible.

The ED criteria for BPs also require a systematic review, however there is no specific reference to any guidance on how to perform such a review. It is recommended that the EFSA guidances on systematic review are also followed for BPs (EFSA 2010, 2011).

It is recognised that a systematic literature review would identify all 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 (e.g. US EPA ToxCast and OECD QSAR Toolbox), which are highly relevant for the identification of ED properties. If available this kind of information must be assessed for its quality (see Section **3.2.2**).

3.2.2. Evaluate the data quality (relevance and reliability)

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. These terms were defined by Klimisch et al*.* (Klimisch, Andreae, and Tillmann 1997) as follows:

Relevance – covering the extent to which data and tests are appropriate for a particular hazard identification or risk characterisation.

Reliability – evaluating 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. 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' (Chapter R.4 (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).

3.2.2.1. Data from standard studies

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

The relevant standard data for the hazard identification of substances with ED properties are described in Chapter **4** and in Levels 2–5 of the OECD CF (**Table 9**).

In order to comply with the standard information requirements of the PPP and BP Regulations all mandatory studies should be carried out according to the latest version of the corresponding test guideline. This is of particular importance when assessing the ED properties of a substance since in recent years a number of test guidelines have been revised to include additional parameters which are relevant for identification of ED properties. In the case of the two-generation reproduction toxicity study (OECD TG 416 (OECD 2001b), even where the studies have been conducted according to the latest version of the test guideline, 'EATS-mediated' adversity or activity will not have been completely investigated since currently the only mammalian test guideline investigating all the relevant 'EATS-mediated' parameters is OECD TG 443.

It is recognised that the available information on a substance generated according to older versions of guidelines (e.g. the repeated dose 28-day oral toxicity Study in rodents (OECD TG 407 (OECD 2008)); the OECD TG 416 or the combined repeated dose toxicity study with the reproduction/developmental toxicity screening tests (OECD TG 422 (OECD 2016b) may be reliable and relevant for the identification of ED properties. However, they are not fully adequate for the identification of ED properties since they are missing parameters highly relevant for the assessment. 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 should be clearly reported as missing information, and may lead to the need to generate additional information.

Additionally, when assessing the relevance of toxicity studies, effects are considered adequately characterised if doses up to the maximum tolerated dose are used. If evidence of that cannot be provided, other equally appropriate limiting doses include those that achieve saturation of exposure or use the maximum feasible dose. Generally speaking, limit doses of 1,000 mg/kg/day are considered appropriate in all cases where indications of saturation of exposure or limited/no absorption are provided. If none of these criteria can be achieved, a dose of 2,000 mg/kg/day or the maximum feasible dose, whichever is lower, should be considered.

For ecotoxicology, the highest test concentration should be set by the maximum tolerated concentration determined from a range finder or from other toxicity data. The maximum tolerated concentration is defined as the highest test concentration of the chemical which results in less than 10% mortality. For tests on aquatic organisms, the maximum solubility in water, or 10 mg/L for chronic (sub-lethal) tests, could be considered.

Evidence only observed in the presence of excessive toxicity should be assessed. As a general rule, in the absence of a dose-response relationship, hazards suggesting an endocrine-mediated effect which is only evident in the presence of systemic excessive toxicity should not be considered as linked to a primary endocrine MoA. In such a case, justification on excessive toxicity should be provided.

When evaluating the standard studies, the 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.

3.2.2.2. 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 not selected by the systematic review.

Elements to be considered when using systematic review

According to the EFSA guidance on submission of scientific peer-reviewed open literature for the approval of pesticide active substances (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 papers 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 relevance and reliability.

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, biological plausibility of results, consistency among substances with similar attributes and effects, etc. For many data requirements, standardised protocols exist and therefore a reasonable approach for evaluation would be to apply validity and quality criteria that are included in the most relevant test guidelines. The methodological quality of studies may alternatively be assessed by applying other criteria on how to classify the studies according to their reliability for use in risk assessments. Compliance with good laboratory practice standards is, however, not to be considered as a reliability criterion.

Non-guideline studies

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

The scientific validity and reliability of a (Q)SAR model is evaluated following the five OECD principles for validation of (Q)SAR models (OECD 2007e). 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 productiveness; and it is related to mechanistic interpretation. In particular, 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 MoA, examples and references that can be used to assess the reliability of the prediction. Additional 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). 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).

Epidemiological data

No framework has been established on how to assess epidemiological information in the regulatory process. In particular, none of the classical criteria used for the evaluation of these studies are included in the current regulatory framework (e.g. study design, use of odds ratios and relative risks, potential confounders, multiple comparisons, assessment of causality).

Multiple studies assessing the association between the use of PPPs and the occurrence of human health adverse effects acknowledge that epidemiological studies suffer from many limitations and large heterogeneity of data and that broad definition of PPPs in the epidemiological studies limited the value of the results, particularly of meta-analyses.

Nevertheless, where a positive association can be observed between PPP exposures and occurrence of potentially endocrine-related effects, this should be considered as relevant and a special effort should be made to assess the reliability of the study (or studies). However, considering the known limitations of the epidemiological studies, negative associations should be taken with caution and they will not dismiss the assessment based on animal test results. Epidemiological outcomes, where available, should be considered a relevant evidence and part of the WoE approach as well as their integration with the experimental toxicological data. EFSA published a scientific opinion on the use of the epidemiological data and a proposal for their integration with experimental data (EFSA 2017).

Field studies, monitoring data and population modelling

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. (EFSA 2009). As regards to evaluation of population modelling, no specific guidance is available. However, a scientific opinion on good modelling practice may give some indications (EFSA 2014) .

In vitro methods

Mechanistic *in vitro* data can potentially provide strong evidence for a relevant biological process, which could provide key information in the assessment, even though only few *in vitro* assays are currently available as an OECD test guideline. Unfortunately, there are currently no broadly accepted frameworks to assess mechanistic *in vitro* data in decision making (NRC 2014; Vandenberg et al. 2016). However, 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. A draft OECD guidance document is available providing more detailed information on the good scientific, technical and quality practices from *in vitro* method development to *in vitro* method implementation for regulatory use (OECD 2017a).

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.** Other relevant sources of information and in **Appendix D –**). Therefore, a case-by-case evaluation of these data can be performed provided that sufficient details are available.

3.2.3. 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 draft assessment, renewal assessment and competent authority reports. The literature review should also be included and transparently reported and evaluated. A summary of the relevant studies retrieved with the literature should be included with an evaluation of their reliability. The applicant should provide summaries of the studies with the dossier. Applicants are strongly recommended to use the OECD harmonised templates[10] when reporting the studies in the summary dossier.

All the parameters which are relevant for the ED assessment, identified in each 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 Error! Reference source not found.). This should also include consideration of general adversity. Additional instructions on the elements (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 and further evaluated. Both data from the

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

mammalian toxicology section and the ecotoxicology section should be tabulated in a single spreadsheet. A screenshot of a part of the Excel data spreadsheet is shown in **Figure 2** as example on how to record the available information.

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

Once all relevant information (e.g. experimental studies, (Q)SAR predictions) has been evaluated as explained in Section **3.2.2**, a WoE approach should be taken to determine whether some of the identified adverse effects are caused by an endocrine modality.

Relevant parameters should be assembled into lines of evidence to determine whether and how they contribute to adverse effects. In parallel, lines of evidence should also be assembled for the assessment of endocrine activity.

A line of evidence is in broad terms a '*set of relevant information grouped to assess a hypothesis*' (EFSA 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.

For the purpose of building lines of evidence, the parameters investigated in the available pieces of evidence are grouped according to their potential to indicate EATS modalities into the groups described in Section **3.1** (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.

The lines of evidence for adverse effects and endocrine activity will be used to postulate putative (endocrine) mode(s) of action and to understand if there is a biologically plausible link between the observed adverse effects and endocrine activity. If available, AOPs could be supportive when assembling line(s) of evidence (see the OECD AOP Knowledge Base (http://aopkb.org/)).

Figure 2. Screenshot of the Excel table provided in Appendix E, showing how to record the available information

														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]
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**
7	Dossier	1958	Chronic toxicity	dog	0; 0.25; 1.25; 2.50; 12.5	mg/kg bw/day	Oral	52	Weeks	Adult														
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			
12	Dossier	1983	Combined Chronic Toxicity/	hamster	0; 0.15; 1.5; 15	mg/kg bw/day	Oral	78	Weeks	Adult														
2	Literature	1985	Repeated Dose 28-Day Oral	mouse	0; 1000; 2000; 4000	mg/kg bw/day	Oral	4	Weeks	Adult												Increase 1000 (hepatoc		
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			
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	
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										
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							
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

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

In the ED criteria, the identification of adverse effects is based on the WHO definition (IPCS/WHO, 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 system and current practices are applicable for deciding whether the observed effects are treatment-related and should be considered adverse. On this basis, 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 judging if the definition of adversity is fulfilled. A substance identified as ED, will by the nature of its endocrine MoA, in many cases display a pattern of effects. In some cases, *in vivo* mechanistic data may contribute to the definition of adversity e.g. hormonal changes linked to a histological finding and/or Level 3 tests using intact (immature) animals might also provide (additional) evidence of adverse effects.

In addition, 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 (e.g. several estrogen sensitive parameters affected in a consistent manner). Therefore, it requires expert judgement to assemble the lines of evidence for adversity. 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.

A line of evidence may consist of a single parameter (e.g. histopathological findings in the testis observed in one or more studies); or a combination of several related parameters (e.g. a combination of thyroid weight and increased incidence of thyroid hyperplasia in studies of different duration; additional information on how to further investigate thyroid concerns is provided in **Appendix A –**). It could also consist of a number of related parameters measured in the same study (e.g. post-implantation loss combined with reduced litter size).

For non-target organisms separate lines of evidence could be assembled for the different species/taxa. In particular, data on fish could be used for assembling lines of evidence for EAS modalities while data on amphibians could be used for assembling lines of evidence for the thyroid modality. 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 EAS modalities and/or the development/growth (hind-limb length, developmental stage, time to metamorphosis, etc.) for the T modality. Data on other taxa (e.g. birds) can, on a case by case basis, be considered as complementary information.

When assembling the line of evidence, any available epidemiological data, field and monitoring studies and ecological population modelling, should be considered. These data can be considered as supportive evidence in the overall WoE for the evaluation of whether an ED is likely to have adverse consequences for humans and/or at the population level. However, they cannot be used to override or dismiss evidence of adversity found in laboratory studies, nor can they replace laboratory studies.

3.3.2. Assembling 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 are potentially adverse effects which due to the nature of the effect and the existing knowledge also provide *in vivo* mechanistic information for at least one EATS modality (as the observed adversity is very likely caused by alteration in one or more of the EATS modalities).

The lines of evidence for endocrine activity could be organised by modality. If data are available, lines of evidence could be organised following the biological level of organisation (cell, tissue, organ).

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

The evaluation of the lines of evidence should be based on the assessment of 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 dataset as a whole.

It is acknowledged that for some endocrine effects, due to the biology of the endocrine system, more complex dose responses (i.e. non-monotonic) may occur. Therefore non-linear dose responses should not by default be dismissed as not supporting the assessment. Nevertheless, though in most of the cases the design of standard *in vivo* toxicity studies (mainly because of the limited number of doses) does not allow to conclude on the presence of a non-monotonic dose-response, evidence of non-monotonicity in *in vitro* studies (where many concentrations can be tested) could provide additional information relevant to supporting the biological plausibility of an endocrine MoA where endocrine-related adversity is observed in Level 4 or 5 studies (EFSA 2017). 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 non-monotonic dose response is considered low. In this context it should be highlighted that a standard toxicity study must detect toxicity in order to be valid (unless tested at the limit dose).

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). 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 justified, for example by the lack of a proper dose spacing and/or differences in study designs.

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.

3.3.4. Reporting the lines of evidence

The lines of evidence should be reported in a tabular format as exemplified in **Table 2** and **Table 3**. More specifically, the lines of evidence should be reported and organised according to their contribution to the assessment. 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 (incl. general systemic toxicity). 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.

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. 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.

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.

Table 2. Example showing how to assemble the lines of evidence for thyroid disruption

	Line of evidence	Parameter	Species	Exposure Weeks	Route of exposure	Dose mg/kg/day	Observed effects	Conclusion	Indicative of
Evidence of endocrine activity	*In silico* prediction	(Q)SAR prediction DEREK					Predicted to Inhibit of iodine transport	Supporting evidence	Thyroid
	In vivo mechanistic	hormonal changes T3, T4	dog	26	oral	13	dose dependent decrease	Sufficient; 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		
	In vivo mechanistic	hind limb length	frog	3	dermal	1.75	dose dependent decrease	Sufficient	Thyroid
		thyroid (histopathology)	frog		dermal	1.75	dose dependent increase		
Evidence of EATS-mediated adversity	EATS mediated parameter	field study	lizard		dermal / dietary	2.5	lizards from exposed locations displayed thyroid follicular lumens with more reabsorption vacuoles than those from reference fields	Supporting; association between exposure and thyroid disruption	Thyroid
	EATS mediated parameter	thyroid (histopathology)	dog	26	oral	13	follicular cell hyperplasia; dose dependent increase		Thyroid

	Line of evidence	Parameter	Species	Exposure Weeks	Route of exposure	Dose mg/kg/day	Observed effects	Conclusion	Indicative of
Evidence of EATS-mediated adversity			hamster	78	oral	15	no effect; highest dose tested 15	Sufficient; observed in 2 species in a dose related manner	
			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 (15) follicular cells hyperplasia/adenoma		
	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	Thyroid
			mouse	78	oral	15	hyperemia; dose dependent increase		
			rat	104	oral	5	Adenoma		
			rat	2 generation	oral	15.64	vacuolated cells		
	EATS mediated parameter	Thyroid	dog	26	oral	13	dose dependent increase		Thyroid

	Line of evidence	Parameter	Species	Exposure Weeks	Route of exposure	Dose mg/kg/day	Observed effects	Conclusion	Indicative of
		(organ weight)	mouse	78	oral	15	dose dependent increase	sufficient; observed in 2 species in a dose related manner	
			rat	4	inhalation	0.32	dose dependent increase		
			rat	104	oral	5	dose dependent increase		
Evidence of general systemic toxicity	General systemic toxicity	Body weight	dog	26	oral	36	decrease (5%)	sufficient; 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%	sufficient; minor effects in relative liver weight in the high dose groups	
			hamster	78	oral	15	no effect; highest dose tested 15		

Line of evidence	Parameter	Species	Exposure Weeks	Route of exposure	Dose mg/kg/day	Observed effects	Conclusion	Indicative of
		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%		
	Kidney weight (relative)	dog	26	oral	36	no effect	Sufficient; 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		

367

368

Table 3. Example showing how to assemble the lines of evidence for aromatase inhibition leading to reproductive dysfunction in fish

	Line of evidence	Parameter	Species/Cell line(s)	Exposure (weeks)	Route of exposure	Doses (mg/L)	Observed effects	Conclusion	Indicative of
Evidence for endocrine activity	in *vitro* mechanistic data	Aromatase activity	H295R				Inhibition	Sufficient	S
			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		
		Androgen receptor binding/activation	Immuno-immobilised human AR				Positive for AR binding		
			Human AR transfected into CHO-K1 cell line (AR activation)				Negative for agonism. Positive for antagonism		
		Estrogen receptor binding/activation	Yeast etrogen screen (activation)				Weak positive for agonism		
			Human ERαor ERβ transfected into CHO cell line				Negative for both agonism and antagonism		
	In *vivo* mechanistic	Hormonal changes:estradiol	*Pimephales promelas*	3	water	0.5	dose dependent decrease	Sufficient. Estradiol decrease observed in a	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)	dose related manner but measured in one study only. 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		
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 follicules; decreased mean ovarian stages scores)	Supportive evidence. The parameter was only measured in one study.	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 fertility. 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		
	General toxicity	Liver histopathology	*Pimephales promelas*	36	water	0.558	Increase nuclear pleomorphism, multi-nucleation, cystic degeneration, necrosis, pigmented macrophages, aggregates and anisocytosis in hepatocytes of males and females:	Insufficient. 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 as explained in Section **3.3**, an analysis of the sufficiency of the dataset with regard to the investigation of either 'EATS-mediated' adversity or EATS-related endocrine activity has to be carried out. According to the current knowledge and available test guidelines, this is the case when all the 'EATS-mediated' parameters foreseen to be investigated by OECD TG 443[11] have indeed been measured and the results included in the dossier. If this is not the case, 'EATS-mediated' adversity may not have been sufficiently investigated and it is not possible to follow this scenario.

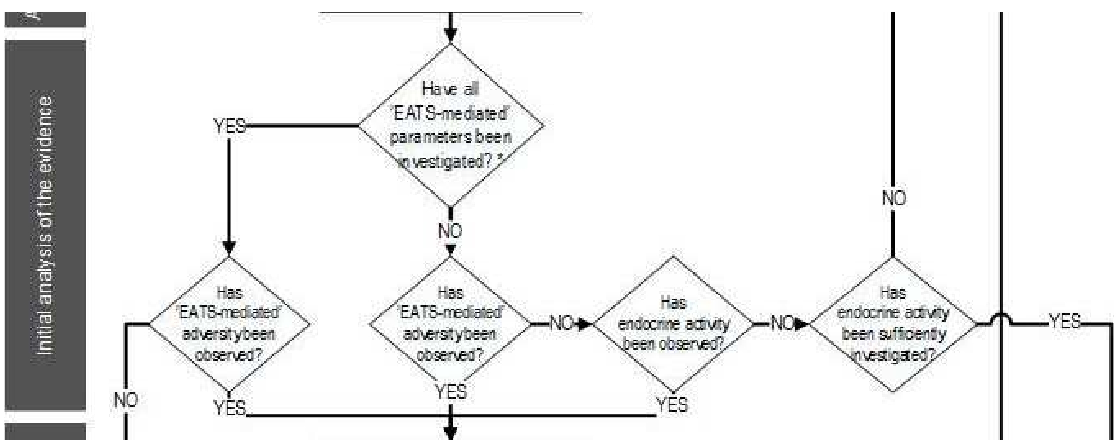
With regard to non-target organisms other than mammals, in order to have all 'EATS-mediated' parameters sufficiently investigated, the 'EATS-mediated' parameters foreseen to be investigated by OECD TG 240 and 241 must have indeed been measured. These two OECD TGs are considered to cover all the EATS modalities in fish and amphibians according to OECD GD 150 and current available test guidelines.

In this section different scenarios providing guidance on how to proceed with the assessment, depending on the information available, are described. A zoom-in of the flowchart presented in Section **3.1** is reported in **Figure 3** and a summary of these scenarios is provided in **Table 4.**

As explained in the assessment strategy (Section **3.1**) it normally should be more efficient to strive for a conclusion on the ED properties with regard to humans and in parallel, using the same database, strive for a conclusion on mammals as non-target organisms; and finally, consider case-by-case, if further assessment is needed to conclude on non-target organisms other than mammals. If the ED criteria are not met for mammals as non-target organisms, only then the assessment should proceed to consider the other taxonomic groups.

Therefore, the scenarios outlined in this section are generic and should be applied in each case as necessary for the assessment of ED properties in relation to humans, mammals as non-target organisms, and non-target organisms other than mammals.

Figure 3. Zoom in on the initial analysis of the evidence from the flowchart in **Figure 1**



[11] i.e. the 'EATS-mediated' parameters investigated in a OECD TG 443 including cohorts 1a and 1b; the extension of the cohort 1b to produce then F2-generation.

Table 4. Overview of the assessment scenarios

The table contains a high level summary of the scenario-specific next steps in the assessment; the scenario descriptions in Sections **3.4.1** and **3.4.2** should be read for full understanding.

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), Available information may be sufficient to conclude on potential for ED properties.
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), Available information may be sufficient to conclude on potential for ED properties.

3.4.1. 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 showed in **Figure 3** "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** (paras 1 and 2) with regard to humans and non-target organisms.

Two scenarios can be foreseen:

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

When no adversity based on 'EATS-mediated' parameters is observed, then it is not possible to perform a MoA analysis because of lack of adversity (i.e. the first condition of the ED criteria is not met). Under these conditions it is possible to conclude that **the substance does not meet the ED criteria with regard to humans**. The same conclusion can be drawn for mammals as non-target organisms.

However, in order to conclude that the ED criteria are not met for other non-target organisms, the 'EATS-mediated' parameters considered by OECD TG 240 and 241 must have been investigated and found negative. If this is the case, it is possible to conclude that **the substance does not meet the ED criteria for non-target organisms.**

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, a MoA analysis is required to establish the biological plausibility of the link between the 'EATS-mediated' adversity and endocrine activity.

This scenario is applicable for the assessment with regard to humans and non-target organisms.

3.4.2. 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 3** "Have all 'EATS-mediated' parameters been investigated?" is NO.

These scenarios cover the cases where the dataset does not include all of the 'EATS-mediated' parameters considered by OECD TG 443 or, in the case of non-target organisms other than mammals, all of the 'EATS-mediated' parameters covered by OECD TGs 240 and 241 (e.g. when a FLCTT study is provided in the dossier). In these situations, adversity based on parameters labelled as 'sensitive to, but not diagnostic of, EATS' parameters cannot be dismissed as not endocrine-related because the 'EATS-mediated' parameters have not been sufficiently investigated.

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 investigated

If the incomplete set of investigated 'EATS-mediated' parameters does not indicate adversity or only information on 'sensitive to, but not diagnostic of, EATS' parameters is available (either indicating or not indicating adversity), as a minimum, endocrine activity must be further investigated.

Three sub-scenarios can be distinguished in this case, depending whether endocrine activity has been observed, or not observed, or not sufficiently investigated:

i) Endocrine activity observed

If the available/generated mechanistic information gives indication of endocrine activity, a MoA analysis is required to establish the biological plausibility of the link between the observed endocrine activity and adverse effect for the postulated MoA(s) (see Section **3.5**). If endocrine activity is observed in *in vitro* mechanistic tests (i.e. level 2) then this would be sufficient as a starting point for the MoA analysis. In **Table 5** the recommended minimum *in vitro* testing battery is reported. As not all 'EATS-mediated' parameters have been investigated, additional information on adversity may need to be generated to enable MoA analysis.

This scenario is applicable for the assessment with regard to humans, mammals as non-target organisms and non-target organisms other than mammals. 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.

ii) No endocrine activity observed, but sufficiently investigated

If the available/generated mechanistic information does not give indication of endocrine activity, it is necessary to check whether endocrine activity for all EATS modalities has been sufficiently investigated. To sufficiently cover the EATS modalities with regard to endocrine activity the level 3 tests: Amphibian Metamorphosis Assay (OECD TG 231, (OECD 2009c); Uterotrophic Bioassay in Rodents (OECD TG 440; (OECD 2007d); and Hershberger Bioassay in Rats (OECD TG 441; (OECD 2009d) must have been conducted; for additional guidance see Chapter **4**. If this is the case and no endocrine activity is observed, then it is not possible to postulate an endocrine MoA, and it can be concluded that **the substance does not meet the ED criteria for humans and non-target organisms.**

The recommended dataset for endocrine activity on mammals and amphibians, as listed in the paragraph above, is generally considered sufficient to cover other non-target organisms, unless information is available indicating a higher sensitivity. These differences should be followed up on a case by case basis e.g. by performing level 3 tests on fish, in order to reach a firm conclusion on non-target organisms.

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, it is needed to generate further information using level 2 and/or level 3 assays (for additional guidance see Chapter **4**) to fully investigate the endocrine activity. If all assays in the level 2 testing battery are negative, this is <u>not</u> sufficient to demonstrate lack of endocrine activity *in vivo* (due to the complexity of the endocrine system and the limitations of the *in vitro* assays). Level 3 assays OECD TG 440 and 441 should be conducted. Special consideration should be given to the thyroid pathway. If the information available from the data set on mammals allows to conclude that the thyroidal endocrine system was not affected, this may be considered as an indication that thyroidal adverse effects in other vertebrate non-target organisms (i.e. amphibians) are unlikely and thus further testing may not be necessary. If such a conclusion cannot be drawn, amphibian testing (i.e. OECD TG 231) should be considered.

Alternatively, on a case-by-case basis, it may be considered more efficient to generate the missing 'EATS-mediated' parameters to enable MoA analysis.

Depending on the outcome of these tests, 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, a MoA analysis is required to establish the biological plausibility of the link between the 'EATS-mediated' adversity and endocrine activity.

This scenario is applicable for the assessment with regard to humans and non-target organisms.

Table 5. Recommended set of *in vitro* testing battery (or equivalents)

Pathway	Assay family	OECD guideline*	EPA guideline	EU method
Estrogen	Transactivation assay	OECD TG 455	OPPTS 890.1300	
Androgen	Transactivation assay	OECD TG 458		
Steroidogenesis	Steroidogenesis	OECD TG 456	OPPTS 890.1550	EU B.57
Steroidogenesis	CYP19		OPPTS 890.1200	

Currently available assays address activity on estrogenic, anti-estrogenic, androgenic, anti-androgenic and steroidogenic modalities.

To limit the number of assays to be conducted, a minimal set could exclude the ER and AR binding assays in favour of the ER (OECD 2012e; US EPA 2009c) and AR (OECD 2016c) transactivation assays. The latter provide information not only on receptor binding potential but also on receptor activation (agonistic) (to elicit a genomic response, requiring the successful interaction with cofactors needed for transcription) or inhibition (antagonistic) as well as the ability of the compound to be taken up by the cell.

In addition, this minimal set should include the H295R cell-based assay (OECD 2011c; US EPA 2009e) investigating the interference with enzymes involved in the synthesis of estrogen and testosterone as well as a specific assay investigating inhibition of aromatase (CYP19), an enzyme involved in the conversion of testosterone to estrogen. The latter assay, although not an OECD TG, is recognised as a US EPA guideline study (US EPA 2009e).

It is noted that there are no *in vitro* assays focusing on thyroid disruption currently available as OECD TGs at Level 2 of the OECD CF. In the absence of suitable in vitro methods, concerns relating to thyroid disruption need to be followed up in vivo (see **Appendix A –**).

3.5. MoA analysis

When adverse effects and/or endocrine activity are identified, the MoA analysis is necessary to demonstrate the biologically plausible link between the two. As described in Section **3.5**, a MoA analysis is required in the scenarios 1b (adversity observed based on 'EATS-mediated' parameters, sufficiently investigated), 2a(i) (no adversity observed based on 'EATS-mediated' parameters, but endocrine activity observed) and 2b (adversity observed based on 'EATS-mediated' parameters, not sufficiently investigated).

Figure 4 illustrates the necessary steps, which are explained below.

The first step of the MoA analysis is to postulate MoA(s) (see Section **3.5.1**).

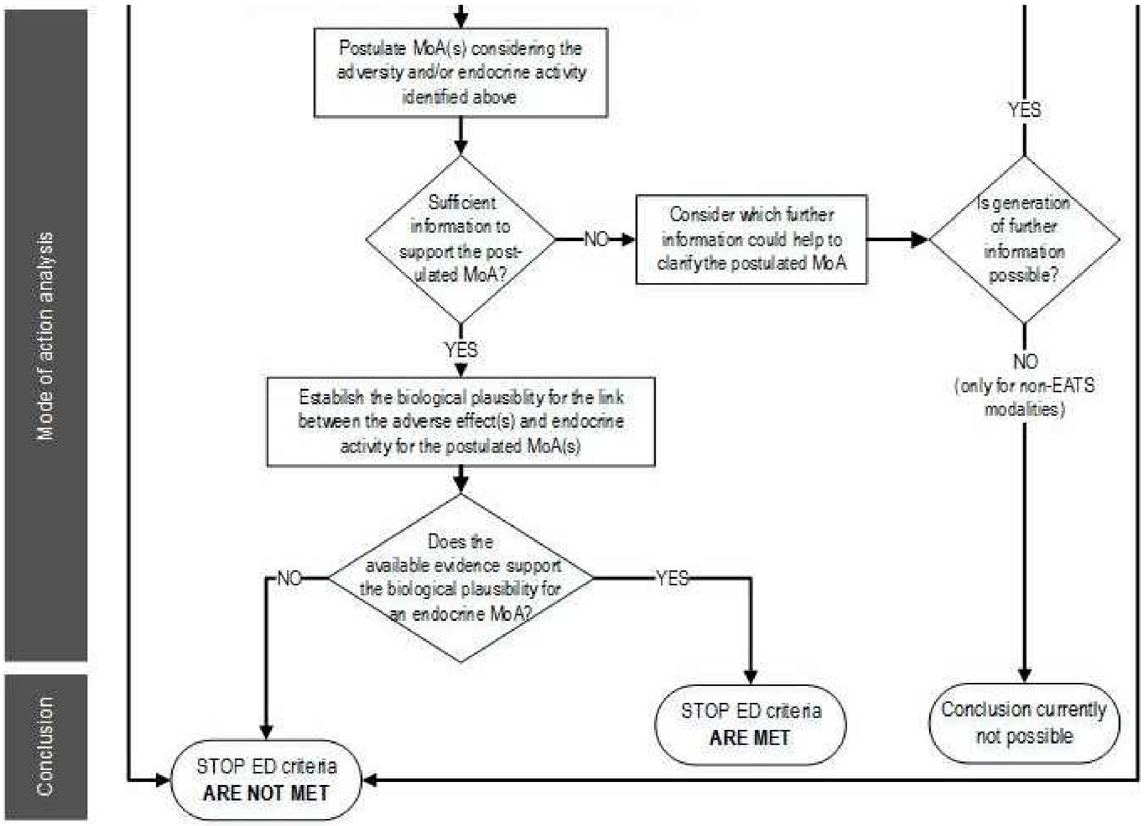
Then it needs to be considered whether the available information on lines of evidence is sufficient to postulate MoA(s).

a) If the available information is sufficient to support the postulated MoA, then it is possible to assess whether there is a biologically plausible link between endocrine activity and the observed adverse effect(s) and subsequently conclude whether the ED criteria are met (see Section **3.5.2**).
b) If the available information is not sufficient to support the postulated MoA, further information is needed to demonstrate the postulated MoA(s).

It is noted that when entering in the MoA analysis with adversity observed based on 'EATS-mediated' parameters, likely further data are not necessary. The 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 4. Zoom in on MoA analysis and conclusion steps from the flowchart in **Figure 1**



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

When adverse effects and/or endocrine activity are identified, the MoA analysis is necessary to demonstrate the biologically plausible link between the two. For this purpose, one or more hypotheses for putative MoA(s) could be developed, covering the observed adverse effect(s) and/or endocrine activity that have triggered the assessment.

A MoA can be described as a series of biological events (i.e. key events (KE)) that result in the specific adverse effect. In the case of endocrine disruption, this sequence at least includes one endocrine mediated KE.

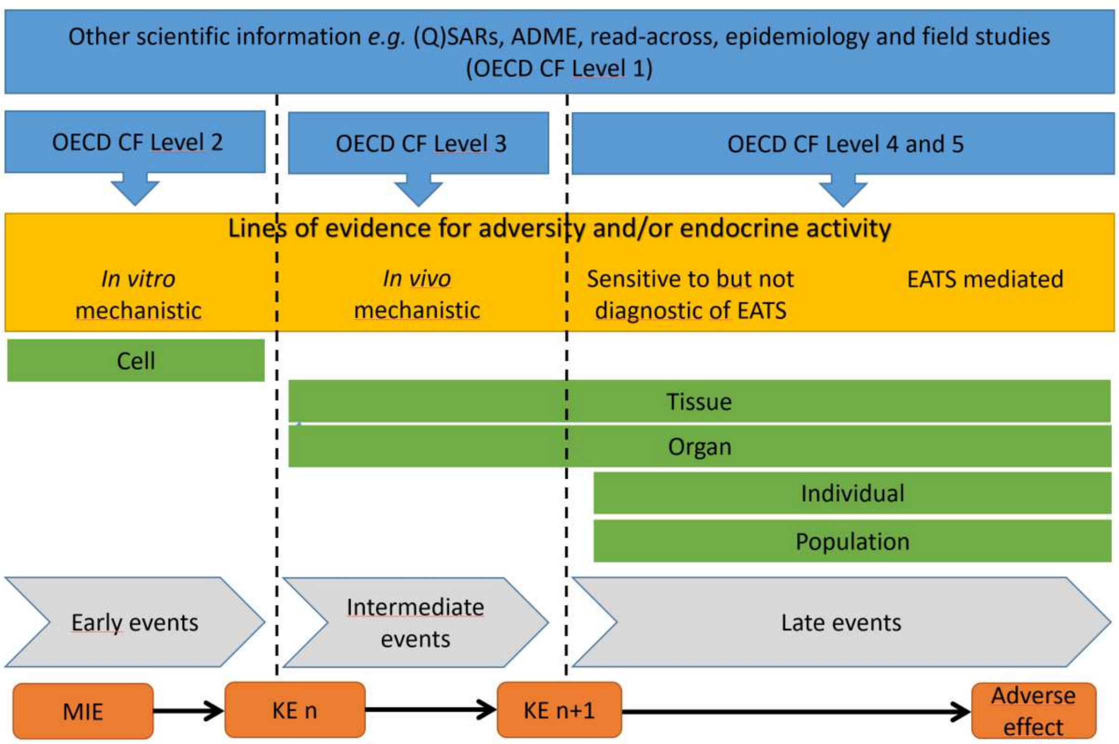
KEs are those events that are considered essential to the induction of the (eco)toxicological response as hypothesised in the postulated MoA. They are empirically observable and measurable steps and can be placed at different levels of the biological organisation (at cell, tissue, organ, individual or population level, see **Figure 5**). 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.

It is not possible to indicate *a priori* how many KEs would be needed to construct a MoA. The level of detail and certainty to support the postulated MoA will depend on the type of information available at the time of the assessment. The postulated 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).

However, there may be situations where the earlier KEs are not needed for the conclusion because of the nature of the adverse effects and the broad knowledge is sufficient to conclude on the biologically plausible link. Indeed, when adversity is indicated by 'EATS-mediated' parameters, the toxicological and endocrinological knowledge may be considered sufficient to conclude on the overall biologically plausible link between the adverse effect and the endocrine activity. A justification should be provided that the observed adverse effect is coherent with broadly accepted pre-existing theory and knowledge (Susser 1991) and that at least one putative endocrine mediated MoA can be described. In this case it is however still necessary to postulate an endocrine MoA and the OECD GD 150 should be applied to link the more likely endocrine pathway resulting in the observed adverse effect.

From the available information assembled into lines of evidence, there will be indications that suggest whether the substance acts via one or more of the EATS 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 5**). Subsequently, the KEs in the putative 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**).

Figure 5. Scheme illustrating how the available information can be organised into lines of evidence to support the postulated mode of action. The arrows linking KEs represent the KE relationships



KE: key event; MIE: molecular initiating event.

Although it might be assumed that endocrine active chemicals will have a single, highly specific mode of endocrine action, this is sometimes not the case. The potential of a substance to elicit different MoAs can obviously lead to difficulties in the interpretation of assay data. If there are indications that a substance may act via multiple MoAs (endocrine or non-endocrine), 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 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.

If an alternative non-endocrine MoA is postulated, it must be properly substantiated. It is however recommended that putative MoA for the endocrine pathways linked to the adverse effect, as proposed in OECD GD 150, would be postulated and duly investigated to fully discharge endocrine mediated MoA.

Table 6. Example of table summarising the key events

[Summary of the hypothesis] The molecular initiating event is unknown, however, the substance increases serum estradiol 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 estradiol	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)

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

If the available information is not sufficient to support the postulated MoA, further information is needed to demonstrate the postulated MoA(s). In principle, any suitable source of information reported in Chapter **4** could be considered to generate the specific additional information necessary.

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 cannot be reached, further data must be generated by the applicant. For example, where contradictory data exist, alternative endocrine and/or a non-endocrine mediated MoA should be postulated and substantiated with empirical data.

In some cases, only evidence on endocrine activity may be available (i.e. scenario 2a(i). In this case, it is very unlikely that any MoA can be postulated; it should therefore be considered which additional information (i.e in vivo level 3, 4 or 5 studies) would be needed to postulate it. 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), it may be necessary to further investigate adversity, therefore *in vivo* Level 3, 4 or 5 studies are expected to be conducted. If no adversity is observed, this would support the lack of an endocrine MoA; if adversity is observed the endocrine MoA would be further substantiated. Targeted mechanistic studies (e.g. Level 2 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.

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 2012c) or the 21-day Fish Assay (OECD TG 230; (OECD 2009b) may be sufficient to further elucidate the intermediate KEs (e.g. estradiol level and VTG).

3.5.2. Establish the biological plausibility for the link between the adverse effect (s) and endocrine activity for the postulated MoA(s)

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) MoA and human relevancy framework (Boobis et al. 2006; Boobis et al. 2008; Meek, Palermo, et al. 2014) provide a methodology for analysing and transparently laying out the evidence for the association of the MoA of a chemical with specific adverse effects. The methodology is applicable to the assessment of any MoA including endocrine-disrupting MoAs. The OECD AOP activity (OECD 2016d, 2017d) also provides a structured framework to integrate the evidence. This framework lays out the sequential progression of KEs from an MIE to the adverse outcome of either human or ecotoxicological relevance. KEs are those that are essential to the progression of the response as hypothesised in the AOP. KEs are connected one to another and this linkage is termed a key event relationship (KER).

In these scientific frameworks the level of evidence required to support the sequence of events leading to adversity might be considered too high a requirement for the hazard identification of an ED for regulatory purposes (JRC 2013). To conclude on the biological plausibility of the link, it may not be necessary to establish the whole sequence and relationship of events leading to the adverse effect. The 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. It is also recognised that the hazard-based identification of endocrine properties is conducted on a case-by-case basis and the amount of evidence needed to establish a biologically plausible relationship will be case-specific. According to the OECD CF and OECD GD 150, 'EATS-mediated' parameters are associated with endocrine MoAs, thus a very high level of understanding will be required to demonstrate that the adverse effect is related to an alternative non-endocrine MoA.

The approach outlined in the IPCS MoA framework has been modified in this guidance to address additional considerations which are necessary for ED assessment.

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

In the context of this guidance, biological plausibility is used in two slightly different contexts: firstly the overall biological plausibility which links the adverse effect and the endocrine activity (in line with the criteria) and secondly the biologically plausible link between two KEs. The primary intent of the biological plausibility for establishing the KER is to provide scientifically credible support for the structural and/or functional relationship between the pair of KEs. Whereas, the overall biological plausibility for an endocrine disrupting MoA, will focus on providing credible support for the link between the adverse effect and the endocrine activity.

Additional elements to support the strength of the putative MoA are analogy, consistency and specificity (see Section **3.5.2.3**). Additionally, human and population relevance needs to be considered (see Sections **3.5.2.4** and **3.5.2.5**).

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 continuum 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).

3.5.2.1. Biological plausibility for the key event relationships

The assessment should consider whether the key event relationship is consistent with what is known about endocrine disruption in general (biological plausibility) and also what is known for the substance specifically.

Biological plausibility. This analysis refers only to the broader knowledge of biology. The putative 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 Chapter **4**, the following questions may be helpful to address this element:

- Is the hypothesis consistent with the broader knowledge of biology?
- Is there a mechanistic relationship between, for example, the KE up and the KE down, 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 WoE for the hazard-based ED identification. 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 key event relationship based on extensive previous documentation and broad acceptance
- Moderate: if the key event relationship 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.

3.5.2.2. 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 2016d).

Table 7. Example of a table which allows analysis of both dose–response and temporal concordance between the key events

[Species X] dose–response and temporal concordance between the key events				
Dose (mg/kg/day)	**KE1** **Increased serum estradiol**	**KE2** **Uterine hypertrophy**	**KE3** **Uterine hyperplasia**	**Adverse effect** **Uterine neoplasia**
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. - indicate no effect; +, ++ and +++ indicate 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 similar or below 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/severity of later KEs 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 (i.e. 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 putative MoA.

Temporal concordance cannot be demonstrated in all cases. In such cases the 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 inconsistent evidence with the expected pattern that can be explained (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.

3.5.2.3. 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 putative 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 experiment conducted in knock out animal 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.

Consistency. This analysis addresses the repeatability of the KEs in the putative MoA in different studies. 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 putative 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?

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 systemic toxicity. The following questions may be helpful to address this element:

- Could the adverse effect be the result of a different MIE (i.e. non-endocrine-mediated)?
- Is the observed adverse effect the result of marked (general) systemic toxicity?

Non-specific, marked systemic toxicity where effects on the endocrine system might be observed along with other toxic effects should not be considered to be the result of an endocrine-disrupting MoA in the absence of any other specific information that might be indicative of a plausible direct endocrine-disrupting MoA.

In the context of this guidance, consistency, analogy and specificity are important elements that support the strength of the MoA. However, they are not specifically weighted as they mainly refer to a single or multiple KE(s) and not to the KER for which the modified Bradford Hill criteria have been applied.

3.5.2.4. 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*'.

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, Palermo, et al. 2014) 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; rather, this guidance is focussing on hazard identification and, as such, priority should be given to 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 should also be excluded.

3.5.2.5. Relevance at population level for non-target organisms (vertebrates)

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 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 ED criteria clarify that relevance at the (sub)population level should be assumed by default in the absence of appropriate scientific data demonstrating non-relevance. Additionally, since the definition of adversity for non-target organisms already considers the (sub)population relevance, the ecotoxicological assessment intrinsically considers impacts at the (sub)population level. With respect to non-target organisms, data on all taxonomic groups, including mammalian data, even if considered not relevant for assessing effects on humans, are in principle considered relevant.

In analogy to human relevance, a substantial amount of information is required to conclude that the observed endocrine-mediated adverse effect is not relevant at the (sub)population level for non-target organisms (vertebrates).

3.5.2.6. Extent of support for the overall assessment of the biologically plausible link

The result of the analysis conducted for the elements in Sections **3.5.2.1**, **3.5.2.2** and **3.5.2.3** should be transparently documented. **Table 8** gives an example of how to report this information.

The assessment of the overall biological plausibility of the link between endocrine activity and adverse effects should 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).

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 establishing the link between the adverse effect and the endocrine activity (Meek, Boobis, et al. 2014; Meek, Palermo, et al. 2014).

It's important to recognize 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, where or 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, Boobis, et al. 2014; Meek, Palermo, et al. 2014). 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).

In all cases, where at least for one KER, the biological plausibility is strong or moderate, the overall biologically plausible link between the adverse effect and endocrine activity should also be considered strong. The resulting weight from the analysis of the empirical support for KERs should be also considered. In absence of dose, temporal and/or incidence concordance, study design(s) should be first re-evaluated for technical correctness. If considered correct, alternative MoA should be considered at this point.

If the overall pattern of evidence leading to the adverse effect is based on 'EATS-mediated' parameters, the toxicology and endocrinology knowledge, is considered sufficient to define the overall biologically plausible link between the adverse effect and the endocrine activity, providing that a justification exists that the observed adverse effect is coherent with broadly accepted pre-existing theory and knowledge (OECD 2012a; Susser 1991) and that at least one putative endocrine mediated MoA can be postulated. Where contradictory data exist, alternative endocrine and/or a non-endocrine mediated MoA should be postulated and substantiated with empirical data.

Table 8. Example summarising the conclusions on the biological plausibility of the link between the adverse effect and the endocrine activity for a postulated mode of action

	Key event relationships (KERs)			
	MIE to KE1	**KE1 to KE2**	**KE2 to KE3**	**KE3 to AE**
Biological plausibility for the KERs	MODERATE It is known that chemically induced inhibition of androgen synthesis can increase the estradiol/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 hypertrophy, hyperplasia and ultimately cancer	See KE1 to KE2	See KE1 to KE2
Empirical support for the KERs	MODERATE – The substance clearly increases serum estradiol 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 (KE2) following hormonal perturbation (KE1) with dose-response and temporal concordance	STRONG – dose/incidence and time concordance is observed for the relationship between KE2 and KE3.	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	**KE1**	**KE2**	**KE3**	**AE**
Essentiality of KEs	No data	MODERATE – There are no stop-recovery studies available. However, based on human clinical experience (see references) an unopposed estrogen action is essential for the tumour development.	See KE1	See KE1	See KE1
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 estradiol is reported for some antifungal agent, but a full MOA was not developed .				

Key event relationships (KERs)	
Specificity	In this case the MIE is unknown, however, the substance clearly increase the levels of estradiol at doses well below those which induce general systemic toxicity.
Identified uncertainties	**Comment**
Uncertainty 1 *[Brief description of the uncertainty]* Lack of a clear understanding of the MIE	Increase in estradiol can be consequent to many MIE.
Uncertainty 2 *[For the empirical support for the KER between the MIE and the KE1, 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 (3 *hormonal assessment only performed for estradiol)*	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 though increased serum estradiol this 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	

3.5.3. 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 putative 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 hypothesised 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 link between endocrine activity and adverse effect is not biologically plausible if the biological plausibility for the KERs is weak and the empirical support is weak.

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:

- Are there endocrine activity and adverse effect(s) relevant for humans which can be biologically plausible linked in an endocrine MoA?
- Are there endocrine activity and adverse effect(s) relevant for non-target organisms which can be biologically plausible linked in an endocrine MoA?

Where no 'EATS-mediated' adversity is observed for a sufficient dataset (scenario **1a,** Section **3.4.1**) or where endocrine activity was fully investigated and found negative for an insufficient dataset (scenario **2a (ii)**, Section **3.4.2**), it is possible to by-pass the MoA analysis and to conclude that the criteria are not met (because an endocrine-related MoA cannot be established if adversity and/or endocrine activity is missing).

In all other scenarios, the conclusion on the ED properties of a substance should be drawn on the basis of the MoA analysis and the biological plausibility of the link between the adverse effects and the endocrine activity.

Where the adversity observed is based on 'EATS-mediated' parameters a MoA analysis is needed to conclude that the ED criteria are met (scenarios **1b**, Section **3.4.1** and **2b**, Section **3.4.2**). In such cases, the MoA analysis is supported by the toxicological and endocrinological knowledge, which is considered sufficient to conclude that an overall biologically plausible link between the 'EATS-mediated' adverse effect and the endocrine activity exists. The conclusion statement should be supported by the scientific justification that the observed 'EATS-mediated' adverse effect is coherent with a broadly accepted pre-existing theory and knowledge.

Where endocrine activity is observed a MoA analysis is required (scenario **2a(i)**, Section **3.4.2**). In this case it may be possible to conclude, based on the observed endocrine activity and existing information on adversity, (e.g. 'sensitive to, but not diagnostic of, EATS' parameters). However, if the available information does not allow to draw a conclusion, additional information on adversity must be generated by exploring the most sensitive endpoints for 'EATS-mediated' adversity (e.g. OECD TG 443). Depending on the results from the additional information on adversity the different corresponding scenarios (i.e. 1a, 1b, or 2b) should be followed. For non-target organisms (e.g. fish) the most common situation might be that adversity is identified on the basis of 'sensitive to, but not diagnostic of, EATS parameters'. 'Sensitive to, but not diagnostic of, EATS' parameters combined with level 2 and level 3 mechanistic information could be sufficient for MoA analysis and to conclude.

Where no 'EATS-mediated' adversity, in an insufficient dataset (scenario **2a (iii)**, Section **3.4.2**), was observed and the endocrine activity was not sufficiently investigated, additional information on 'EATS-mediated' adversity and/or endocrine activity have to be provided. Depending on the results from the additional information on adversity the different corresponding scenarios (i.e. 1a, 1b) should be followed. An alternative to generating additional information on 'EATS-mediated' adversity is to sufficiently investigate the endocrine activity in the EATS modalities (see Section **3.4.2**). If this alternative is followed and the generated information does not show endocrine activity, then a MoA analysis is not possible due to lack of endocrine activity. Consequently, it can be conclude that ED criteria are not met.

If the MoA analysis supports the biological plausibility of the link between the observed adverse effects and endocrine activity for at least one MoA among those postulated, the substance is considered to meet the ED criteria. If the biological plausibility of the link between the endocrine activity and the adverse effect(s) is not demonstrated for any of the postulated MoA(s), the substance is considered not to meet the ED criteria.

Where the available information is sufficient to establish a non-EATS endocrine MoA, in such cases the MoA analysis set out in this guidance should be followed to conclude whether the ED criteria are met.

It is possible that, by entering the MoA analysis, the supporting available information would be not sufficient to conclude on criteria as described above for EATS modalities. A critical 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' MoA is not feasible and that the biological plausibility is highly uncertain. In such cases, conclusion is currently not possible.

In all the cases where data are not provided for performing ED assessment (e.g. for performing a MoA analysis) and this is not considered justifiable, a potential concern would be identified.

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

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

4. Information sources for endocrine disruptor identification

In this chapter, the sources of information that may be used and helpful for the assessment and identification of the endocrine disrupting properties of a substance are described. These information sources comprise non-test methods, in vitro and in vivo test methods, and other information.

OECD Conceptual Framework and OECD GD 150

This chapter is largely based on the 2012 '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 2012a) and the draft of its revision from July 2017 (OECD 2017b). The OECD GD 150 provides widely accepted consensus guidance on the interpretation of effects measured in relevant OECD Test Guidelines (OECD TGs), which may arise as a consequence of perturbations of EATS-modalities, and how these effects might be evaluated to support ED identification.

Annex II of OECD GD 150 provides the OECD Conceptual Framework for Testing and Assessment of Endocrine Disrupters (OECD CF, see **Table 9**). The OECD CF lists the OECD Test Guidelines and standardized 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.

Table 9. OECD conceptual framework (draft 2017)

	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 standardized or non-standardized 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), yeast estrogen screen (ISO 19040-1,2&3) • 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 (See OECD GD No. 211 Describing Non-Guideline In Vitro Test Methods)

	Mammalian Toxicology[3]	Non-Mammalian Toxicology[3]
Level 3 *In vivo* assays providing data about selected endocrine mechanism(s) / pathway(s)[1]	• Uterotrophic assay (OECD TG 440) • Hershberger assay (OECD TG 441)	• Amphibian metamorphosis assay (AMA) (OECD TG 231) • Fish short term reproduction assay (FSTRA) (OECD TG 229)[2] • 21 day fish assay (OECD TG 230) • Androgenized 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 [2]	• Repeated dose 28-day study (OECD TG 407) • Repeated dose 90-day study (OECD TG 408) • Pubertal development and thyroid Function assay in peripubertal male rats (PP male	• Fish sexual development test (FSDT) (OECD TG 234) • Larval amphibian growth & development assay (LAGDA) (OECD TG 241) • Avian reproduction assay (OECD TG 206)

	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) • Subchronic dermal toxicity: 90-day study (OECD TG 411) • Subchronic inhalation toxicity: 90-day study (OECD TG 413) • Repeated dose 90-day oral toxicity study in non-rodents (OECD TG 409)	• 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)[2] • *Potamopyrgus antipodarum* reproduction test (OECD TG 242)[4] • *Lymnaea stagnalis* reproduction test (OECD TG 243)[4] • Chironomid toxicity test (OECD TG 218-219) [4] • Daphnia reproduction test (with male induction) (OECD TG 211) [4] • Earthworm reproduction test (OECD TG 222, 2004) [4] • Enchytraeid reproduction test (OECD TG 220, 2004) [4] • Sediment water lumbriculus toxicity test using spiked sediment (OECD TG 225, 2007) [4] • Predatory mite reproduction test in soil (OECD TG 226, 2008) [4] • Collembolan reproduction test in soil (TG OECD 232, 2009) [4]
Level 5 *In vivo* assays providing more comprehensive data on adverse effects on endocrine relevant endpoints over more extensive parts of the life cycle of the organism [2]	• Extended one-generation reproductive toxicity study (OECD TG 443)[5] • 2-Generation reproduction toxicity study (OECD TG 416 most recent update)	• Fish lifecycle toxicity test (FLCTT) • Medaka extended one-generation reproduction test (MEOGRT) (OECD TG 240) • Avian 2 generation toxicity test in the Japanese quail (ATGT) • Sediment water chironomid life cycle toxicity test (OECD TG 233) [4] • Daphnia multigeneration test for assessment of EDCs (draft OECD TG) [4] • Zebrafish extended one generation reproduction test (ZEOGRT) (draft OECD TG)

[1] Some assays may also provide some evidence of adverse effects.

[2] Some effects can be sensitive to more than one mechanism and may be due to non-ED mechanisms.

[3] Depending on the guideline/protocol used, the fact that a substance may interact with a hormone system in these assays does not necessarily mean that when the substance is used it will cause adverse effects in humans or ecological systems.

[4] At present, these invertebrate assays solely involve apical endpoints which are able to respond to some endocrine disruptors and some non-EDs. Those in Level 4 are partial lifecycle tests, while those in Level 5 are full- or multiple lifecycle tests.

[5] The EOGRT study (OECD TG 443) is preferable for detecting endocrine disruption because it provides an evaluation of a number of endocrine endpoints in the juvenile and adult F1, which are not included in the 2-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. 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. on 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 adverse effect to provide clear answers as to whether a compound has the ability to act via endocrine-mediated modalities. 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 measure apical endpoints that are considered predictive of adverse effects but not necessarily suitable to identify how the effects arise (i.e. by what MoA). Mechanistic data can be retrieved also from CF Level 4 and 5 tests. 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 current version of OECD GD 150, the test methods are grouped in three parts (A, B and C) according to the extent of guidance provided for effects interpretation. The test methods listed under Part A are established test methods which have been in wide use as validated OECD or national test guidelines for which guidance is provided, whereas the test methods listed under Part B have not yet received full validation for endocrine outcomes, or are TGs that are not primarily designed for testing endocrine disruption. Lastly, test methods listed under Part C are those listed in the OECD CF, but for which no guidance is currently provided, either because there is insufficient experience in their use or because they are thought not to offer significant advantages over existing tests. As more ED-relevant test methods are developed into TGs or endocrine parameters added to existing TGs 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 TGs or US Environmental Protection Agency (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 and non-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 **3.2.2**). In general, any non-standard study providing information on relevant EATS-effects (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 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 2012a)).

Finally, it is important to bear in mind while carrying out the ED assessment (Chapter **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

While the primary data sources will be the data generated using standardised test methods and the systematic literature review according to the data requirements of the specific regulatory framework, other sources and types of information to be considered include the following:

- Databases of compiled data (see **Appendix D –**)

- Published literature (see Section **3.2.1**)
- (Q)SAR model outputs (see Section **4.1**)
- Read-across and category approaches (see Section **4.1**)
- Human (epidemiological) data (see Section **4.4.1**)
- Field studies, from controlled field experiments (see Section **4.4.2**)

A general overview of some relevant databases of compiled data (not exhaustive) is given in **Table 10**. More information can be found in **Appendix D –**.

Table 10. Other relevant sources of information

<table>
<tr><td colspan="2">Regulatory documents connected to other EU Regulations beyond the BP and PPP Regulations (e.g. REACH, Cosmetic Product Regulation)</td></tr>
<tr><td rowspan="17">Databases specifically related to endocrine active or endocrine-disrupting properties</td><td>Endocrine active substances information system (EASIS) (EC JRC)</td></tr>
<tr><td>ToxCast (US EPA)</td></tr>
<tr><td>ToxCast ER prediction model (US EPA)</td></tr>
<tr><td>SIN (Substitute it now!) List (International chemical secretariat)</td></tr>
<tr><td>The endocrine disruption exchange (TEDX)</td></tr>
<tr><td>Endocrine disruptor screening program, EDSP21 (US EPA)</td></tr>
<tr><td>Endocrine disruptor knowledge base, EDKB database (US FDA)</td></tr>
<tr><td>Estrogenic activity database, EADB (US FDA)</td></tr>
<tr><td>Toxicology data network (Toxnet) developmental and reproductive toxicology database (DART)</td></tr>
<tr><td>NURSA (nuclear receptor signalling atlas)</td></tr>
<tr><td>OECD (Q)SAR toolbox (OECD, ECHA)</td></tr>
<tr><td>AOP knowledge base (OECD)</td></tr>
<tr><td>ToxRefDB (US EPA)</td></tr>
<tr><td>eChem portal (OECD)</td></tr>
<tr><td>COSMOS database - an EU project developing methods for determining the safety of cosmetic ingredients for humans, without the use of animals, using computational models</td></tr>
<tr><td>Danish (Q)SAR Database</td></tr>
<tr><td>(Q)SAR Data Bank</td></tr>
</table>

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 (absorption, distribution, metabolism and excretion) 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.

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 2013b; JRC 2014; Lo Piparo and Worth 2010).

In silico predictions may be used as a means of generating supporting information for EATS modalities within a WoE approach. In particular, by providing information on the molecular initiating event (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, Edwards, and Roitberg 2008), docking studies (Warren et al. 2006), and 3D-(Q)SARs like the comparative molecular field analysis (CoMFA) (Cramer, Patterson, and Bunce 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, pre-steps 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 mode ls cover different types of receptors (e.g. ER, AR, THR) 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, Dobler, and Smiesko 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 Error! Reference source not found.).

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 (JRC 2014; Wu et al. 2013). For ease of use, profilers are typically implemented in software tools, such as the OECD (Q)SAR Toolbox (Dimitrov et al. 2016; OECD 2014) and the Chemotyper (Yang et al. 2015) (see **Appendix D –**).

Table 11. Software tools for predicting endocrine activity

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; THR = Thyroid hormone receptor.

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

[a] AHR; [b] AHR, glucocorticoid, liver X, mineralocorticoid, peroxisome proliferator-activated receptor γ, enzymes CYP450 3A4 and 2A13; [c] GR, LXR, PPAR, RXR; [d] PPAR, AR, AHR, ER, GR, PR, FXR, LXR, PXR, THR, VDR, RXR. [e] PXR.

Attention should be paid in the interpretation of results to understand the specific basis and scope of the prediction for each ED pathway. 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 2008, 2017c). 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 dataset or when determining the KEs in a putative MoA. Nevertheless, such data cannot be used to conclude that there is no concern for ED properties, although it may be used to trigger further testing.

As an adaptation of the data requirements according to Annex IV, Section 1.5 of the BP Regulation (EU 2012), 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 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 (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 adverse effect 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 selected endocrine pathways fall under Level 2 of the OECD CF for the testing and assessment of ED (OECD 2012b). The assays currently listed in the OECD CF Level 2 are specifically those that detect one particular endocrine modality only, focusing on the estrogenic and 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 disruption 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 suggests the use of 18 different *in vitro* assays (ER binding, dimerization, chromatin binding, transcriptional activation and ER-dependent cell proliferation) to predict agonist/antagonist activity (Browne et al. 2015; Judson et al. 2015), although reanalysis of the data set suggests a limited number of assays might provide the same prediction (Burgoon 2017; Judson et al. 2017). Guidance on the reporting of defined approaches has been developed by OECD (OECD 2017e).

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 (BG1Luc Estrogen Receptor Transactivation Test Method for Identifying Estrogen Receptor Agonists and Antagonists; OECD TG 457; (OECD 2012f). 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.

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, for low affinity chemicals, mammalian-based test systems that focus on human hERα might not effectively predict effects in fish and reptiles (Ankley et al. 2016).

Currently, only a few assays have OECD-adopted TGs, although several relevant assays are under consideration for TG development. It is therefore expected that much of the *in vitro* data will be obtained from the scientific literature and will be from non-TG methods. While preference might be on TG studies, data generated by other relevant *in vitro* assays should always be considered, providing that the principle of the assay is clearly described and that the assays are shown to be robust and reproducible based on available validation data (e.g. by using the criteria set out in the performance-based TGs for transactivation assays or validation principle as addressed in the OECD draft guidance document on good *in vitro* method and practices (GIVIMP;(OECD 2017a)). An OECD guidance document is in place on the reporting of non-standardised *in vitro* assays (i.e. non-test guidelines) (OECD 2017c) 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 OECD TGs that are valid for, and can more easily be extended to encompass, multiple assays. Performance-based TGs are now in place for ER binding assays (OECD TG 493; (OECD 2015e) and ER transactivation assays (OECD TG 455; (OECD 2012e), while a performance-based TG 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**	**US EPA OPPTS 890.1250 / OECD TG 493 *****	**US EPA OPPTS 890.1150**	**OECD TG 458 ****	**US EPA OPPTS 890.1200**	**OECD TG 456 (EU B.57)**
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-EcoScreenTM cell line	Human recombinant microsomes	Human H295R cells
Indicative of:	E	E	A	A	S	S
Androgen receptor binding/transactivation			x	X		
Aromatase					x	
Estrogen receptor binding/transactivation	x	x				
Steroidogenesis (estradiol and/or testosterone synthesis)						x

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.
** This TG was not validated when OECD GD 150 was published. However, in OECD GD 150 a stably transfected human AR transactivation assay (AR STTA) was listed in Section B. This assay subsequently became validated and was named OECD TG 458 (OECD 2016c). Therefore TG 458 is now included in this table.
*** In OECD GD 150 the only available ER binding assay was the US EPA OPPTS 890.1250 (US EPA 2009b). Afterwards, another validation study was conducted and led to OECD TG 493 (OECD 2015e).

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 been drafted. 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 2017a). 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.

When interpreting the results of in vitro tests, the lack of a metabolic system, as well as the other ADME properties, should be considered. In part his is because *in vitro* systems usually consist of (a monolayer) of one cell type that focuses on a specific pathway. In general, they lack the complexity of the combinations of cells *in vivo* and ADME properties. To partly overcome this limitation, several *in vitro* can be run by incorporating (part of the) metabolising systems, as a surrogate to the potential metabolized into an active, less active or inactive substance/metabolite which might explain the apparent discrepancy between in vitro and in vivo results. Activities on including a metabolisation step are currently on the OECD TG program (OECD 2017h).

As mentioned above, while 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 (nor battery) 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 disruption activity on the endocrine modality under investigation. However, consistent negative *in vitro* effects (in multiple systems) can be interpreted as an indication of a lack of endocrine disruption activity for a specific endocrine modality and as such can be used to support a 'ED criteria are not met" conclusion, if it can be substantiated that the compound is available to the test system and does not undergo metabolic activation.

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**, 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 effects 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 it is tabulated in **Table 13**, **Table 14**, **Table 15**, **Table 16** and **Table 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 *in vivo* test methods are described in the level 3 to 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 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 rat 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.

In these assays, animals with minimal endogenous estrogen/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. As such, Level 3 assays are incapable of revealing the full spectrum of possible ED 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 TGs 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 TGs allow for the detection of other endocrine modalities (e.g. disruption of 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 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. 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) 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.

The uterotrophic assay is a short-term assay (3 days), using oral gavage or subcutaneous routes. 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 excretion).

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 2006; Schapaugh et al. 2015).

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 epididymes 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 (T3 and T4), LH, FSH and testosterone. The weanling assay does not include glans penis.

The assay uses immature weanling or castrated peripubertal 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.

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-alpha-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).

The weights of the optional organs (adrenal) provide information not only on androgen modality, but also on systemic toxicity. 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. 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.

The Hershberger assay is a short-term assay (10 days), using oral gavage or subcutaneous injection.

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. The effects are also grouped in the category 'EATS–mediated'.

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

Section A lists parameters from tests for which guidance is provided in OECD GD 150.

		Section A	
Test guideline		OECD TG 440 (Level 3)	OECD TG 441+OECD GD 115 (Level 3)
Test duration		4 days	11 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 #:**		
Adrenals weight*	N		x (optional)
Cowper's glands weight (Hershberger)	A		x
Epididymis weight*	E, A, S		x
Estradiol level	E, A, S		x
FSH level*	E, A, S		x (optional)
Glans penis weight (Hershberger)	A		x
Keratinisation and cornification of vagina (UT assay)	E	x	
LABC weight (Hershberger)*	A		x
LH level*	E, A, S		x (optional)
Proliferation of endometrial epithelium (UT assay)	E	x	
Prostate weight (Hershberger)*	A		x
Seminal vesicles weight (Hershberger)*	A		x
Steroidogenesis (genes/enzyme changes)	E, A, S		x
T3 and T4 level*	T		x
Testis weight*	E, A, S		x
Testosterone level*	E, A, S		x (optional)
Thyroid histopathology (Hershberger)*	A		x
Uterus histopathology (UT assay)*	E	x	
Uterus weight (UT assay)*	E, A	x	
Vaginal opening	E, A	x	

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.

* 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).

*^ 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.

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 sensitive to more than one MoA. 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 5 tests also include parameters indicative of endocrine activity. The list of relevant parameters, based on OECD GD 150 and JRC screening methodology, is shown in Table **14**.

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

The 28-day repeat dose toxicity test (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 for no effect. Dosing should begin as soon as possible after weaning and, in any case, before the animals are nine weeks old.

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 2017f)) exposures

Preferred species: rat

When interpreting the histopathological data of the ovaries (follicular, thecal, and granulosa cells), uterus, cervix and vagina, possible asynchrony of the estrus cycle should be taken into account.

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

The assay has not been validated to detect ED, but it does contain many parameters that are suitable for the determination of 'EATS-mediated' effects and effects 'sensitive to, but not diagnostic of, EATS' modalities, even if some endocrine-sensitive parameters are missing (e.g. thyroid hormones, functional measurement of estrous cyclicity). 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 1998a) is likely to be more sensitive than OECD TG 407.

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 1998b)), subchronic dermal toxicity: 90-day study (OECD TG 411 (OECD 1981b)), and subchronic inhalation toxicity: 90-day study (OECD TG 413 (OECD 2017g)).

Preferred species: rat

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

The prenatal developmental toxicity study (OECD TG 414 (OECD 2001a)) involves repeated dosing of pregnant females and therefore potential exposure of the developing fetus. The revised version of the

TG adopted in 2001 includes more parameters than the previous version, but was not specifically designed to detect ED. 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).

The OECD GD 150 does not provide guidance on the interpretation of some parameters measured in this TG. Therefore the grouping of the parameters has been assigned for the purpose of this guidance.

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.

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 TGs 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 TGs. 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 pre-mating dosing period in males, fertility may not be a particular sensitive indicator of testicular toxicity. Therefore, a detailed histological examination of the testes (i.e. staging) is essential.

Regarding thyroid hormone, measurement of T4 is mandatory in the 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 thyroid-stimulating hormone (TSH) should be measured as 'total'.

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 foetus. It includes some parameters that may detect endocrine disruption (e.g. abnormalities of male and female genitalia).

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.

OECD GD 150 does not provide guidance on the interpretation of some parameters measured in this TG. Therefore the grouping of the parameters has been assigned for the purpose of this guidance.

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 estrous 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.

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 2009d)) or female (OPPTS 890.1450 (US EPA 2009f)) 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 2012a).

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. These tests have been considered to be of low reliability, based on a retrospective analysis of the performance criteria of the assays (Schapaugh et al. 2015).

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

The two-generation reproduction toxicity test (OECD TG 416 (OECD 2001b)) assesses endocrine-related parameters in a less comprehensive way that the other level 5 assay (OECD TG 443 (OECD 2012d)), and although some 'EATS-mediated' parameters like estrous 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 estrus cycles (in order to detect any adverse effects on estrus 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 2012d) has been designed to cover specific life stages not covered by other assays and to test for effects that may occur as a result of pre- and postnatal exposure to chemicals. 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, anogenital distance index at birth, age of vaginal opening and preputial separation. According to the TG, 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 TG. 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 pre-mating period (minimum of two weeks) and a two-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.

Thyroid hormone 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. 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 TG recommends that parameters involving pup mammary glands of both sexes be included, when validated.

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

Vaginal opening and first vaginal estrus 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 estrus, (i.e. pseudo-precocious puberty) which may or may not continue as the animal matures. Thus, evaluating the first vaginal estrus 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 estrus 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 estrus measurements, that body weight can influence these parameters. Another parameter which should be investigated in relation to effect on estrus 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. Further, sex hormones play an important role in development of sexual dimorphism of the brain. Substances interfering with the sex hormonal balance 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 (Adori et al. 2010; Arredouani 2014; Cutolo et al. 2002; Trigunaite, Dimo, and Jorgensen 2015).

Preferred species: rat

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

The table is divided into three 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 that are not primarily designed for detection of endocrine disruption, for which limited guidance is given in OECD GD 150; and Section C lists parameters from tests listed in the OECD CF but for which no guidance is currently provided in OECD GD 150 because there is insufficient experience in their use

		Section A						**Section B**					**Section C**	
Test guideline		**OECD TG 407 (Level 4)**	**OECD TG 415 (Level 4)**	**OECD TG 416 (Level 5)**	**OECD TG 443 (Level 5)**	**US EPA OPPTS 890.1500 (Level 4)**	**US EPA OPPTS 890.1450 (Level 4)**	**OECD TG 408 (Level 4)**	**OECD TG 451-3 (Level 4)**	**OECD TG 421 (Level 4)**	**OECD TG 422 (Level 4)**	**Adult Male Assay (Level 4)**	**OECD TG 414 (Level 4)**	**OECD TG 426 (Level 4)**
Test duration		28 days (plus 14 days recovery period)	16–19 weeks	29 weeks	30 weeks	30 days	20 days	90 days	between 12 and 18 months in mouse or 24 in rat	11 weeks	11 weeks	15 days	from implantation to the day prior to the scheduled caesarean section (days 5-15 in rodent, 6-18 in rabbits)	from GD 6 to PND 21
Life stages		adult (P)	adult (P) and F1	adult (P), F1 and F2	adult (P), F1 and eventually also F2	juvenile male	juvenile female	adult (P)	adult (P)	adult (P) and F1	adult (P) and F1	adult (P)	fetus	fetus and F1
Species / *in vitro* test system		rat	mouse, rat	mouse, rat	rat	Rat	rat	rat	mouse, rat	rat	rat	rat	rat, rabbit	rat
Parameter name	**Indicative of[#]:**													

Accessory sex glands weight	E, A, S											x		
Accessory sex organs histopathology	E, A, S							x	x	x				
Age at first estrus	E, A						x							
Age at balano-preputial 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	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			
Coagulating gland weight	E, A, S	x		X	x	x				x	x	x		
Colloid area (thyroid histopathology)	T	x				x	x				x (optional)			
Cowper's gland weight										x (optional)	x (optional)			
Epididymis histopathology	E, A, S	x	x (optional)	X	x	x		x	x	x	x	x		
Epididymis weight*	E, A, S	x		X	x	x		x	x	x	x	x		
Estradiol level												x		
Estrus cyclicity	E, A, S	X Optional (at necropsy by vaginal smears)		X	x		x			x	x			
Follicle stimulating hormone (FSH) level*	E, A, S											x		
Follicular cell height (thyroid histopathology)	T	x		X		x	x				x			
Glans penis weight										x (optional)	x (optional)			
Genital abnormalities	E, A, S		x	X	x					x	x			
LABC weight*	E, A, S					x				x (optional)	x (optional)			
Luteinising hormone (LH) level *	E, A, S											x		
Mammary gland histopathology (male)	E, A, S	x (optional)			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	x		x	x	x	x	x			
Ovary weight	E, A, S	x (paired) (optional)		X	x		x	x	x	x	x			
Oviduct histopathology	E, A, S		optional		x									
Prolactin level												x		
Prostate histopathology (with seminal vesicles and coagulating glands)	E, A, S	x	X (optional)	X	x			x		x	x			
Prostate weight*	E, A, S	x		X	x	x			x	x	x	x		
Seminal vesicles histopathology	E, A, S	x	X (optional)	X	x						x			
Seminal vesicles weight*	E, A, S	x		X	x	x				x	x	x		
Sperm morphology	E, A, S			X	x									
Sperm motility	E, A, S			X	x									
Sperm numbers	E, A, S			X	x									
T3 and/or T4 level*	T	x (optional)			x	x	x			x	x	x		
Testis histopathology	E, A, S	x	X (optional)	X	x	x		x	x	x	x	x		
Testis weight*	E, A, S	x		X	x	x		x	x	x	x	x		
Testosterone/Dihydrotestosterone level*	E, A, S					x						x		
Thyroid histopathology*	T	x		X	x	x	x	x	x	X (optional)	X (optional)	x		
Thyroid-stimulating hormone level (TSH)	T	x (optional)			x	x	x			x	x	x		
Thyroid weight	T	x (optional)		x	x	x	x		x	x (optional)	x (optional)	x		
Uterus histopathology (with cervix)*	E, A, S	x	X (optional)	X	x		x	x	x	X (optional)	x			
Uterus weight (with cervix)*	E, A, S	X (optional)	x	X	x		x	x	x	x	x		x † (gravid uterus)	

Vagina histopathology	E, A, S	x	X (optional)	X	x			x	x		x			
Vaginal smears	E, A, S	x (optional)		X	x					x	X			
Adrenals histopathology	N	x			x			x	x		x			
Adrenals weight*	N	x		X	x	x	x	x	x		X			
Brain weight	N	x		X	x			x	x		x			x
Dystocia	N		x	X	x					x				
Fertility	N			X	x					x	x			
Fetal development (or physical development of the foetuses?)	N		x							x	x		x †	x
Gestation length	N		x	X	x					x	x			
Litter size	N		x	X	x					x	x			x †
Litter viability	N		x	x	x					x	x			
Litter/pup weight	N		x	X	x					x	x		x †	
Number of implantations, corpora lutea	N			X	x					x	x		x †	

#: 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.

*: 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).

†: when these parameters are measured in OECD TG 414 and/or 426 the OECD GD 150 does not provide guidance on their interpretation. Therefore, the interpretation shown in this table and in the corresponding text has been assigned by the authors of this guidance document.

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, gonado-somatic index) 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 normal 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, 2012c). However, abnormal behaviour or appearance could also be clinical signs of general toxicity, or due to other MoAs. Therefore, interpretation of such behaviours needs to be linked to other effects in order to ascertain if they are linked to an endocrine activity or even adverse effects.

Other parameters, such as vitellogenin and spiggin production, secondary sexual characteristic, sex ratio, and gonad or thyroid histopathology can inform on 'EATS-mediated' effects and are detailed below.

Vitellogenin

Vitellogenin (VTG) is normally produced by the liver as a precursor of yolk proteins in female fish, amphibian and bird under estrogenic regulation (Slater, Redeuilh, and Beato 1991). VTG is not produced by male under natural condition, and 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 TGs.

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 (EFSA 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 an OECD guidance document as a screening test for androgen antagonism (OECD GD 148 (OECD 2011a)).

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 yet 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 according to (OECD 2009b) and expert judgement. 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 an endpoint reflecting sex reversal, and can in principle be affected by oestrogens, anti-oestrogens, 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 amphibian 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 can also play a role in the sex determination and the sex ratio, which should be taken into account when interpreting the results (Ospina-Alvarez and Piferrer 2008), although this should not be an issue when testing under controlled laboratory condition.

It is acknowledged that sex ratio is an apical endpoint relevant at the population level that is 'EATS-mediated'. Sex ratio is also relevant for amphibians and birds.

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. Moreover, GSI endpoint can be impacted secondarily through the cortisol-mediated stress response pathway as it has been observed that female Mozambique tilapia (*Oreochromis mossambicus*) implanted with cortisol to simulate chronic stress had reduced oocyte size and 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, Collette, and Facey 1997), and that inter-individual variation in ovarian weight can be high during the spawning cycle (OECD 2004). GSI is therefore a highly variable measure in fish and should be interpreted with caution. 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, 2015b) 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), 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, 2012a).

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 endpoint 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, 2015b), 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. Thyroid histopathology can also be carried out on bird, for which guidance is given in OCSPP 890.2100 (US EPA 2009a). Potential changes should be evaluated in: 1) overall thyroid size; 2) the overall size and shape of follicles; 3) the overall size and relative number of thyroid follicular epithelial cells; and 4) the relative quantity and quality of colloid.

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 pre-hatch 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). 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).

4.3.2.2.1 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 MoA. These are the fish short-term reproduction assay (OECD TG 229 (OECD 2012c)), the 21-day fish assay (OECD TG 230 (OECD 2009b)) and its variant the androgenised female stickleback screen (OECD GD 148 (OECD 2011a)). It should be noted that all three fish tests primarily give information on potential endocrine MoAs in adult fish, although some of those test can also give information on relevant adverse effect (e.g. fecundity in combination with VTG and possibly SSC). Test conditions and measured parameters are briefly described below and summarised in **Table 15**. In addition, two 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, and the juvenile medaka anti-androgen screening assay (JMASA).

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

In the OECD TG 229 fish short-term reproduction assay (OECD 2012c) sexually mature male and spawning female fish are exposed to a chemical for 21 days. Two 'EATS-mediated' parameters are measured in both males and females: VTG and SSC. Induction of plasma VTG levels in male fish serves 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 endpoint. Gonad histopathology can be evaluated to assess the reproductive fitness of the test animals and to add to the WoE of other endpoints 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 screening Level 3 test for endocrine MoA, it can also show ED-mediated adverse effects, which implies that the combined effects might be sufficient in some cases to reach a conclusion without additional testing. 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 especially this sensitive window.

Validated species: Fathead minnow (*Pimephales promelas*); Japanese medaka (*Oryzias latipes),* partially validated for the zebrafish (*Danio rerio;* 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 TG, 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 'EATS-mediated' endpoint 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-hour 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 hours 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 aromatizable 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.

Species: cyp19a1bGFP zebrafish (*Danio rerio*).

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

This test, currently under validation at the OECD, is designed to identify androgen antagonists and chemicals interfering with androgen biosynthesis.

The assay is based on male juvenile medaka (*Oryzias latipes*), which develop 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 male 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*).

4.3.2.2.2 OECD CF level 4 and 5 tests

There are three *in vivo* tests guidelines for identification of endocrine adverse effects in fish at the level 4 and 5 of the OECD CF: the medaka extended one-generation reproduction test or MEOGRT (OECD TG 240 (OECD 2015c)) at level 5, the fish life cycle toxicity test (US EPA OPPTS 850.1500 (US EPA 2009d), which has not been validated) at level 5, and the fish sexual development test (OECD TG 234 (OECD 2011b)) at Level 4. The list of relevant parameters that give indications on the ED properties, based on OECD GD 150 and JRC screening methodology, is shown in **Table 15**. 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), 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).

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 2011b), 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 endpoint. 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 on EDs identification and MoA. 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 to sex ratio changes caused by a specific endocrine mechanism than other. In sticklebacks, the validation data available so far showed that on this species alterations of phenotypic sex ratio by the test substances were uncommon (OECD TG 234). Therefore, absence of observed changed in sex ratio in stickleback would not be sufficient to disregard a substance's endocrine potential in fish and in general, this species should not be used for conducting a new study. An effect on sex ratio in TG 234 shows that the test chemical causes an adverse apical effect, is a developmental toxicant, and is probably also an ED, in absence of general systemic toxicity (OECD GD 150).

Measurements of VTG and sex ratio can in combination demonstrate the endocrine MoA, more particularly estrogenic, androgenic and aromatase inhibition; and to a lesser extent the effects of estrogen and androgen antagonists can also be seen (OECD TG 234). As an example, a low level of VTG can also be expressed in males; therefore, depending on the analytical detection limit (LOD), a decrease in males can also be observed. However, given the low biological significance of such an observation at the population level, it can only be informative on MoA and should always be combined with other data (i.e. sex ratio and change of VTG in females) for interpretation. The combined measurement of VTG and sex ratio also give, in the same test, information on both mechanism and adverse effect relevant at the population level. Additionally, gonadal histopathology is an optional 'EATS-mediated' endpoint; body length and weight should be measured and survival, hatching success, abnormal behaviour and morphological abnormalities should be monitored.

It has to be noted that, as this test does not cover the reproductive life stage of the fish, chemicals that are suspected to affect reproduction should be examined in a test that covers it.

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 measures potential adverse effects on population-relevant parameters, including survival, gross development, hatching, time to spawn and reproduction. Additionally, observations of behaviour and morphological abnormalities should be made daily.

Moreover, if there is evidence for a chemical having potential endocrine-disrupting activity (e.g. androgenic or estrogenic activity in other tests and assays) other useful information is obtained by measuring mechanistic parameters such as hepatic VTG mRNA or VTG protein, phenotypic SSC such as characteristic male anal fin papillae as related to genetic sex, and evaluating kidney, liver and gonad histopathology. The Japonese medaka is the appropriate species for use in this test guideline, because of the possibility to determine its genetic sex. This is based on the presence or absence of the medaka male sex-determining gene *dmy*. Such mechanistic parameters can assist in determining whether any effect is endocrine-mediated or is linked to systemic and other toxicity and to help better understanding any responses. Therefore, they must be interpreted in relation to non-endocrine-specific parameters and population-relevant parameters.

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 endpoints would be the same, taking into account the biological differences between the species (e.g. the absence of validated SSC in zebrafish). Ultimately, the choice of the species should depend on the endpoint-related sensitivity of each test species and species-specific characteristics.

Validated species: Japanese medaka (*Oryzias latipes*)

Fish life cycle toxicity tests (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 never been validated. Therefore, when new testing is necessary, a test carried out according to a validated OECD test guideline would be preferred. 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). It does not include endpoints specific to a particular EATS modality, but they can be added. Limited data are obtained from the F1 generation in the test. Of particular interest in the context of estrogens, androgens and steroidogenesis disruptors 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 on endpoints that might be sensitive to, but not diagnostic of, endocrine disruption 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.

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)

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.

		Section A					Section B
Test guideline		**OECD TG 229 (Level 3)**	**OECD TG 230 (Level 3)**	**OECD TG 240 (Level 5)**	**OECD TG 234 (Level 4)**	**US EPA OPPTS 850.1500** (Level 5)**	**OECD GD 148 Androgenised female stickleback screen (Level 3)**
Test duration		21 days	21 days	133 days	60 days post-hatch	100-190 days	21 days
Life stages		Sexually mature male and spawning female (F0)	Sexually mature male and spawning female (F0)	From sexually mature males and females of F0 to hatching of the F2	From newly fertilised egg until completion of sexual differentiation (F0)	Freshly fertilised eggs of F0 to juvenile stage of F1	Sexually mature female (F0)
Species		Fathead minnow, Japanese medaka, zebrafish	Fathead minnow, Japanese medaka, zebrafish	Medaka; can be adapted to zebrafish (ZEOGRT, under validation)	Japanese medaka, three-spined stickleback, zebrafish, fathead minnow (partially validated)	Fathead minnow or sheepshead minnow (marine). Can be adapted to medaka and zebrafish	Stickleback
Parameter name	**Indicative of #:**	**OECD TG 229**	**OECD TG 230**	**OECD TG 240**	**OECD TG 234**	**US EPA OPPTS 850.1500****	**Androgenised female stickleback screen (GD 148)**
Male SSC in females	E, A, S	X	X	X	X[a]		
Male SSC in males	E, A, S	X	X	X	X[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

Testosterone level	E, A, S			X[b]			
Estradiol level	E, A, S			X[b]			
Specific gonad histopathology*	E, A, S	X		X	X		
Sex ratio (female biased)	E, A			X	X	X	
Sex ratio (male biased)	E, A, S			X	X	X	
Behaviour	N	X	X	X	X	X	X
Length	N			X	X	X	
Morphological abnormalities	N	X	X	X	X		X
Gonado-somatic index	N			X			
Embryo time to hatch	N			X			
Reproduction (fecundity, fertility)	N	X		X		X	
Survival	N	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	
Body weight	N			X	X	X	X

[#] Based on draft OECD GD 150 of July 2017 (OECD 2017b), indicative of: the (E)strogen-; (A)ndrogen-; (S)teroidogenesis-; or (T)hyroid- modalities; (N)ot assignable to a specific modality.
* Histological examination of the gonads should enable identification of intersex (presence of testis-ova) and undifferentiated fish; detailed guidance on specific gonad histopathology examination in fish is given in (OECD 2010).
** No endpoints specific to a particular EATS modality are included at present but they could be added if validated.
[a] When medaka is the test species.
[b] Hormone measurements are not mentioned in the TG240 but are mentioned in the OECD GD 150 as endpoints of this TG.

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 endpoints 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.

4.3.2.3.1 OECD CF level 3 tests

Amphibian metamorphosis assay (OEC TG 231; OPPTS 891100, CF Level 3)

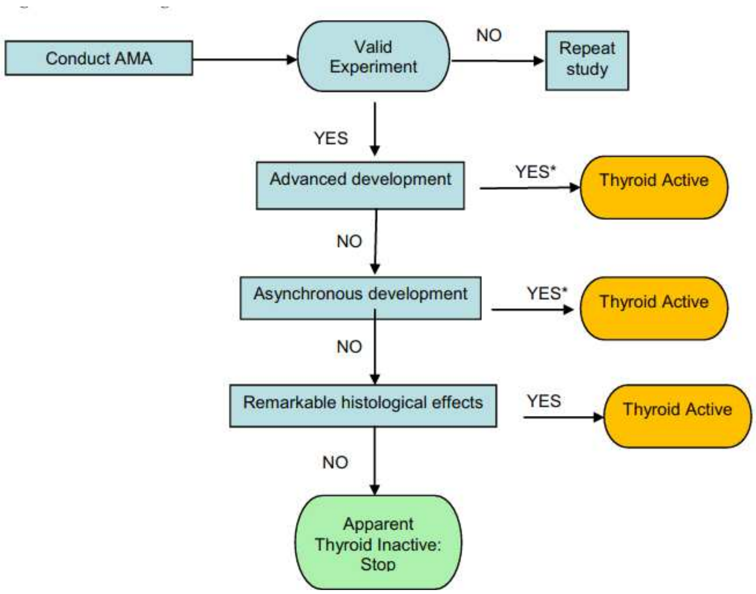
The AMA was developed to identify substances affecting the function of the HPT axis in vertebrates. The test is conducted with larval stages (tadpoles) of *Xenopus laevis* exposed for 21 days. The developmental stage, hind limb length, snout to vent length measurement and wet weight are the apical endpoints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

The apical endpoints hind-limb length and 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 generalized 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 hind-limb 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 6**.

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

Figure 6. Decision tree for evaluating thyroidal effects in the AMA (from OECD TG 231 (OECD 2009c)).



*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 endpoints are required.

Xenopus embryonic thyroid signalling assay XETA (CF level 3)

This 72-hour *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 a Green fluorescent protein fluorescence in the tadpoles, each translucent tadpole expressing a basal fluorescence. In contact with a thyroid disruptor, the green fluorescent protein is down- or up-regulated, which allows the chemical effect on the thyroid system to be assessed.

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

4.3.2.3.2 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 **4.3.2.2.1** for details). In addition, the LAGDA allows the detection of endocrine effects on the development of the reproductive system, and emphasis is given to population-relevant endpoints (i.e. mortality, development, growth and reproductive development).

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 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 endpoints mediated by endocrine activity on the HPG axis, as well as the endpoint histopathology of gonads and reproductive ducts. Change in levels of VTG provide information about a substance interfering with the sex hormone system (E, A, S) (optional).

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

Histopathology examination of the liver (i.e. decreased glycogen vacuolation) and kidneys (i.e. mineralisation and tubule dilation) can indicate effects not diagnostic of EATS (OECD 2015b). The potential relationship between the histological changes observed and the treatment on the one hand, and a potential endocrine disruption effect on the other hand should be considered on a case-by-case basis based on a WoE approach (OECD 2015a).

In addition, mortality, abnormal behaviour and growth endpoint (length and weight) 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.

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	
Test guideline		**OECD TG 231 (Level 3)**	**OECD TG 241 (Level 4)**
Test duration		21 days	16 weeks
Life stages		Tadpole NF (NF 51)	Embryo, tadpoles, early juvenile
Species		*Xenopus laevis*	*Xenopus laevis*
Parameter name	**Indicative of #:**	**OECD TG 231**	**OECD TG 241**
Hind-limb length	T	X	
Developmental stage	T	X	
Plasma level of VTG	E, A, S		X
Thyroid histopathology (amphibian)*	T	X	X
Histopathology (gonad, reproductive ducts)*	E, A		X
Sex ratio (phenotypic (gonad histology), genetic)	E, A		X
Time to metamorphosis (NF stage 62)	T		X
Body weight	N	X	X
Snout-vent length/Growth	N	X	X
Malformations	N	X	X
Mortality	N	X	X
Behaviour	N	X	X
Histopathology (liver, kidney)*	N		X
Liver weight (liver somatic index;)	N		X
.			

#: 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.
* 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.

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. 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, the undertaken validation process initiated by OECD could not go to its end, and 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 toxicity test (OECD TG 206, CF Level 4)

The avian reproduction toxicity test (OECD TG 206 (OECD 1984)) gives a list of endocrine-sensitive parameters 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 (Berg et al. 2004; De Wit 2006; Lundholm 1997). It is noted that OECD TG 206 recommends gross pathology examinations, although further details on this assessment are not reported. 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, subadults 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. The 14-day-old survivors per F1 generation hen, representing the second generation (F2), is the primary biological endpoint of 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, it has to be noted before to conduct this test that it was considered insufficient according to OECD standards and could not be validated, and that its use has considerable animal welfare implications.

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)

The table is divided into two sections: Section A lists parameters from tests for which guidance is provided in OECD GD 150; Section Blists 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

		Section A	Section B
Test guideline		**OECD TG 206 (Level 4)**	**US EPA OCSPP 890.2100 ** (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 #:**	**OECD TG 206**	**US EPA OCSPP 890.2100 ** **
Estradiol, testosterone and thyroid hormone levels measurements (egg yolk, adult, thyroid hormone from thyroid gland)	E,A,T		X
Histopathology (thyroid gland, gonad)*	E,A,T		X
Sex ratio of chicks	E,A		X
Secondary sexual characteristic (Plumage)	E, A		X
Gross pathology	N	X	X
Hatchability	N	X	X
Egg fertility (ED 8)	N		X
Eggshell thickness	N	X	X
Eggshell strength (Newton)	N		X
Egg viability (% viable embryo of egg set)	N	X	
Embryo viability (ED 15)			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)*	N		X

Based on the draft OECD GD 150 of July 2017 (OECD 2017b), indicative of: the (E)strogen-; (A)ndrogen-; (S)teroidogenesis-; or (T)hyroid- modalities; (N)ot assignable to a specific modality.
* 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, reprodcutive tissues should be examined in the next highest until indications of overt toxicity are not observed.
** This TG is not validated by OECD.
ǂ Embryonic day

4.4. Epidemiological data, field studies and population models

4.4.1. Epidemiological data

According to Regulation (EU) No 283/2013 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' (EU 2013). Likewise, in the BP Regulation concerning the making available on the market and use of BPs (EU 2012), 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 (EU 2009) , 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 2013a). 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 (EFSA 2013b), 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. Target 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 (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. For instance, a key question which could be addressed by such models is the degree of reproductive impairment which is likely to trigger consequences at the population level. Because the data needs are so great across so many compounds and so many taxa, development of population modelling may be a possible practical approach to determine whether adverse effects at population level are likely (Marty et al. 2017). The advantage of modelling is that different environmental situations can be simulated and extrapolation in time is possible. It is, however, noted that at present such models 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). Moreover, a large amount of data is needed to build a substance-specific model. Although there is currently no generally accepted models and no common agreement on which endpoints need to be included, 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 2014). Therefore, while the mentioned tools might provide supportive information to be integrated in a WoE approach, they currently cannot be used to dismiss the population relevance of an adverse effect in a hazard assessment context.

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 to be the most sensitive indicator of chemically induced effects. In the context of this guidance, 'detailed histopathological examination' 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 (XETA, EASZY and JMASA) with fish and amphibians are currently in the validation process (see Sections **4.3.2.2.1** and **4.3.2.3.1**). It is recommended that those three are performed together with the *in vitro* battery, once fully validated. This will reduce the uncertainty linked to the extrapolation of mechanistic information from mammalian to other vertebrate species.

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.

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[12], a number of AOPs exist for 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 should be explored.

Considering the current knowledge in fish endocrinology and the availability of standard test methodologies, further investigations are recommended into the possibility of including additional parameters related to modalities other than EAS in the existing test guidelines.

Further exploration of the possibility of including measurements of thyroidal hormones in the OECD TG 231 and 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.

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

12 https://aopwiki.org/

6. References

Adori, M., E. Kiss, Z. Barad, K. Barabas, E. Kiszely, A. Schneider, D. Kovesdi, E. Sziksz, I. M. Abraham, J. Matko, and G. Sarmay. 2010. 'Estrogen augments the T cell-dependent but not the T-independent immune response', *Cell Mol Life Sci*, 67: 1661-74.

Andersson, C., I. Katsiadaki, K. Lundstedt-Enkel, and J. Orberg. 2007. 'Effects of 17alpha-ethynylestradiol on EROD activity, spiggin and vitellogenin in three-spined stickleback (Gasterosteus aculeatus)', *Aquat Toxicol*, 83: 33-42.

Ankley, G. T., C. A. LaLone, L. E. Gray, D. L. Villeneuve, and M. W. Hornung. 2016. 'Evaluation of the scientific underpinnings for identifying estrogenic chemicals in nonmammalian taxa using mammalian test systems', *Environ Toxicol Chem*, 35: 2806-16.

Arredouani, M. S. 2014. 'New insights into androgenic immune regulation', *Oncoimmunology*, 3.

Becker, R. A., G. T. Ankley, S. W. Edwards, S. W. Kennedy, I. Linkov, B. Meek, M. Sachana, H. Segner, B. Van Der Burg, D. L. Villeneuve, H. Watanabe, and T. S. Barton-Maclaren. 2015. 'Increasing Scientific Confidence in Adverse Outcome Pathways: Application of Tailored Bradford-Hill Considerations for Evaluating Weight of Evidence', *Regul Toxicol Pharmacol*, 72: 514-37.

Benigni, R., C. L. Battistelli, C. Bossa, A. Giuliani, and O. Tcheremenskaia. 2017. 'Endocrine Disruptors: Data-based survey of in vivo tests, predictive models and the Adverse Outcome Pathway', *Regul Toxicol Pharmacol*, 86: 18-24.

Berg, C., A. Blomqvist, L. Holm, I. Brandt, B. Brunstrom, and Y. Ridderstrale. 2004. 'Embryonic exposure to oestrogen causes eggshell thinning and altered shell gland carbonic anhydrase expression in the domestic hen', *Reproduction*, 128: 455-61.

Boobis, A. R., S. M. Cohen, V. Dellarco, D. McGregor, M. E. Meek, C. Vickers, D. Willcocks, and W. Farland. 2006. 'IPCS framework for analyzing the relevance of a cancer mode of action for humans', *Crit Rev Toxicol*, 36: 781-92.

Boobis, A. R., J. E. Doe, B. Heinrich-Hirsch, M. E. Meek, S. Munn, M. Ruchirawat, J. Schlatter, J. Seed, and C. Vickers. 2008. 'IPCS framework for analyzing the relevance of a noncancer mode of action for humans', *Crit Rev Toxicol*, 38: 87-96.

Browne, P., R. S. Judson, W. M. Casey, N. C. Kleinstreuer, and R. S. Thomas. 2015. 'Screening Chemicals for Estrogen Receptor Bioactivity Using a Computational Model', *Environ Sci Technol*, 49: 8804-14.

Burgoon, Lyle D. 2017. 'Autoencoder Predicting Estrogenic Chemical Substances (APECS): An improved approach for screening potentially estrogenic chemicals using in vitro assays and deep learning', *Computational Toxicology*, 2: 45-49.

Caslin, T. M., and J. O. Wolff. 1999. 'Individual and demographic responses of the gray-tailed vole to vinclozolin', *Environmental Toxicology and Chemistry*, 18: 1529-33.

Cramer, R. D., D. E. Patterson, and J. D. Bunce. 1988. 'Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins', *J Am Chem Soc*, 110: 5959-67.

Creasy, D. M. 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: 408-15.

Cronin, M. T. D., and A. P. Worth. 2008. '(Q)SARs for predicting effects relating to reproductive toxicity', *Qsar & Combinatorial Science*, 27: 91-100.

Cutolo, M., B. Seriolo, B. Villaggio, C. Pizzorni, C. Craviotto, and A. Sulli. 2002. 'Androgens and estrogens modulate the immune and inflammatory responses in rheumatoid arthritis', *Neuroendocrine Immune Basis of the Rheumatic Diseases Ii, Proceedings*, 966: 131-42.

Dang, Z. 2016. 'Interpretation of fish biomarker data for identification, classification, risk assessment and testing of endocrine disrupting chemicals', *Environment International*, 92-93: 422-41.

De Wit, CA. 2006. 'Effects of endocrine disrupters in wild birds and mammals.' in T Grotmol, A Bernhoft, GS Eriksen and TP Flaten (eds.), *Endocrine Disrupters* (The Norwegian Academy of Science and Letters: Oslo).

Dimitrov, S. D., R. Diderich, T. Sobanski, T. S. Pavlov, G. V. Chankov, A. S. Chapkanov, Y. H. Karakolev, S. G. Temelkov, R. A. Vasilev, K. D. Gerova, C. D. Kuseva, N. D. Todorova, A. M. Mehmed, M. Rasenberg, and O. G. Mekenyan. 2016. 'QSAR Toolbox - workflow and major functionalities', *SAR QSAR Environ Res*: 1-17.

ECHA. 2008. "Guidance for the implementation of REACH, Chapter R.6: QSARs and grouping of chemicals." In, 134. Helsinki: ECHA (European Chemicals Agency).

———. 2011. "Guidance for the implementation of REACH, Chapter R.4: Evaluation of available information." In, 23. Helsinki: ECHA (European Chemicals Agency).

———. 2017a. "Guidance on the Biocidal Products Regulation: Volume III Human Health, Assessment + Evaluation (Parts B+C)." In, 361. Helsinki: ECHA (European Chemicals Agency).

———. 2017b. "Guidance on the Biocidal Products Regulation: Volume IV Environment, Assessment & Evaluation (Parts B+C) " In, 416. Helsinki: ECHA (European Chemicals Agency).

———. 2017c. "Read-Across Assessment Framework (RAAF)." In, 60. Helsinki: ECHA (European Chemicals Agency).

Edwards, S. W., Y. M. Tan, D. L. Villeneuve, M. E. Meek, and C. A. McQueen. 2016. 'Adverse Outcome Pathways-Organizing Toxicological Information to Improve Decision Making', *J Pharmacol Exp Ther*, 356: 170-81.

EEA. 2012. "The Impacts of Endocrine Disrupters on Wildlife, People and their Environments – The Weybridge+15 (1996–2011) Report." In, 116. Copenhagen: EEA.

EFSA. 2006. 'Temperature-dependent sex determination in fish revisited: prevalence, a single sex ratio response pattern, and possible effects of climate change.' in S Ostlund-Nilsson, I Mayer and F Anne (eds.), *Biology of the three spined stickleback* (CRC Press: Huntingford).

———. 2009. 'Guidance on Risk Assessment for Birds and Mammals on request from EFSA', *EFSA Journal*, 7: 1438.

———. 2010. 'Application of systematic review methodology to food and feed safety assessments to support decision making', *EFSA Journal*, 8: 1637.

———. 2011. 'Submission of scientific peer-reviewed open literature for the approval of pesticide active substances under Regulation (EC) No 1107/2009', *EFSA Journal*, 9: 2092.

———. 2013a. 'Guidance on tiered risk assessment for edge-of-field surface water', *EFSA Journal*, 11: 268.

———. 2013b. '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*, 11: 3132.

———. 2014. 'Scientific Opinion on good modelling practice in the context of mechanistic effect models for risk assessment of plant protection products', *EFSA Journal*, 12: 3589.

———. 2017. 'Guidance on the use of the weight of evidence approach in scientific assessments', 15: 69.

Draft EFSA Guidance on Uncertainty in Scientific Assessments. https://www.efsa.europa.eu/sites/default/files/160321DraftGDUncertaintyInScientificAssessment.pdf

EU. 2002. 'Regulation (EC) No 178/2002 of the European Parliament and of the Council of 28 January 2002 laying down the general principles and requirements of food law, establishing the European Food Safety Authority and laying down procedures in matters of food safety. ', *Official Journal of the European Union*, L 31/1.

———. 2009. '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.', *Official Journal of the European Union*, L 309/1.

———. 2012. '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. ', *Official Journal of the European Union*, L167 / 1.

———. 2013. '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. ', *Official Journal of the European Union*, L 93.

———. 2017a. '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. L 301/1.', *Official Journal of the European Union*, L301.

———. 2017b. 'DRAFT COMMISSION REGULATION (EU) .../... of XXX setting out scientific criteria for the determination of endocrine disrupting properties

and amending Annex II to Regulation (EC) 1107/2009.'. https://ec.europa.eu/health/sites/health/files/endocrine_disruptors/docs/2016_pppcriteria_en.pdf

Fan, X., and L. Wu. 2016. 'The impact of thyroid abnormalities during pregnancy on subsequent neuropsychological development of the offspring: a meta-analysis', *J Matern Fetal Neonatal Med*, 29: 3971-6.

Foo, JTW, and TJ Lam 1993. 'Serum cortisol response to handling stress and the effect of cortisol implantation on testosterone level in the tilapia, *Oreochromis mossambicus*.', *Aquaculture*, 115: 145-58.

Ghassabian, A., H. El Marroun, R. P. Peeters, V. W. Jaddoe, A. Hofman, F. C. Verhulst, H. Tiemeier, and T. White. 2014. 'Downstream effects of maternal hypothyroxinemia in early pregnancy: nonverbal IQ and brain morphology in school-age children', *J Clin Endocrinol Metab*, 99: 2383-90.

Grim, K. C., M. Wolfe, T. Braunbeck, T. Iguchi, Y. Ohta, O. Tooi, L. Touart, D. C. Wolf, and J. Tietge. 2009. 'Thyroid histopathology assessments for the amphibian metamorphosis assay to detect thyroid-active substances', *Toxicol Pathol*, 37: 415-24.

Helfman, G, B Collette, and D Facey. 1997. *The diversity of fishes.* (Blackwell Science: Malden, MA).

JRC. 2013. *Key scientific issues relevant to the identification and characterisation of endocrine disrupting substances - Report of the Endocrine Disrupters Expert Advisory Group. JRC Science for Policy Reports* (Publications Office of the European Union: Luxembourg).

———. 2014. *Alternative methods for regulatory toxicology – a state-of-the-art review* (Publications Office of the European Union, Luxembourg: Luxembourg).

———. 2016. 'Screening methodology to identify potential endocrine disruptors according to different options in the context of an impact assessment', Publications Office of the European Union, Luxembourg. http://publications.jrc.ec.europa.eu/repository/bitstream/JRC101950/jrc%20screening%20methodology%20for%20ed%20impact%20assessment%20%28online%29.pdf.

Judson, R. S., K. A. Houck, E. D. Watt, and R. S. Thomas. 2017. 'On selecting a minimal set of in vitro assays to reliably determine estrogen agonist activity', *Regul Toxicol Pharmacol*, 91: 39-49.

Judson, R. S., F. M. Magpantay, V. Chickarmane, C. Haskell, N. Tania, J. Taylor, M. H. Xia, R. L. Huang, D. M. Rotroff, D. L. Filer, K. A. Houck, M. T. Martin, N. Sipes, A. M. Richard, K. Mansouri, R. W. Setzer, T. B. Knudsen, K. M. Crofton, and R. S. Thomas. 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: 137-54.

Kleinstreuer, N. C., P. Ceger, E. D. Watt, M. Martin, K. Houck, P. Browne, R. S. Thomas, W. M. Casey, D. J. Dix, D. Allen, S. Sakamuru, M. Xia, R. Huang, and R. Judson. 2017. 'Development and Validation of a Computational Model for Androgen Receptor Activity', *Chem Res Toxicol*, 30: 946-64.

Klimisch, H. J., M. Andreae, and U. Tillmann. 1997. 'A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data', *Regulatory Toxicology and Pharmacology*, 25: 1-5.

Kohler, H. R., and R. Triebskorn. 2013. 'Wildlife Ecotoxicology of Pesticides: Can We Track Effects to the Population Level and Beyond?', *Science*, 341: 759-65.

Kramer, V. J., M. A. Etterson, M. Hecker, C. A. Murphy, G. Roesijadi, D. J. Spade, J. A. Spromberg, M. Wang, and G. T. Ankley. 2011. 'Adverse Outcome Pathways and Ecological Risk Assessment Bridging to Population-Level Effects', *Environmental Toxicology and Chemistry*, 30: 64-76.

Kwon, H. C., S. H. Choi, Y. U. Kim, S. O. Son, and J. Y. Kwon. 2005. 'Androgen action on hepatic vitellogenin synthesis in the eel, Anguilla japonica is suppressed by an androgen receptor antagonist', *J Steroid Biochem Mol Biol*, 96: 175-8.

Li, J. Z., and P. Gramatica. 2010. 'Classification and Virtual Screening of Androgen Receptor Antagonists', *Journal of Chemical Information and Modeling*, 50: 861-74.

Lo Piparo, E., and A. Worth. 2010. "Review of (Q)SAR Models and Software Tools for Predicting Developmental and Reproductive Toxicity." In *JRC Scientific and Technical Reports*.

Lundholm, C. D. 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: 113-28.

Marty, M. S., A. Blankinship, J. Chambers, L. Constantine, W. Kloas, A. Kumar, L. Lagadic, J. Meador, D. Pickford, T. Schwarz, and T. Verslycke. 2017. 'Population-relevant endpoints in the evaluation of endocrine-active substances (EAS) for ecotoxicological hazard and risk assessment', *Integr Environ Assess Manag*, 13: 317-30.

McGee, T. D., J. Edwards, and A. E. Roitberg. 2008. 'Preliminary molecular dynamic simulations of the estrogen receptor alpha ligand binding domain from antagonist to apo', *Int J Environ Res Public Health*, 5: 111-4.

Meek, M. E., A. Boobis, I. Cote, V. Dellarco, G. Fotakis, S. Munn, J. Seed, and C. Vickers. 2014. 'New developments in the evolution and application of the WHO/IPCS framework on mode of action/species concordance analysis', *J Appl Toxicol*, 34: 1-18.

Meek, M. E., C. M. Palermo, A. N. Bachman, C. M. North, and R. Jeffrey Lewis. 2014. 'Mode of action human relevance (species concordance) framework: Evolution of the Bradford Hill considerations and comparative analysis of weight of evidence', *J Appl Toxicol*, 34: 595-606.

Nakamura, M., T. Kobayashi, X. T. Chang, and Y. Nagahama. 1998. 'Gonadal sex differentiation in teleost fish', *Journal of Experimental Zoology*, 281: 362-72.

Nanda, I., U. Hornung, M. Kondo, M. Schmid, and M. Schartl. 2003. 'Common spontaneous sex-reversed XX males of the medaka Oryzias latipes', *Genetics*, 163: 245-51.

Nelson, K. R., A. L. Schroeder, G. T. Ankley, B. R. Blackwell, C. Blanksma, S. J. Degitz, K. M. Flynn, K. M. Jensen, R. D. Johnson, M. D. Kahl, D. Knapen, P. A. Kosian, R. Y. Milsk, E. C. Randolph, T. Saari, E. Stinckens, L. Vergauwen, and D. L. Villeneuve. 2016. 'Impaired anterior swim bladder inflation following exposure to the thyroid peroxidase inhibitor 2-mercaptobenzothiazole part I: Fathead minnow', *Aquat Toxicol*, 173: 192-203.

NRC. 2014. *Review of EPA's Integrated Risk Information System (IRIS) Process* (Press TNA; Washington, DC).

OECD. 1981a. *Test No. 410: Repeated Dose Dermal Toxicity: 21/28-day Study* (OECD Publishing).

———. 1981b. *Test No. 411: Subchronic Dermal Toxicity: 90-day Study* (OECD Publishing).

———. 1983. *Test No. 415: One-Generation Reproduction Toxicity Study* (OECD Publishing).

———. 1984. *Test No. 206: Avian Reproduction Test* (OECD Publishing).

———. 1998a. *Test No. 408: Repeated Dose 90-Day Oral Toxicity Study in Rodents* (OECD Publishing).

———. 1998b. *Test No. 409: Repeated Dose 90-Day Oral Toxicity Study in Non-Rodents* (OECD Publishing).

———. 1999. "The Application of the GLP Principles to Field Studies." In. Paris: OECD Publishing.

———. 2001a. *Test No. 414: Prenatal Development Toxicity Study* (OECD Publishing).

———. 2001b. *Test No. 416: Two-Generation Reproduction Toxicity* (OECD Publishing).

———. 2002. "Report of the SETAC/OECD workshop on avian toxicity testing. Series on testing and assessment number 5." In. Paris: OECD.

———. 2004. *Detailed Review Paper on Fish Screening Assays for the Detection of Endocrine Active Substances* (OECD Publishing).

———. 2006. *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. Series on Testing & Assessment No. 68* (OECD Publishing: Paris).

———. 2007a. "Detailed review paper for avian two-generation toxicity test. Series on testing and assessment number 74." In. Paris: OECD.

———. 2007b. "Guidance Document on the Uterotrophic bioassay – Procedure to test for antioestrogenicity. Series on Testing & Assessment No. 71." In. Paris: OECD Publishing.

———. 2007c. *Test No. 426: Developmental Neurotoxicity Study* (OECD Publishing).

———. 2007d. *Test No. 440: Uterotrophic Bioassay in Rodents* (OECD Publishing).

———. 2008. *Test No. 407: Repeated Dose 28-day Oral Toxicity Study in Rodents* (OECD Publishing).

———. 2009a. "Guidance Document on the weanling Hershberger bioassay in rats: a short-term screening assay for (anti)androgenic properties. Series on Testing & Assessment No. 115." In. Paris: OECD Publishing.

———. 2009b. *Test No. 230: 21-day Fish Assay* (OECD Publishing).

———. 2009c. *Test No. 231: Amphibian Metamorphosis Assay* (OECD Publishing).

———. 2009d. *Test No. 441: Hershberger Bioassay in Rats* (OECD Publishing).

———. 2009e. *Test No. 451: Carcinogenicity Studies* (OECD Publishing).

———. 2009f. *Test No. 452: Chronic Toxicity Studies* (OECD Publishing).

———. 2009g. *Test No. 453: Combined Chronic Toxicity/Carcinogenicity Studies* (OECD Publishing).

———. 2010. "Guidance document on the diagnosis of endocrine-related histopathology in fish gonads. OECD Series on Testing and Assessment no. 123." In, 114. Paris: OECD Publishing.

———. 2011a. *Guidance Document on the Androgenised Female Stickleback screen, Series on Testing and Assessment No. 148* (OECD Publishing: Paris).

———. 2011b. *Test No. 234: Fish Sexual Development Test* (OECD Publishing).

———. 2011c. *Test No. 456: H295R Steroidogenesis Assay* (OECD Publishing).

———. 2012a. "Guidance Document on standardised test guidelines for evaluating chemicals for endocrine disruption. Series on Testing and Assessment No. 150." In. Paris: OECD Environment, Health and Safety Publications, Paris.

———. 2012b. "OECD Conceptual Framework for Testing and Assessment of Endocrine Disrupters." In. Paris: OECD.

———. 2012c. *Test No. 229: Fish Short Term Reproduction Assay* (OECD Publishing).

———. 2012d. *Test No. 443: Extended One-Generation Reproductive Toxicity Study* (OECD Publishing).

———. 2012e. *Test No. 455: Performance-Based Test Guideline for Stably Transfected Transactivation In Vitro Assays to Detect Estrogen Receptor Agonists and Antagonists* (OECD Publishing).

———. 2012f. *Test No. 457: BG1Luc Estrogen Receptor Transactivation Test Method for Identifying Estrogen Receptor Agonists and Antagonists* (OECD Publishing).

———. 2013a. "Guidance Document supporting OECD Test Guideline 443 on the extended one-generation reproductive toxicity test. Series on Testing & Assessment No. 151." In. Paris: OECD Publishing.

———. 2013b. *Test No. 210: Fish, Early-life Stage Toxicity Test* (OECD Publishing).

———. 2014. *Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models; Series on Testing and Assessment No 69*

(OECD Publishing, Paris).

———. 2015a. *Guidance Document on histopathology techniques and evaluation (Part I) for the larval amphibian growth and development assay (LAGDA) Series on Testing & Assessment No. 228* (OECD Publishing: Paris).

———. 2015b. *Guidance Document on histopathology techniques and evaluation (Part II) for the larval amphibian growth and development assay (LAGDA) Series on Testing & Assessment No. 228* (OECD Publishing: Paris).

———. 2015c. *Test No. 240: Medaka Extended One Generation Reproduction Test (MEOGRT)* (OECD Publishing).

———. 2015d. *Test No. 241: The Larval Amphibian Growth and Development Assay (LAGDA)* (OECD Publishing).

———. 2015e. *Test No. 493: Performance-Based Test Guideline for Human Recombinant Estrogen Receptor (hrER) In Vitro Assays to Detect Chemicals with ER Binding Affinity* (OECD Publishing).

———. 2016a. *Test No. 421: Reproduction/Developmental Toxicity Screening Test* (OECD Publishing).

———. 2016b. *Test No. 422: Combined Repeated Dose Toxicity Study with the Reproduction/Developmental Toxicity Screening Test* (OECD Publishing).

———. 2016c. *Test No. 458: Stably Transfected Human Androgen Receptor Transcriptional Activation Assay for Detection of Androgenic Agonist and Antagonist Activity of Chemicals* (OECD Publishing).

———. 2016d. *Users' Handbook supplement to the Guidance Document for developing and assessing Adverse Outcome Pathways* (OECD Publishing).

———. 2017a. 'DRAFT Guidance document on Good In Vitro Method Practices (GIVIMP) for the development and Implementation of In Vitro Methods for Regulatory use in Human Safety Assessment'. http://www.oecd.org/env/ehs/testing/OECD_Draft_GIVIMP_in_Human_Safety_Assessment.pdf.

———. 2017b. 'DRAFT Guidance Document on standardised test guidelines for evaluating chemicals for endocrine disruption. Series on Testing and Assessment No. 150.', OECD Environment, Health and Safety Publications, Paris. https://www.oecd.org/env/ehs/testing/GD150_Jul2017_clean_v2.pdf.

———. 2017c. *Guidance Document for Describing Non-Guideline In Vitro Test Methods; Series on testing and assessment No 211* (OECD Publishing).

———. 2017d. "Guidance document on Developing and Assessing Adverse Outcome Pathways. Series on Testing and Assessment No. 184." In. Paris: OECD Publishing.

———. 2017e. *Guidance Document on the Reporting of Defined Approaches to be Used Within Integrated Approaches to Testing and Assessment; Series on testing and assessment No 225* (OECD Publishing).

———. 2017f. *Test No. 412: Subacute Inhalation Toxicity: 28-Day Study* (OECD Publishing).

———. 2017g. *Test No. 413: Subchronic Inhalation Toxicity: 90-day Study* (OECD Publishing).

———. 2017h. 'Work plan for the Test Guidelines Programme (TGP)'. http://www.oecd.org/chemicalsafety/testing/TGP%20work%20plan_August%202017.pdf.

Ospina-Alvarez, N., and F. Piferrer. 2008. 'Temperature-dependent sex determination in fish revisited: prevalence, a single sex ratio response pattern, and possible effects of climate change', *PLoS One*, 3: e2837.

Palace, V. P., R. E. Evans, K. G. Wautier, K. H. Mills, P. J. Blanchfield, B. J. Park, C. L. Baron, and K. A. Kidd. 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: 1920-35.

Panaye, A., J. P. Doucet, J. Devillers, N. Marchand-Geneste, and J. M. Porcher. 2008. 'Decision trees versus support vector machine for classification of androgen receptor ligands', *Sar and Qsar in Environmental Research*, 19: 129-51.

Polzonetti-Magni, AM, G Mosconi, F Palermo, E Cottone, S Forno, and MF Franzoni. 2004. 'Amphibians as a model for studying xenoestrogen-dependent neuroendocrine disruption', *Ital. J. Zool.*: 95-100.

Renjith, P., and K. Jegatheesan. 2015. 'Machine learning classifier algorithms to predict endocrine toxicity of chemicals', *International Journal of Toxicological and Pharmacological Research*, 7: 303-08.

Ribay, K., MT. Kim, W. Wang, D. Pinolini, and H. Zhu. 2016. 'Predictive modelling of estrogen receptor binding agents using advanced cheminformatics tools and massive public data', *Frontiers in Environmental Science*, 4.

Schapaugh, A. W., L. G. McFadden, L. M. Zorrilla, D. R. Geter, L. D. Stuchal, N. Sunger, and C. J. Borgert. 2015. 'Analysis of EPA's endocrine screening battery and recommendations for further review', *Regul Toxicol Pharmacol*, 72: 552-61.

Scholz, S., and N. Kluver. 2009. 'Effects of endocrine disrupters on sexual, gonadal development in fish', *Sex Dev*, 3: 136-51.

Slater, E. P., G. Redeuilh, and M. Beato. 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', *Mol Endocrinol*, 5: 386-96.

Stinckens, E., L. Vergauwen, A. L. Schroeder, W. Maho, B. R. Blackwell, H. Witters, R. Blust, G. T. Ankley, A. Covaci, D. L. Villeneuve, and D. Knapen. 2016. 'Impaired anterior swim bladder inflation following exposure to the thyroid peroxidase inhibitor 2-mercaptobenzothiazole part II: Zebrafish', *Aquat Toxicol*, 173: 204-17.

Susser, M. 1991. 'What is a cause and how do we know one? A grammar for pragmatic epidemiology', *Am J Epidemiol*, 133: 635-48.

Trigunaite, A., J. Dimo, and T. N. Jorgensen. 2015. 'Suppressive effects of androgens on the immune system', *Cellular Immunology*, 294: 87-94.

US EPA. 2009a. *Endocrine Disruptor Screening Program Test Guidelines 890.2100: Avian Two-Generation Toxicity Test in the Japanese Quail* (US EPA: Cincinnati (OH)).

———. 2009b. *Endocrine Disruptor Screening Program Test Guidelines - OPPTS 890.1250: Estrogen Receptor Binding Assay Using Rat Uterine Cytosol (ER-RUC)* (US EPA Risk Assessment Forum, Washington (DC)).

———. 2009c. *Endocrine Disruptor Screening Program Test Guidelines - OPPTS 890.1300: Estrogen Receptor Transcriptional Activation (Human Cell Line (HeLa-9903))* (US EPA Risk Assessment Forum, Washington (DC): Washington (DC)).

———. 2009d. *Endocrine Disruptor Screening Program Test Guidelines - OPPTS 890.1500: Pubertal Development and Thyroid Function in Intact Juvenile/Peripubertal Male Rats* (US EPA Risk Assessment Forum, Washington (DC): Washington (DC)).

———. 2009e. *Endocrine Disruptor Screening Program Test Guidelines - OPPTS 890.1550: Steroidogenesis (Human Cell line - H295R)* (US EPA Risk Assessment Forum, Washington (DC): Washington (DC)).

———. 2009f. *ndocrine Disruptor Screening Program Test Guidelines - OPPTS 890.1450: Pubertal Development and Thyroid Function in Intact Juvenile/Peripubertal Female Rats* (US EPA Risk Assessment Forum, Washington (DC): Washington (DC)).

Vandenberg, L. N., M. Agerstrand, A. Beronius, C. Beausoleil, A. Bergman, L. A. Bero, C. G. Bornehag, C. S. Boyer, G. S. Cooper, I. Cotgreave, D. Gee, P. Grandjean, K. Z. Guyton, U. Hass, J. J. Heindel, S. Jobling, K. A. Kidd, A. Kortenkamp, M. R. Macleod, O. V. Martin, U. Norinder, M. Scheringer, K. A. Thayer, J. Toppari, P. Whaley, T. J. Woodruff, and C. Ruden. 2016. 'A proposed framework for the systematic review and integrated assessment (SYRINA) of endocrine disrupting chemicals', *Environ Health*, 15: 74.

Vedani, A., M. Dobler, and M. Smiesko. 2012. 'VirtualToxLab - a platform for estimating the toxic potential of drugs, chemicals and natural products', *Toxicol Appl Pharmacol*, 261: 142-53.

Wang, M., L. Chai, H. Zhao, M. Wu, and H. Wang. 2015. 'Effects of nitrate on metamorphosis, thyroid and iodothyronine deiodinases expression in Bufo gargarizans larvae', *Chemosphere*, 139: 402-9.

Warren, G. L., C. W. Andrews, A. M. Capelli, B. Clarke, J. LaLonde, M. H. Lambert, M. Lindvall, N. Nevins, S. F. Semus, S. Senger, G. Tedesco, I. D. Wall, J. M. Woolven, C. E. Peishoff, and M. S. Head. 2006. 'A critical assessment of docking programs and scoring functions', *Journal of Medicinal Chemistry*, 49: 5912-31.

WHO/IPCS. 2002. "Global Assessment of the State-of-the-Science of Endocrine Disruptors." In. Geneva: World Health Organization.

WHO/ICPS. 2009. Principles and Methods for the Risk Assessment of Chemicals in Food. Environmental Health Criteria 240, publicly available at http://www.who.int/foodsafety/publications/chemical-food/en/

Wu, S., J. Fisher, J. Naciff, M. Laufersweiler, C. Lester, G. Daston, and K. Blackburn. 2013. 'Framework for identifying chemicals with structural features associated with the potential to act as developmental or reproductive toxicants', *Chem Res Toxicol*, 26: 1840-61.

Yang, C., A. Tarkhov, J. Marusczyk, B. Bienfait, J. Gasteiger, T. Kleinoeder, T. Magdziarz, O. Sacher, C. H. Schwab, J. Schwoebel, L. Terfloth, K. Arvidson, A. Richard, A. Worth, and J. Rathman. 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: 510-28.

Zhang, L., A. Sedykh, A. Tripathi, H. Zhu, A. Afantitis, V. D. Mouchlis, G. Melagraki, I. Rusyn, and A. Tropsha. 2013. 'Identification of putative estrogen receptor-mediated endocrine disrupting chemicals using QSAR- and structure-based virtual screening approaches', *Toxicol Appl Pharmacol*, 272: 67-76.

Zhao, C. Y., R. S. Zhang, H. X. Zhang, C. X. Xue, H. X. Liu, M. C. Liu, Z. D. Hu, and B. T. Fan. 2005. 'QSAR study of natural, synthetic and environmental endocrine disrupting compounds for binding to the androgen receptor', *SAR QSAR Environ Res*, 16: 349-67.

Appendix A – Additional considerations on how to assess the potential for thyroid disruption

Abbreviations

Triiodothyronine (T3); thyroxine (T4); thyroid hormone (TH); thyroid-stimulating hormone (TSH); thyrotropin-releasing hormone (TRH); hypothalamic–pituitary–thyroid axis (HPT axis); thyroxine-binding globulin (TBG); transthyretin (TTR); thyroglobin (TG); developmental neurotoxicity (DNT).

Background

The thyroid gland and its associated hormones are of interest for regulatory toxicology due to its important role in metabolism, growth and development. 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. This negative feedback also occurs when levels of TSH are high, by supressing TRH production.

The hypothalamic-pituitary-thyroid axis (HPT axis) has been conserved across evolution in all 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 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; 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.

Using the current understanding of thyroid physiology and toxicology[13] it is proposed that the following be applied when interpreting data from experimental animals:

1. It is presumed that substances that alter the circulating levels of T3 and/or T4 with concurrent histopathological findings in the thyroid would pose a hazard for human thyroid hormone insufficiency in adults as well as pre- and post-natal neurological development of offspring.
2. It is presumed that 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 presumed 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 investigations can inform more specifically on the mode of action of the thyroid-disruption and its human relevance.

[13] European workshop on Thyroid disruption organised by the European Commission and ANSES held in Paris 29-31 March 2017 (European Commission 2017).

Investigation of increase in thyroid hormone metabolism in the liver

In cases where changes in TH levels or in thyroid follicular cell histopathology are observed in rodents (particularly in the rat) in the absence of such effects in other tested animal species (e.g. dog), human relevance of such effects could be further investigated. 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 (Capen 1997; Curran and DeGroot 1991; Ennulat et al. 2010).

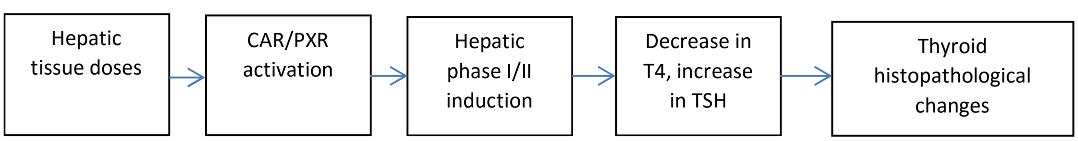
To investigate whether liver enzyme induction is responsible for the effects seen on TH levels or thyroid histopathology and weight, as well as the likely human relevance of the effect, the following information is needed:

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 toxicity study should be considered. This study should measure TSH, T3 and T4 and, where possible, additional data on liver induction (e.g. measurement of UDPGT).

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 and humans. Enzymes activities should be investigated in the context of the IPCS mode of action and human relevancy framework (Boobis et al. 2006) investigating significant quantitative species differences.

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 potential for inhibition of the sodium-iodide symporter (NIS) (Cianchetta et al. 2010; Kogai and Brent 2012) or thyroid peroxidase (TPO) (Kambe and Seo 1997; Wu, Beland, and Fang 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/standardized in vitro assays do not exist to investigate all these different pathways.

An example of putative mode of action is reported below:

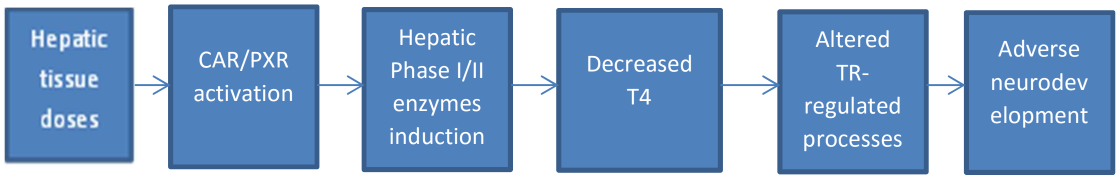


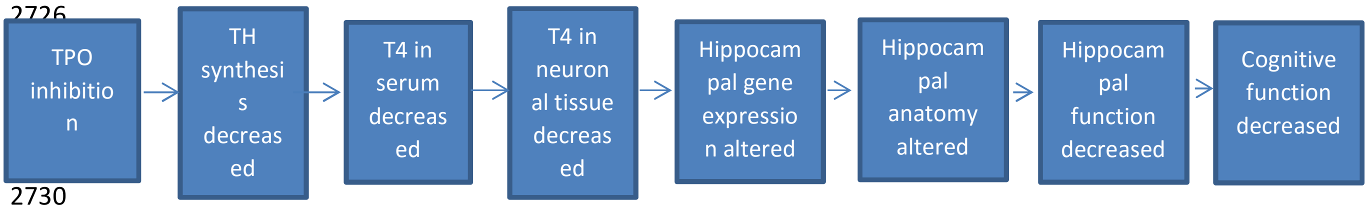
The assessment of quantitative differences in hepatic induction can therefore be used to provide evidence of non-relevance to human.

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 other histological changes and/or hormonal evidence of hypothyroidism is a relatively frequent observation in experimental toxicological studies, particularly in rodents. 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. 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 TGs 421/422, 426, 416, 443) (OECD 2001, 2012, 2016b, 2016a) 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 foetus and new-born) 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).

Examples of putative modes of action are reported below:





Further investigations of thyroid disruption

An in-depth understanding of the fundamental principles that regulate TH homeostasis 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

Bianco, A. C., G. Anderson, D. Forrest, V. A. Galton, B. Gereben, B. W. Kim, P. A. Kopp, X. H. Liao, M. J. Obregon, R. P. Peeters, S. Refetoff, D. S. Sharlin, W. S. Simonides, R. E. Weiss, G. R. Williams, Approaches American Thyroid Association Task Force on, Economy Strategies to Investigate Thyroid Hormone, and Action. 2014. 'American Thyroid Association Guide to investigating thyroid hormone economy and action in rodent and cell models', *Thyroid*, 24: 88-168.

Boobis, A. R., S. M. Cohen, V. Dellarco, D. McGregor, M. E. Meek, C. Vickers, D. Willcocks, and W. Farland. 2006. 'IPCS framework for analyzing the relevance of a cancer mode of action for humans', *Crit Rev Toxicol*, 36: 781-92.

Capen, C. C. 1997. 'Mechanistic data and risk assessment of selected toxic end points of the thyroid gland', *Toxicol Pathol*, 25: 39-48.

Cianchetta, S., J. di Bernardo, G. Romeo, and K. J. Rhoden. 2010. 'Perchlorate transport and inhibition of the sodium iodide symporter measured with the yellow fluorescent protein variant YFP-H148Q/I152L', *Toxicol Appl Pharmacol*, 243: 372-80.

Curran, P. G., and L. J. DeGroot. 1991. 'The effect of hepatic enzyme-inducing drugs on thyroid hormones and the thyroid gland', *Endocr Rev*, 12: 135-50.

Ennulat, D., D. Walker, F. Clemo, M. Magid-Slav, D. Ledieu, M. Graham, S. Botts, and L. Boone. 2010. 'Effects of hepatic drug-metabolizing enzyme induction on clinical pathology parameters in animals and man', *Toxicol Pathol*, 38: 810-28.

European Commission. 2017. 'Supporting the organisation of a workshop on thyroid disruption –Final Report', Publications Office of the European Union, Luxembourg. https://publications.europa.eu/en/publication-detail/-/publication/472d2c88-a8b1-11e7-837e-01aa75ed71a1/language-en.

Janssen, S. T., and O. E. Janssen. 2017. 'Directional thyroid hormone distribution via the blood stream to target sites', *Mol Cell Endocrinol*, 458: 16-21.

Kambe, F., and H. Seo. 1997. 'Thyroid-specific transcription factors', *Endocr J*, 44: 775-84.

Kogai, T., and G. A. Brent. 2012. 'The sodium iodide symporter (NIS): regulation and approaches to targeting for cancer therapeutics', *Pharmacol Ther*, 135: 355-70.

OECD. 2001. *Test No. 416: Two-Generation Reproduction Toxicity* (OECD Publishing).

———. 2012. *Test No. 443: Extended One-Generation Reproductive Toxicity Study* (OECD Publishing).

———. 2016a. *Test No. 421: Reproduction/Developmental Toxicity Screening Test* (OECD Publishing).

———. 2016b. *Test No. 422: Combined Repeated Dose Toxicity Study with the Reproduction/Developmental Toxicity Screening Test* (OECD Publishing).

US EPA. 2005. 'Guidance for Thyroid Assays in Pregnant Animals, Fetuses and Postnatal Animals, and Adult Animals', US EPA, Office of Pesticide Programs. https://www.epa.gov/sites/production/files/2015-06/documents/thyroid_guidance_assay.pdf.

Wu, Y., F. A. Beland, and J. L. Fang. 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', *Toxicol In Vitro*, 32: 310-9.

Appendix B – Recommendations for design, conduction and technical evaluation of hormonal studies

Abbreviations

European Union (EU); Follicle-stimulating hormone (FSH); luteinising hormone (LH); triiodothyronine (T3); thyroxine (T4); thyroid-stimulating hormone (TSH); Repeated dose 28-day oral toxicity study in rodents (OECD TG 407); post-natal day (PND); radioimmunoassay (RIA).

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 7days) is generally required to reach a steady state for the response and adaptation of hormone dependent organs, if they are included in the investigation (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 estrous 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.

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.

Free or bound fraction to be measured. A high volume of serum (approximately 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 pre-treated (e.g. ultracentrifugation or dialysis). Chromatography or equally sensitive techniques should be applied for detection of free hormone in adults; furthermore, the applicability of RIA for the pups is questionable in terms of sensitivity (personal communication).

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 foetuses 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 foetuses.

Sex. Both sexes can be used for measurement of thyroid hormones. Synchronisation of females is not a pre-requisite for thyroid hormonal assay.

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

Wilcoxon test, two-sided (power 75%; $p < 0.05$)

Rats 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

CV: coefficient of variation

Animal care. Animal care and housing should fulfil the requirements according to current EU legislation (Directive 2010/63/EU revising Directive 86/609/EEC on the protection of animals used for scientific purposes). 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. All of the samples of one study should be taken in the shortest possible time (not more than 2 hours). 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 CO2 should be avoided for animal welfare reasons and due to interference with the concentrations of the thyroid hormones in exposed animals.

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 minutes per animal under anaesthesia and 1 minute 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 minutes to acclimatize. Extended acclimatisation for up to 24 hours is not necessary.

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 foetuses, 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 minutes 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, pre-treatment 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 and the upper range of the assay sensitivity should be established.

b) Reproducibility of the assay should be assessed and the coefficients of the inter- and intra-assay variation should be calculated. 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.

c) Repeatability of the assay should be proven.

d) The type of applied quality control samples (e.g. spiked samples, biological control samples, reference range etc.) should be recorded.

e) The performance of the assay with a particular matrix (serum or plasma) should be assessed.

f) A validation study, conducted with a positive control (reference compound) should be available to establish the laboratory's proficiency in performing the assay.

g) Stability of the sample under selected storage conditions should be validated.

h) Validation of the assay should be carried out for each species separately.

i) If the measurements of the free fraction of T3 and T4 are conducted in mechanistic studies, pre-treatment of samples is required, followed by chromatographic detection of the non-bound fractions of the hormones. Cross-reactivity of antibodies used in the assay should be established at least at the level of the kit manufacturer.

j) 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 size should be detected by comparing to values in the concomitant control group. However, each laboratory conducting thyroid hormone analyses should develop their own historical control range. 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 standardized housing/dietary/environmental conditions.

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 estradiol, 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 estradiol, 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 estrous 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 estrous cycle.

Consideration of effects of circadian rhythm. Blood sampling should be accomplished in a 3-hour 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 randomization of females, the stage of estrous 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 minutes. 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

Bianco, A. C., G. Anderson, D. Forrest, V. A. Galton, B. Gereben, B. W. Kim, P. A. Kopp, X. H. Liao, M. J. Obregon, R. P. Peeters, S. Refetoff, D. S. Sharlin, W. S. Simonides, R. E. Weiss, G. R. Williams, Approaches American Thyroid Association Task Force on, Economy Strategies to Investigate Thyroid Hormone, and Action. 2014. 'American Thyroid Association Guide to investigating thyroid hormone economy and action in rodent and cell models', *Thyroid*, 24: 88-168.

OECD. 2008. *Test No. 407: Repeated Dose 28-day Oral Toxicity Study in Rodents* (OECD Publishing).

Sandow, J. 2006. 'Assays in Endocrine Safety Pharmacology.' in Springer Edition (ed.), *Drug Discovery and Evaluation: Safety and Pharmacokinetic Assays* (Springer-Verlag Berlin: Heidelberg, New York).

Appendix C – Information requirements for active substances under the Biocidal Products and Plant Protection Products Regulations 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**[1]
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 (Level 4)	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 (Level 4)	Information requirement	Information requirement
Subchronic repeated dose toxicity study (90 days; OECD TG 409), in a non-rodent species. Preferred species is dog (Level 4)	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 (Level 4)	Information requirement[2]	Information requirement[2]
Further repeat dose studies (Level 4)	Triggered	Triggered

	PPP	BP[1]
Reproductive toxicity		
Pre-natal developmental toxicity study (OECD TG 414) in a first species, rabbit is preferred (Level 4)	Information requirement	Information requirement
Pre-natal developmental toxicity study (OECD TG 414) in a second species, rat is preferred (Level 4)	Information requirement[3]	Triggered
Developmental neurotoxicity (OECD TG 426; Level 4)	Triggered	Triggered
Two-generation reproductive toxicity study (OECD TG 416), in rats (Level 5)	Information requirement[4]	Information requirement[4]
Extended one-generation reproduction toxicity (OECD TG 443) including the second generation and neurotoxicity and immunotoxicity cohorts (Level 5)	See notes 4,5	See notes 4,5
Carcinogenicity		
Carcinogenicity testing in a first species (OECD TG 451), rat is the preferred species (Level 4)	Information requirement[6]	Information requirement[6]
Carcinogenicity testing in a second species (OECD TG 451), mouse is the preferred species (Level 4)	Information requirement[6]	Information requirement[6]
Endocrine-disrupting properties[7]		
H295R Steroidogenesis assay (OECD TG 456 Level 2)	Triggered	Triggered
Stably transfected human estrogen receptor alpha transcriptional activation assay for detection of estrogenic agonist-activity of chemicals (OECD TG 455 Level 2)	Triggered	Triggered
Uterotrophic assay (mechanistic *in vivo* tests) (OECD TG 440 Level 3)	Triggered	Triggered

	PPP	BP[1]
Hershberger assay (mechanistic *in vivo* test) (OECD TG 441 Level 3)	Triggered	Triggered
Peripubertal male and female assays (OPPTS 890.1500 and 890.1450 Level 4)	Triggered	Triggered
15-day intact adult male rat assay (US EPA 2007 Level 4)	Triggered	Triggered
Relevant human health data	Information requirement	Information requirement
Epidemiological studies on the general population	Information requirement	Information requirement
Literature data[8]	Information requirement	Information requirement in the ED criteria

Notes

1 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'.

2 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.

3 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).

4 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).

5 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).

6 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).

7 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.

8 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).

3027 ## C.2. Ecotoxicological data

	PPP	BP[1]
Effects on birds and other terrestrial vertebrates		
Subchronic and reproductive toxicity to birds (OECD TG 206 Level 4)	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.	Triggered If 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 organisms Triggered
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 shall be discussed with the national competent authorities.	Indication of endocrine activity Triggered
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 hours via hydrolysis).	Triggered
Fish full life cycle test (EPA OPPTS 850.1500-level 5)	Triggered if there is concern regarding ED properties identified in the screening testing battery.	Triggered
Endocrine-disrupting properties for aquatic organisms[2]		
Fish short-term reproduction assay (OECD TG 229 Level 3)[3]	Screening test battery always required unless ED properties can be excluded	Not an information requirement

	based on information on toxicity profile and mode of action.	
21-day fish assay: a short-term screening for estrogenic and androgenic activity, and aromatase inhibition (OECD TG 230 Level 3)	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-level 3)	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 Level 3)	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[4]	Information requirement.	Information requirement in the ED criteria

Notes

1 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'.

2 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.

3 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 230.

4 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).

References

EFSA. 2011. 'Submission of scientific peer-reviewed open literature for the approval of pesticide active substances under Regulation (EC) No 1107/2009', *EFSA Journal*, 9: 2092.

Appendix D – Databases, software tools and literature-derived (Q)SARs

3050 ## D.1. Databases with information on endocrine activity

Database	Link	Availability	Description
Endocrine Disruptor Knowledge Base (EDKB) database (US 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 1800 chemicals, as well as chemical structure search capabilities. Among the data are an ER binding dataset (containing 131 ER binders and 101 non-ER binders), and an AR binding dataset (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) (US 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.

Database	Link	Availability	Description
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 (pre-generated 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-estradiol.
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))

Database	Link	Availability	Description
			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)	http://www.oecd.org/ehs/eChemPortal	Freely avalable	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.
SIN (Substitute it now!) List (International chemical secretariat)	http://sinlist.chemsec.org	Freely available	The database contains chemicals that have been identified by the International chemical secretariat (ChemSec) as being SVHCs, based on the criteria defined in REACH article 57. The list includes accordingly three categories: CMR substances; PBT and vPvB substances; substances of equivalent concern, which include endocrine disrupting chemicals.
TEDX List of Potential Endocrine Disruptors (The endocrine disruption exchange (TEDX))	https://endocrinedisruption.org/interactive-tools/tedx-list-of-potential-endocrine-disruptors/search-the-tedx-list	Freely available	The TEDX List of Potential Endocrine Disruptors identifies chemicals that have shown evidence of endocrine disruption in scientific research. Peer-reviewed research showing effects on endocrine signalling is identified in publicly available scientific literature. The list includes chemicals with at least one study demonstrating endocrine disrupting properties.
AOP Knowledge Base in e.AOP.Portal (OECD)	https://aopkb.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.

Database	Link	Availability	Description
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 datasets. It includes (Q)SAR prediction results for ER binding and developmental toxicity.

3053 ## D.2. Software tools for predicting endocrine activity

Software	Link	Availability	Effect addressed	Description
Endocrine Disruptor Knowledge Base (EDKB) database (US 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.)	http://www.simulations-plus.com	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 IC50 estradiol/IC50 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 datasets 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, S	Quantitative prediction of rat ER binding affinity and AhR binding affinity.
TIMES (Laboratory of Mathematical Chemistry, Bourgas University)	http://oasis-lmc.org	Commercial	E, A, S	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).

Software	Link	Availability	Effect addressed	Description
VirtualToxLab (Vedani and Smiesko 2009; Vedani et al. 2009)	http://www.biograf.ch	Commercial	E, A, T, S	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.
OECD (Q)SAR Toolbox (OECD, ECHA)	https://www.qsartoolbox.org	Freely available	E	The OECD (Q)SAR Toolbox (Dimitrov et al. 2016; OECD 2014b, 2014a) 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	Web service for predicting endocrine disruption potential of molecules, entering structure/SMILES information {Kolsek, 2014 #253}. Includes docking to 18 crystal structures of 14 different nuclear receptors (e.g. AR, ER, GR, LXR, PPAR, RXR, TR).

Software	Link	Availability	Effect addressed	Description
EU project COSMOS KNIME workflow	https://knimewebportal.cosmostox.eu; model executable in the browser of the WebPortal	Freely available	E, A, T, S	Prediction of potential NR binding (PPAR, AR, AHR, ER, GR, PR, farnesoid X receptor (FXR), LXR, PXR, THR, 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, S	The Danish (Q)SAR database is a repository of pre-generated 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, Sild, and Maran 2015) is a repository for (Q)SAR and QSPR models and datasets. 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.

Software	**Link**	**Availability**	**Effect addressed**	**Description**
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.

3057 ## D.3. Literature-derived (Q)SAR models for predicting nuclear receptor binding

Model reference	**Effect addressed**	**Method / type of model**	**Dataset 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 [3H]–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, plasticizers, 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			

Model reference	Effect addressed	Method / type of model	Dataset size and applicability
(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, Ruuskanen, and Tuppurainen 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* assays. Activity patterns across the *in vitro* assays are used to predict ER agonist or antagonist bioactivity and discriminate from assay-specific interference and cytotoxicity.
(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 and Cronin 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 dataset
(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

Model reference	Effect addressed	Method / type of model	Dataset size and applicability
(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, Vedani, and Dobler 2004)	ER binding	Multidimensional (Q)SAR (Raptor)	116 chemicals from NCTR dataset
(Marini, Roncaglioni, and Novic 2005)	ER binding	Various multivariate methods e.g. a back-propagation neural network	132 heterogeneous compounds
(Mansouri et al. 2016; Marini, Roncaglioni, and Novic 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 and Serafimova 2009)	ER binding	COREPA modelling approach combined with metabolic simulation	645 chemicals, including 497 steroid and environmental chemicals and 148 chemicals synthesised for medicinal purposes

Model reference	Effect addressed	Method / type of model	Dataset size and applicability
(Mukherjee, Saha, and Roy 2005)	ER binding	(Q)SAR based on multiple linear regression	25 triphenylacrylonitriles
(Netzeva, Saliner, and Worth 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.
(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, Netzeva, and Worth 2006)	Human ERα binding	Models developed using quantum similarity methods	117 aromatic chemicals
(Salum Lde, Polikarpov, and Andricopulo 2007))	ERα modulators	3D (Q)SAR (CoMFA) and 2D Hologram (Q)SAR	Two training sets containing either 127 or 69 compounds
(Salum, Polikarpov, and Andricopulo 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

Model reference	Effect addressed	Method / type of model	Dataset size and applicability
(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, Dobler, and Lill 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%).

Model reference	Effect addressed	Method / type of model	Dataset size and applicability
(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 datasets form the NRLiSt BDB (http://nrlist.drugdesign.fr)
(Lill, Dobler, and Vedani 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, Steinmetz, and Cronin 2016; Steinmetz et al. 2015)	NR binding: PPAR, AR, AhR, ER, GR, PR, FXR, LXR, PXR, THR, 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. 2016; 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., M. Nakai, Y. Yakabe, M. Takatsuki, M. Mizutani, M. Matsuo, and Y. Shimohigashi. 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 QSAR Environ Res*, 16: 323-37.

Al Sharif, M., I. Tsakovska, I. Pajeva, P. Alov, E. Fioravanzo, A. Bassan, S. Kovarich, C. Yang, A. Mostrag-Szlichtyng, V. Vitcheva, A. P. Worth, A. N. Richarz, and M. T. Cronin. 2016. 'The application of molecular modelling in the safety assessment of chemicals: A case study on ligand-dependent PPARgamma dysregulation', *Toxicology*.

Asikainen, A. H., J. Ruuskanen, and K. A. Tuppurainen. 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', *Environ Sci Technol*, 38: 6724-9.

Browne, P., R. S. Judson, W. M. Casey, N. C. Kleinstreuer, and R. S. Thomas. 2015. 'Screening Chemicals for Estrogen Receptor Bioactivity Using a Computational Model', *Environ Sci Technol*, 49: 8804-14.

Demyttenaere-Kovatcheva, A., M. T. Cronin, E. Benfenati, A. Roncaglioni, and E. Lopiparo. 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', *J Med Chem*, 48: 7628-36.

Dimitrov, S. D., R. Diderich, T. Sobanski, T. S. Pavlov, G. V. Chankov, A. S. Chapkanov, Y. H. Karakolev, S. G. Temelkov, R. A. Vasilev, K. D. Gerova, C. D. Kuseva, N. D. Todorova, A. M. Mehmed, M. Rasenberg, and O. G. Mekenyan. 2016. 'QSAR Toolbox - workflow and major functionalities', *SAR QSAR Environ Res*: 1-17.

Ding, D., L. Xu, H. Fang, H. Hong, R. Perkins, S. Harris, E. D. Bearden, L. Shi, and W. Tong. 2010. 'The EDKB: an established knowledge base for endocrine disrupting chemicals', *BMC Bioinformatics*, 11 Suppl 6: S5.

Dybdahl, M., N. G. Nikolov, E. B. Wedebye, S. O. Jonsdottir, and J. R. Niemela. 2012. 'QSAR model for human pregnane X receptor (PXR) binding: screening of environmental chemicals and correlations with genotoxicity, endocrine disruption and teratogenicity', *Toxicol Appl Pharmacol*, 262: 301-9.

Fang, H., W. Tong, L. M. Shi, R. Blair, R. Perkins, W. Branham, B. S. Hass, Q. Xie, S. L. Dial, C. L. Moland, and D. M. Sheehan. 2001. 'Structure-activity relationships for a large diverse set of natural, synthetic, and environmental estrogens', *Chem Res Toxicol*, 14: 280-94.

Ghafourian, T., and M. T. Cronin. 2005. 'The impact of variable selection on the modelling of oestrogenicity', *SAR QSAR Environ Res*, 16: 171-90.

Hong, H., H. Fang, Q. Xie, R. Perkins, D. M. Sheehan, and W. Tong. 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 QSAR Environ Res*, 14: 373-88.

Hong, H., J. Shen, H. W. Ng, S. Sakkiah, H. Ye, W. Ge, P. Gong, W. Xiao, and W. Tong. 2016. 'A Rat alpha-Fetoprotein Binding Activity Prediction Model to Facilitate Assessment of the Endocrine Disruption Potential of Environmental Chemicals', *Int J Environ Res Public Health*, 13: 372.

Hong, H., W. Tong, Q. Xie, H. Fang, and R. Perkins. 2005. 'An in silico ensemble method for lead discovery: decision forest', *SAR QSAR Environ Res*, 16: 339-47.

Huang, R., M. Xia, S. Sakamuru, J. Zhao, S. A. Shahane, M. Attene-Ramos, T. Zhao, C. P. Austin, and A. Simeonov. 2016. 'Modelling the Tox21 10 K chemical profiles for in vivo toxicity prediction and mechanism characterization', *Nat Commun*, 7: 10425.

Islam, M. A., S. Nagar, S. Das, A. Mukherjee, and A. Saha. 2008. 'Molecular design based on receptor-independent pharmacophore: Application to estrogen receptor ligands', *Biological & Pharmaceutical Bulletin*, 31: 1453-60.

Judson, R. S., F. M. Magpantay, V. Chickarmane, C. Haskell, N. Tania, J. Taylor, M. H. Xia, R. L. Huang, D. M. Rotroff, D. L. Filer, K. A. Houck, M. T. Martin, N. Sipes, A. M. Richard, K. Mansouri, R. W. Setzer, T. B. Knudsen, K. M. Crofton, and R. S. Thomas. 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: 137-54.

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 modelling*, 54(4): 1254-67

Kramer, V. J., and J. P. Giesy. 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: 141-61.

Kurunczi, L., E. Seclaman, T. I. Oprea, L. Crisan, and Z. Simon. 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: 1275-81.

Lagarde, N., S. Delahaye, J. F. Zagury, and M. Montes. 2016. 'Discriminating agonist and antagonist ligands of the nuclear receptors using 3D-pharmacophores', *Journal of Cheminformatics*, 8.

Lill, M. A., M. Dobler, and A. Vedani. 2005. 'In silico prediction of receptor-mediated environmental toxic phenomena - Application to endocrine disruption', *Sar and Qsar in Environmental Research*, 16: 149-69.

Lill, M. A., A. Vedani, and M. Dobler. 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: 6174-86.

Mansouri, K., A. Abdelaziz, A. Rybacka, A. Roncaglioni, A. Tropsha, A. Varnek, A. Zakharov, A. Worth, A. M. Richard, C. M. Grulke, D. Trisciuzzi, D. Fourches, D. Horvath, E. Benfenati, E. Muratov, E. B. Wedebye, F. Grisoni, G. F. Mangiatordi, G. M. Incisivo, H. Hong, H. W. Ng, I. V. Tetko, I. Balabin, J. Kancherla, J. Shen, J. Burton, M. Nicklaus, M. Cassotti, N. G. Nikolov, O. Nicolotti, P. L. Andersson, Q. Zang, R. Politi, R. D. Beger, R. Todeschini, R. Huang, S. Farag, S. A. Rosenberg, S. Slavov, X. Hu, and R. S. Judson. 2016. 'CERAPP: Collaborative Estrogen Receptor Activity Prediction Project', *Environ Health Perspect*, 124: 1023-33.

Marini, F., A. Roncaglioni, and M. Novic. 2005. 'Variable selection and interpretation in structure-affinity correlation modeling of estrogen receptor binders', *Journal of Chemical Information and Modeling*, 45: 1507-19.

Mekenyan, O., and R. Serafimova. 2009. "Mechanism-Based Modeling of Estrogen Receptor Binding Affinity. A Common Reactivity Pattern (COREPA) Implementation. Endocrine Disruption Modeling." In, 259–93. Boca Raton, FL: CRC Press.

Mellor, C. L., F. P. Steinmetz, and M. T. Cronin. 2016. 'The identification of nuclear receptors associated with hepatic steatosis to develop and extend adverse outcome pathways', *Crit Rev Toxicol*, 46: 138-52.

Mukherjee, S., A. Saha, and K. Roy. 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', *Bioorg Med Chem Lett*, 15: 957-61.

Netzeva, T. I., A. G. Saliner, and A. P. Worth. 2006. 'Comparison of the applicability domain of a quantitative structure-activity relationship for estrogenicity with a large chemical inventory', *Environmental Toxicology and Chemistry*, 25: 1223-30.

Ng, H. W., W. Q. Zhang, M. Shu, H. Luo, W. G. Ge, R. Perkins, W. D. Tong, and H. X. Hong. 2014. 'Competitive molecular docking approach for predicting estrogen receptor subtype alpha agonists and antagonists', *BMC Bioinformatics*, 15.

OECD. 2014a. *The Guidance Document for Using the OECD (Q)SAR Application Toolbox to Develop Chemical Categories According to the OECD Guidance on Grouping Chemicals; Series on Testing and Assessment: No 102* (OECD Publishing).

———. 2014b. *Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models; Series on Testing and Assessment No 69*

(OECD Publishing, Paris).

Ribay, K., MT. Kim, W. Wang, D. Pinolini, and H. Zhu. 2016. 'Predictive modelling of estrogen receptor binding agents using advanced cheminformatics tools and massive public data', *Frontiers in Environmental Science*, 4.

Ruusmann, V., S. Sild, and U. Maran. 2015. 'QSAR DataBank repository: open and linked qualitative and quantitative structure-activity relationship models', *J Cheminform*, 7: 32.

Saliner, A. G., T. I. Netzeva, and A. P. Worth. 2006. 'Prediction of estrogenicity: validation of a classification model', *SAR QSAR Environ Res*, 17: 195-223.

Salum, L. B., I. Polikarpov, and A. D. Andricopulo. 2008. 'Structure-based approach for the study of estrogen receptor binding affinity and subtype selectivity', *Journal of Chemical Information and Modeling*, 48: 2243-53.

Salum Lde, B., I. Polikarpov, and A. D. Andricopulo. 2007. 'Structural and chemical basis for enhanced affinity and potency for a large series of estrogen receptor ligands: 2D and 3D QSAR studies', *J Mol Graph Model*, 26: 434-42.

Shen, J., L. Xu, H. Fang, A. M. Richard, J. D. Bray, R. S. Judson, G. Zhou, T. J. Colatsky, J. L. Aungst, C. Teng, S. C. Harris, W. Ge, S. Y. Dai, Z. Su, A. C. Jacobs, W. Harrouk, R. Perkins, W. Tong, and H. Hong. 2013. 'EADB: an estrogenic activity database for assessing potential endocrine activity', *Toxicological Sciences*, 135: 277-91.

Soderholm, A. A., J. Viiliainen, P. T. Lehtovuori, H. Eskelinen, D. Roell, A. Baniahmad, and T. H. Nyronen. 2008. 'Computationally identified novel diphenyl- and phenylpyridine androgen receptor antagonist structures', *Journal of Chemical Information and Modeling*, 48: 1882-90.

Steinmetz, F. P., C. L. Mellor, T. Meinl, and M. T. Cronin. 2015. 'Screening Chemicals for Receptor-Mediated Toxicological and Pharmacological Endpoints: Using Public Data to Build Screening Tools within a KNIME Workflow', *Mol Inform*, 34: 171-8.

Taha, M. O., M. Tarairah, H. Zalloum, and G. Abu-Sheikha. 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.

Tamura, H., Y. Ishimoto, T. Fujikawa, H. Aoyama, H. Yoshikawa, and M. Akamatsu. 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 & Medicinal Chemistry*, 14: 7160-74.

Todorov, M., E. Mombelli, S. Ait-Aissa, and O. Mekenyan. 2011. 'Androgen receptor binding affinity: a QSAR evaluation', *Sar and Qsar in Environmental Research*, 22: 265-91.

Tong, W. D., W. Xie, H. X. Hong, L. M. Shi, H. Fang, and R. Perkins. 2004. 'Assessment of prediction confidence and domain extrapolation of two structure-activity relationship models for predicting estrogen receptor binding activity', *Environmental Health Perspectives*, 112: 1249-54.

Tsakovska, I., M. Al Sharif, P. Alov, A. Diukendjieva, E. Fioravanzo, M. T. D. Cronin, and I. Pajeva. 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: 7651-66.

Vedani, A., M. Dobler, and M. A. Lill. 2005. 'Combining protein modeling and 6D-QSAR. Simulating the binding of structurally diverse ligands to the estrogen receptor', *J Med Chem*, 48: 3700-3.

Vedani, A., and M. Smiesko. 2009. 'In silico toxicology in drug discovery - concepts based on three-dimensional models', *Altern Lab Anim*, 37: 477-96.

Vedani, A., M. Smiesko, M. Spreafico, O. Peristera, and M. Dobler. 2009. '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: 167-76.

Vinggaard, A. M., J. Niemela, E. B. Wedebye, and G. E. Jensen. 2008. 'Screening of 397 chemicals and development of a quantitative structure--activity relationship model for androgen receptor antagonism', *Chem Res Toxicol*, 21: 813-23.

Yang, C., A. Tarkhov, J. Marusczyk, B. Bienfait, J. Gasteiger, T. Kleinoeder, T. Magdziarz, O. Sacher, C. H. Schwab, J. Schwoebel, L. Terfloth, K. Arvidson, A. Richard, A. Worth, and J. Rathman. 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: 510-28.

Zhang, L., A. Sedykh, A. Tripathi, H. Zhu, A. Afantitis, V. D. Mouchlis, G. Melagraki, I. Rusyn, and A. Tropsha. 2013. 'Identification of putative estrogen receptor-mediated endocrine disrupting chemicals using QSAR- and structure-based virtual screening approaches', *Toxicol Appl Pharmacol*, 272: 67-76.

Zhao, C. Y., R. S. Zhang, H. X. Zhang, C. X. Xue, H. X. Liu, M. C. Liu, Z. D. Hu, and B. T. Fan. 2005. 'QSAR study of natural, synthetic and environmental endocrine disrupting compounds for binding to the androgen receptor', *SAR QSAR Environ Res*, 16: 349-67.

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