B. General guidance on endocrine assessment, assays and endpoints

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

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

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

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

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

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

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

B.1. Considerations on the assays addressed

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

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

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

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

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

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

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

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

B.1.1.1. QSAR models

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

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

B.1.1.2. Integrated approaches and models

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

B.1.1.3. Integrated approaches to testing and assessment

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

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

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

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

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

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

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

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

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

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

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

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

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

B.1.2.2 Metabolising systems in in vitro *assays*

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

B.3. Cross-species extrapolations

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

B.6. Regulatory experience of endocrine assessment

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

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

B.6.1. Regulatory experience in the United States

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

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

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

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

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

B.6.2. Regulatory experience in the European Union

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

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

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

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

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

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

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

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

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

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

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

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

B.6.4. Other regulatory experience

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

Notes

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

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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