European Commission

JRC TECHNICAL REPORTS

Expert survey on identification of gaps in available test methods for evaluation of endocrine disruptors

Stephanie Bopp, Malgorzata Nepelska, Marlies Halder and Sharon Munn

2017

Joint Research Centre

EUR 28592 EN

JRC Science Hub
https://ec.europa.eu/jrc

JRC106244

EUR 28592 EN

PDF ISBN 978-92-79-68778-5 ISSN 1831-9424 doi:10.2760/844100

Luxembourg: Publications Office of the European Union, 2017

How to cite this report: Bopp S, Nepelska M, Halder M, Munn S; Expert survey on identification of gaps in available test methods for evaluation of endocrine disruptors; JRC Technical Report, EUR 28592 EN, doi:10.2760/844100.

Contents

Abstract

According to the 2012 WHO/UNEP publication 'State of the Science of Endocrine Disrupting Chemicals' research into endocrine disrupting chemicals over the last decade has indicated that, despite the progress achieved in development and validation of test methods for evaluation of endocrine disruptors, there are still several gaps that need to be addressed. Considering the expected significant amount of work needed to fill the gaps and the limited resources available, it will be important to set priorities for the upcoming period (next 5-10 years) for the development and validation of test methods. Thus there is a need to focus the European input to the OECD test guideline programme to effectively enhance the identification of chemical substances with endocrine disrupting properties whilst making best use of existing resources.

With this objective in mind, DG Environment, supported by JRC, is organising a European expert workshop on setting priorities for further development and validation of test methods for evaluating endocrine disruption. The workshop will take place on 30 May - 01 June 2017 in Brussels. The deliberations will focus on what is necessary and achievable in the context of resources, timescales and animal welfare considerations.

In preparation for the workshop, JRC has drawn up a questionnaire to gather input from experts in the field on key issues to be used as a basis for the further discussions at the workshop. An online survey with the title "Identifying gaps in available test methods for evaluation of endocrine disruptors" was performed on the EU Survey platform and open for commenting from 19/05/2015 until 15/06/2015. A selected group of experts (EFSA Scientific Committee and WG on EDs, ECHA ED WG and RAC, WNT (European members from OECD webpage), Experts identified in Annex 3 of the "Roadmap for setting priorities for further development and validation of test methods and testing approaches for evaluating endocrine disruptors") was invited to participate in the survey.

Experts were asked to rank endocrine related diseases/disorders regarding the possibility to predict them with existing test methods (TMs). They were further asked to rank diseases/disorders regarding the need to develop new test methods to better cover those. Experts were then requested to provide their views on including further tests based on those discussed in the OECD (2012) "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 disruptors" and their views on the current OECD Conceptual Framework and proposals for improvements. Forty experts representing 15 countries and different stakeholder groups (authorities; academia; civil society organisation; industry) replied.

The purpose of this report is to present the detailed survey results. Multiple choice questions were evaluated and where possible quantitative rankings were performed. In addition, the survey respondents provided a lot of valuable information in numerous free text comments. Those are included in the report in tables as they were received, without editing them, unless personal information had to be removed. Brief summaries of the main points raised are added after each section.

1 Survey context and aim

According to the 2012 WHO/UNEP publication 'State of the Science of Endocrine Disrupting Chemicals' research into endocrine disrupting chemicals over the last decade has indicated that, despite the progress achieved in development and validation of test methods for evaluation of endocrine disruptors, there are still several gaps that need to be addressed. Considering the expected significant amount of work needed to fill the gaps and the limited resources available, it will be important to set priorities for the upcoming period (next 5-10 years) for the development and validation of test methods. Thus there is a need to focus the European input to the OECD test guideline programme to effectively enhance the identification of chemical substances with endocrine disrupting properties whilst making best use of existing resources. With this objective in mind, DG Environment, supported by JRC, is organising a European expert workshop on setting priorities for further development and validation of test methods for evaluating endocrine disruption, which will be held on 30 May – 1 June 2017 in Brussels.

The workshop will produce a comprehensive overview of the existing test methods on endocrine disruptors (EDs), and in doing so identify gaps and most importantly set priorities for future work on test method development. In preparation for the workshop, JRC has drawn up a questionnaire to gather input from experts in the field on key issues the outcome of which is to be used as a basis for the discussions at the workshop. In more depth, the aim of this survey was to gather information on approaches, experiences and future directions in setting priorities for further development and validation of test methods and testing approaches for evaluating endocrine disruptors.

In the process of the workshop preparation, the survey outcome was further evaluated and is currently used to develop a thought-starter document. The thought-stater document aims at providing an overview of available test methods (TMs) as regards endocrine disrupting properties, identifying gaps in TMs for identification of known ED properties, identifying relevant existing TMs and testing approaches to be improved, identifying needed enhancements of existing TMs and testing approaches, identifying needs for development of new TMs and testing approaches and proposal for prioritisation methodology.

2 Survey design

The online EU Survey "Identifying gaps in available test methods for evaluation of endocrine disruptors" was launched on 19/05/2015 and closed on 15/06/2015. A selected group of experts (EFSA Scientific Committee and WG on EDs, ECHA ED WG and RAC, WNT (European members from OECD webpage), Experts identified in Annex 3 of the "Roadmap for setting priorities for further development and validation of test methods and testing approaches for evaluating endocrine disruptors") were invited to participate in the in the survey.

The questionnaire consisted of the following sections:

- Information on respondent;
- Priorities for the development of test methods to predict/identify substances that may cause endocrine related diseases/disorders;
- Criteria for setting priorities for test method development and enhancement;
- Coverage of endocrine pathways and related tests/endpoints that might be added and currently under discussion in OECD;
- Comments on the OECD conceptual framework for testing and assessment of endocrine disruptors (as revised in 2012).

The experts could use the predefined set of answers as well as comment in free text fields. The free text comments are shown in tables as they were received, without editing. Only personal information was removed in line with the privacy statement of the survey.

3 Survey results

3.1 Information on respondents (Questions B1-B8)

Responses were received from 40 experts representing 15 countries and different stakeholder groups (15 representing authorities; 11 academia; 3 civil society organisation; 3 industry; 8 other affiliations). For details see **Figure 1**: and **Figure 2**. 38 of 40 respondents provided their name, affiliation and e-mail address. Half of the respondents replied on behalf of their organisation and the other half in their private capacity.

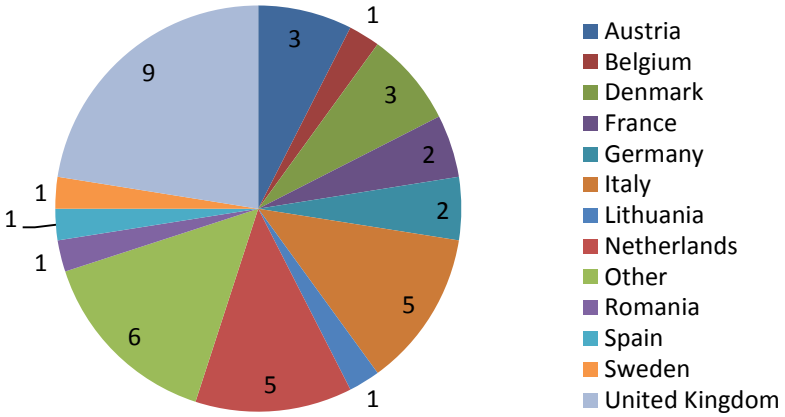

Figure 1: Country of respondent. Survey participants were from 12 different EU countries and 3 non-EU countries (indicated as "Other": USA, Switzerland, Norway).

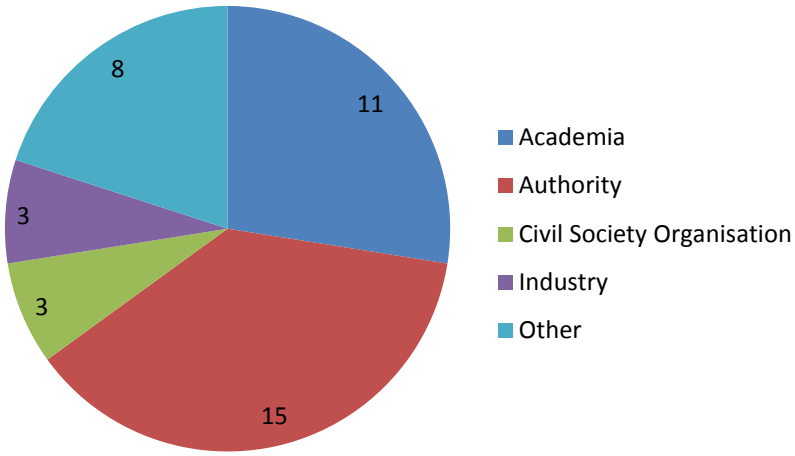

Figure 2: Survey respondents' affiliation. The category "other" was further specified by respondents as e.g. self-employed, retired, NGO, Social no profit cooperative.

3.2 Priorities for the development of test methods to predict/identify substances that may cause endocrine related diseases/disorders

The first part of the survey (survey section C) aimed at identifying the priorities for the development of test methods to predict/identify substances that may cause endocrine related diseases/disorders (**Table 1**).

Table 1: List of diseases/disorders addressed in the survey (according to WHO/UNEP, 2012)

Disease/disorder	**Examples**
Human female reproductive health	pubertal timing, fibroids, endometriosis, polycystic ovary syndrome, fecundity and fertility
Wildlife female reproductive health	uterine fibroids, lower reproductive success, egg shell thinning, imposex in snails
Human male reproductive health	testicular cancer, semen quality, cryptorchidism, hypospadias, reduced anogenital distance, testicular dysgenesis syndrome
Wildlife male reproductive health	feminising disorders, cryptorchidism, low fertility and reproduction rates
Thyroid-related disorders / neurodevelopmental disorders in children	cognitive and behavioural deficits, attention deficit/hyperactivity disorder (ADHD), autism
Hormone related cancers	breast, endometrial, ovarian, prostate, thyroid
Adrenal disorders in humans	adrenocortical hyperplasia, development of fetal adrenal cortex, effects on function of hypothalamic-pituitary-adrenal (HPA) axis
Adrenal disorders in wildlife	uterine occlusions/strictures, leiomyoma, adrenocortical hyperplasia, elevated baseline corticosterone levels
Metabolic disorders	obesity, diabetes, metabolic syndrome
Immune related disorders in humans	inflammatory disorders, immune cancers such as lymphoma and leukaemia, childhood respiratory disease
Immune related disorders in wildlife	increased susceptibility to infectious diseases

The experts were asked to indicate:

- whether any of the existing test methods could predict any of endocrine related diseases/disorders (**Table 1**; C.1);
- their priority for development of new/enhanced test methods for relevant endocrine related diseases/disorders (**Table 1**; C.2);
- any other disease/disorder of concern and not listed in **Table 1** and a need to develop new test methods (C.3).

3.2.1 Prediction of endocrine related diseases/disorders with existing test methods (C.1)

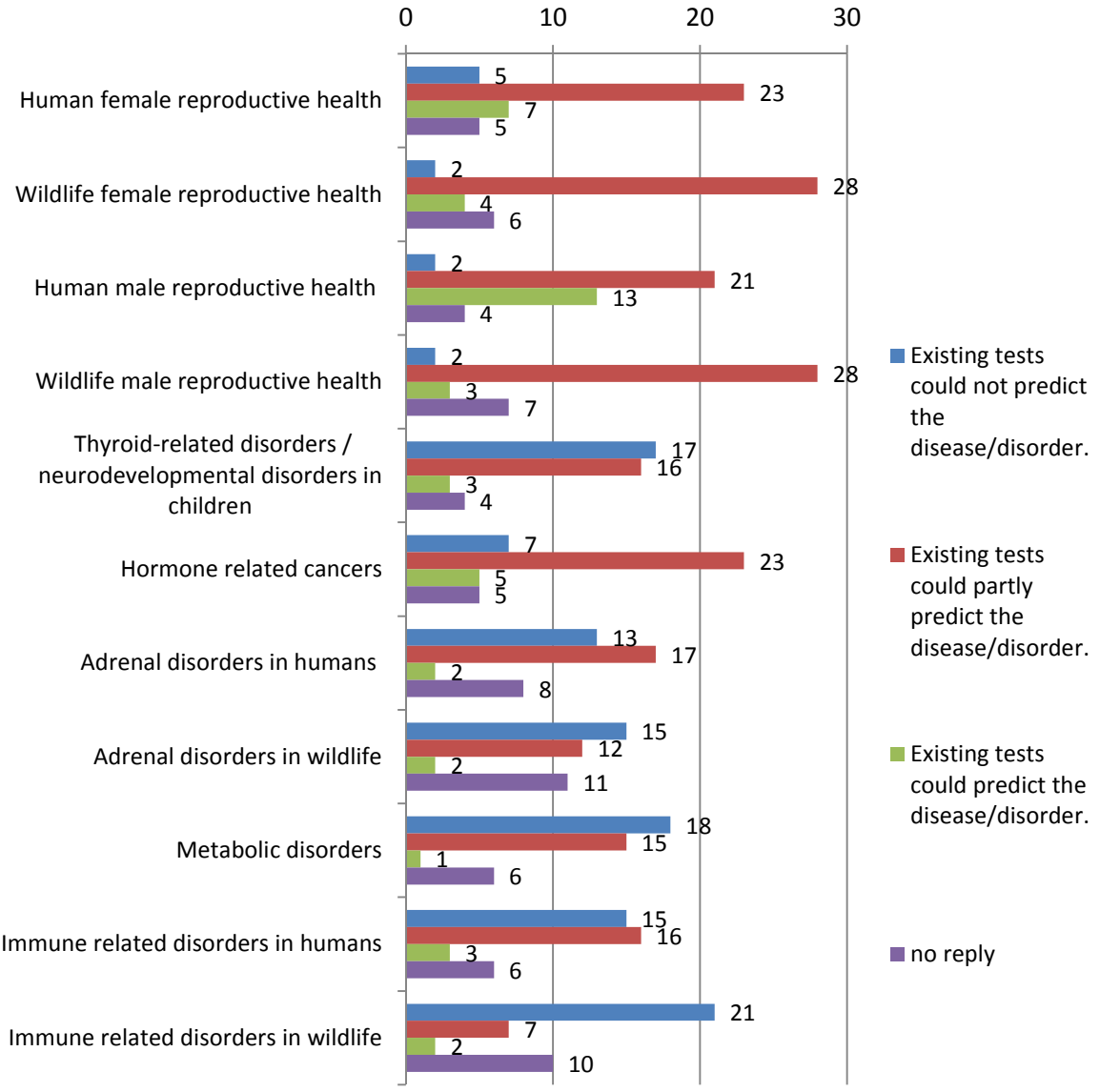

Figure 3: Replies of experts to question C.1: Are you aware of existing test methods for predicting the development of the diseases/disorders listed in Table 1?

In order to establish the priority for the areas in which test methods should be developed, JRC used the following methodology for ranking:

Percentages of answers of a certain category were calculated by dividing by the overall number of answers (excluding no replies which were 4-11 of 40) and then ratios for ranking were calculated as presented in the following equations:

(Equation 1) $\frac{\#\ existing\ tests\ could\ not\ predict\ the\ disease/disorder}{\#\ existing\ tests\ could\ predict\ the\ disease/disorder}$

(Equation 2) $\frac{\#\ existing\ tests\ could\ not\ predict\ +\ \#\ existing\ tests\ could\ partly\ predict\ the\ disease/disorder}{\#\ existing\ tests\ could\ predict\ the\ disease/disorder}$

The outcome of the ranking is given in **Table 2**.

Table 2: Ranking of replies of experts to question C.1 "Are you aware of existing test methods for predicting the development of the diseases/disorders listed below (i.e. in Table 1)?" Ranking is based on ratios of the number of answers indicating that existing test methods could not predict / could partly predict / could predict the respective disease/disorder, as described in Equation 1 and 2 above.

	$\frac{\#\ could\ not\ predict}{\#\ could\ predict}$ **(Eq 1)**		$\frac{\#\ could\ not\ or\ partly\ predict}{\#\ could\ predict}$ **(Eq 2)**	
	Ranking	Ratio	Ranking	Ratio
Metabolic disorders	1	18.0	1	33.0
Immune related disorders in wildlife	2	10.5	3	14.0
Adrenal disorders in wildlife	3	7.5	4	13.5
Adrenal disorders in humans	4	6.5	2	15.0
Thyroid-related disorders / neurodevelopmental disorders in children	5	5.7	5	11.0
Immune related disorders in humans	6	5.0	6	10.3
Hormone related cancers	7	1.4	9	6.0
Human female reproductive health	8	0.7	10	4.0
Wildlife male reproductive health	8	0.7	7	10.0
Wildlife female reproductive health	10	0.5	8	7.5
Human male reproductive health	11	0.2	11	1.8

Ranking: 1-3 highest concern; 4-6 medium concern; 7-9 lower concern; 10-11 lowest concern

According to the replies and the derived ranking, the best coverage with existing TMs is available for human and wildlife reproductive health. The highest concern with less TM availability to predict those was identified for metabolic disorders, immune related disorders in wildlife and adrenal disorders in wildlife and humans.

3.2.2 Priority for development of new/enhanced test methods for relevant endocrine related diseases/disorders (C.2)

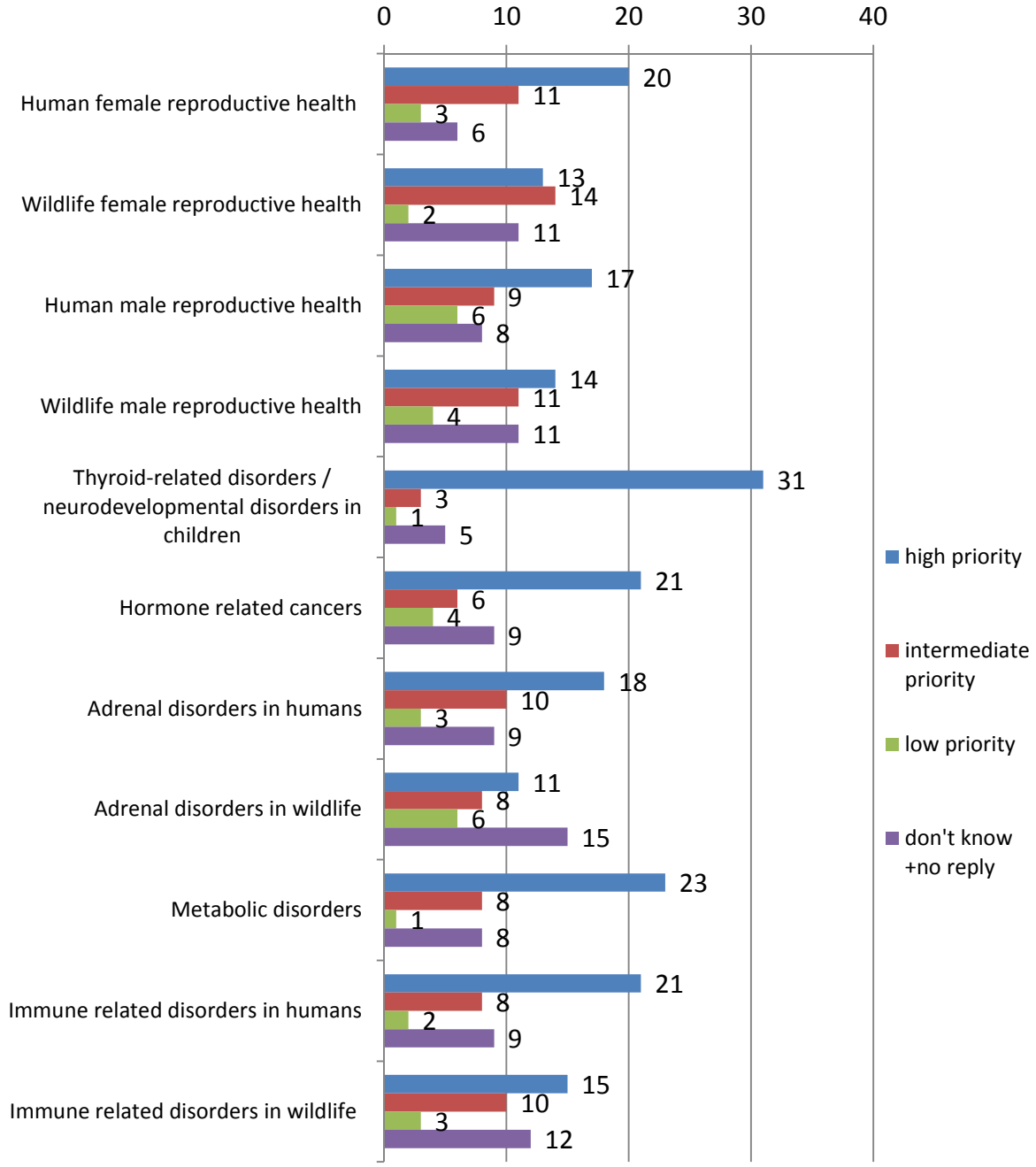

Figure 4: Replies of experts to question C.2: Please indicate your priority for developing test methods for predicting the development of diseases/disorders listed in the table below (especially for those where you answered "could partly predict" or "could not predict"; in the above table (i.e. Figure 3)). (Considering where new or enhanced test methods could lead to the greatest contribution to the protection of human health and the environment in either a short or long-term perspective).

JRC ranked the replies applying the following methodology:

Percentages of answers of a certain category were calculated by dividing by the overall number of answers (excluding no replies which were 5-15 of 40) and then ratios for ranking were calculated as presented in the following equations:

(Equation 3) $\frac{\#\ high\ priority}{\#\ low\ priority}$

(Equation 4) $\frac{\#\ high+intermediate\ priority}{\#\ low\ priority}$

Table 3: Ranking of replies to C.2: Please indicate your priority for developing test methods for predicting the development of diseases/disorders listed in Table 1. Ranking is based on ratios of the number of answers indicating high priority and intermediate priority compared to low priority, as described in Equation 3 and 4 above.

	$\frac{\#\ high\ priority}{\#\ low\ priority}$ **(Eq 3)**		$\frac{\#\ high+intermediate\ priority}{\#\ low\ priority}$ **(Eq 4)**	
	Ranking	Ratio	Ranking	Ratio
Thyroid-related disorders / neurodevelopmental disorders in children	1	31	1	34
Metabolic disorders	2	23	2	31
Immune related disorders in humans	3	11	3	15
Human female reproductive health	4	7	5	14
Wildlife female reproductive health	5	7	4	10
Adrenal disorders in humans	6	6	6	9
Hormone related cancers	7	5	8	8
Immune related disorders in wildlife	8	5	7	7
Wildlife male reproductive health	9	4	9	6
Human male reproductive health	10	3	10	4
Adrenal disorders in wildlife	11	2	11	3

Ranking: 1-3 highest priority; 4-6 medium priority; 7-9 lower priority; 10-11 lowest priority

Table 4 compares the ranking obtained for questions C.1 and C.2 regarding what could be predicted with existing test methods (C.1) and the priorities for developing new test methods (C.2). The diseases/disorders with the biggest gap in related test methods are not necessarily the ones with the highest priorities for filling these gaps. E.g. thyroid related disorders are already partly covered with the existing TMs but the highest priority for further filling the remaining gaps. In other areas like e.g. immune related disorders and adrenal disorders in wildlife large gaps were identified, however they are not under the highest priority for TM development.

Table 4: Comparison of ranking of replies to C1 and C2 as presented in **Table 2** and **Table 3**.

	C.1: Gaps in existing TM based on (Eq 1)	**C.2: Priority for filling gaps based (Eq 3)**
Thyroid-related disorders / neurodevelopmental disorders in children	5	1
Metabolic disorders	1	2
Immune related disorders in humans	6	3
Human female reproductive health	8	4
Wildlife female reproductive health	10	5
Adrenal disorders in humans	4	6
Hormone related cancers	7	7
Immune related disorders in wildlife	2	8
Wildlife male reproductive health	8	9
Human male reproductive health	11	10
Adrenal disorders in wildlife	3	11

Ranking regarding concern/priority: 1-3 highest; 4-6 medium; 7-9 lower; 10-11 lowest

3.2.3 Expert opinions on any other disease/disorder not listed and possible needs for new test methods (C.3)

In addition to the two specific questions, experts were invited to name any other disease/disorder of concern and not listed in Table 1 and a need to develop new test methods (**Table 5**).

Table 5: Free text answers to question C.3: Would you like to add any other disease/disorder of concern that is likely linked to endocrine disrupting substances and where there is a need for the development of new test methods or enhancement of existing test methods?

General comment
1. As mentioned in the introduction for the question, the levels of evidence for a link of the listed diseases with exposure to chemicals are varying. The diseases mentioned can have a variety of causes and the respective diseases are well known to be multifactorial and a causal link between such diseases and chemicals in general and endocrine disrupting substances specifically is not substantiated. Risk factors (not related to endocrine disrupters) for such diseases cannot be dismissed as cause for the diseases. This needs to be considered in balancing the need for additional test methods and in prioritization against test method development in other areas.
Comments specific for wildlife toxicity
2. Screening methods for endocrine disruption in **invertebrates**
3. Too little is known about sensitivity of (certain species of) **birds** to certain EDCs relative to other taxa. There is a need to identify which core characteristics (traits) make birds unique in their potential responses to EDCs

4. Which MOAs and associated specific endpoints give rise to effects in **birds** that would not be seen in mammals or other taxa. Which targeted, step-wise in vitro and small scale in vivo protocols should be developed to specifically address these differences between birds and other taxa.
5. thyroid related disorders in wildlife - neurodevelopmental disorders in human and wildlife
6. **Neurodevelopmental disorders** might be of interest for wild life species, as well, given the wide variety of **neurotoxicants** in the environment such as axonic excitotoxins (pyrethroids), acetylcholine esterase inhibitors (organophosphates and carbamate pesticides). This has been shown for zebrafish regarding altered locomotor behavior (Kluever et al. Environ. Sci. Technol., 2015, 49 (11), pp 7002-7011). Moreover, thyroid disruptors might have also neurodevelopmental effects affecting behavior of wild life species (various environmentally relevant thyroid disruptors such as PCBs shown to be also neurotoxicants, although the causative link is not yet fully elucidated).
7. Moreover, **thyroid disruptors** might have also **neurodevelopmental effects** affecting behavior of wild life species (various environmentally relevant thyroid disruptors such as PCBs shown to be also neurotoxicants, although the causative link is not yet fully elucidated).
Comments specific for human/mammalian toxicity:
8. PCOS, ovarian, testicular and adrenal hormonal disorders
9. Prostate adenomas and carcinomas; mammary gland tumors
10. neurodevelopmental disorders in human
11. premature reproductive senescence, not detected by current TMs
12. **epigenetic MoA** is repeatedly highlighted by OECD working groups as relevant to investigate for the ED context as a longer term goal (3x)
13. **Non-genotoxic carcinogenicity**, also beyond ED MoA, is also spotted as high priority issue by OECD WNT (2x)
14. Disorders, which are not mentioned in table C.2, are **premature senescence and cancers** arising after exposure in utero
15. EDCs and **bone disorders** have been included in Section 2.9 of the WHO/UNEP report. Bone disorders like osteoporosis have been considered as one of the metabolic diseases. Many pathways like traditional EATS pathways in the OECD CF and those included in DRP 178 like PPARs, VDR, RXR, GR etc are involved in the regulation of bone remodeling. Other pathways like TGF-ß; etc influence also bone formation and resorption. There are many in vitro, ex vivo and in vivo mammalian bone remodeling methods in the medical field. Zebrafish is also used to study bone remodeling. There is a need for the development of new test methods. Possibilities exist for enhancing existing test methods, like including histology of bone tissue in the repeated/reproductive toxicity tests, including some parameters over bone remodelling in the developmental toxicity test.
16. There are no standardized methods for **metabolic disorders** and measurements of obesity, There are apart from the DIT cohort in TG 443 no standardized methods to measure reduced immune function
17. For **thyroid-related disorders** the definitive identification of thyroid active chemicals is an effect on organ weight and histopathological changes. However, some of these changes are related to modes of action that have limited relevance for human health. Test method development should include clarification of MoAs that are non-human relevant. In addition thyroid hormone measures should be used in conjunction with changes in thyroid weight and histopathology, not as stand-alone measures. This is particularly true for measures in perinatal animals. The fluctuation of thyroid hormone levels in this age of rats (PND13) is very dynamic due to large individual differences. Pups at PND 13 are in the middle of dramatic growth period and small number of pups will be subjected to measure. Since hormone levels at PND 13 fluctuate easily based on secondary effects by growth retardation and several stress conditions, it is difficult to detect the direct effects on the endocrine system. In order to prevent unnecessary concerns, serum levels for thyroid hormones should be evaluated with

caution.
Comments on criteria for ranking
18. Highest priority in responses to C2 is given to those fields where **more mechanistic knowledge** and methods exist, **but MoA information from these methods** is incomplete and approaches for testing a larger number of substances are not yet available. As soon as IATAs have been established for these fields, the experience and resources shall be used to extend to other fields.
19. In the prioritization, it is necessary to take into account the **maturity of relevant test methods** in regard to scientific credibility, relevance, predictivity (sensitivity, specificity), readiness for standardization and foreseen regulatory use.
20. Test method development as well as regulatory acceptance and demand for these tests should be restricted to those tests that have **accepted relevance to humans** and or can be bounded by human relevant diagnostic boundaries. Efforts to develop test methods, or expand existing test methods to include additional endpoint measurements for which we have no knowledge base for understanding the biological or toxicological meaning of the outcome should be of low priority.
21. Need to improve EATS first *in vitro* so **metabolically competent**

JRC Summary of the comments in free text fields of Question C.3 regarding other endocrine related diseases

The general comments in this section were reflecting concern about the evidence for linking the listed diseases with exposure to chemicals, that might be difficult to establish as these "diseases are well known to be multifactorial. Moreover the "Risk factors (not related to endocrine disrupters) for such diseases cannot be dismissed as cause for the diseases". The experts asked for consideration of these facts in balancing the need for additional test methods and in prioritization against test method development in other areas.

The comments regarding the ecotoxicity highlighted the need of screening methods for endocrine disruption in invertebrates, differential sensitivity of certain species of birds (and birds differences to other taxa) to EDC and need to identify core characteristics (traits) that make birds unique in their potential responses to EDCs. A concern was also expressed for thyroid disrupters in context of neurodevelopmental effects, as various environmentally relevant thyroid disruptors such as PCBs show to be also neurotoxicants, although the causative link is not yet fully elucidated.

In relation to human health the epigenetic and non-genotoxic carcinogenicity MoA were highlighted as relevant to investigate in the ED context as a longer term goal.

Other disorders/diseases that were not addressed in the survey were added, i.e. premature senescence and cancers arising after exposure *in utero*. In the area of metabolic disorders, bone disorders and obesity were identified as areas with high priority for development of the new or enhanced existing test methods. Experts suggested that possibilities exist for enhancing existing test methods, like including histology of bone tissue in the repeated/reproductive toxicity tests, including some parameters over bone remodelling in the developmental toxicity test.

Additionally the lack of the standardised methods to measure reduced immune function was pointed out.

In the area of thyroid-related disorders the experts suggested that the test method development should include clarification of MoAs that are non-human relevant and that serum levels for thyroid hormones should be evaluated with caution in order to prevent unnecessary concerns.

3.3 Criteria for prioritising the development of new test methods or enhancement of existing test methods

In section D of the survey, experts were asked to:

- Rate the relevance of criteria proposed for prioritising tests relevant for ED identification when considering an overall set of test methods (D.A) and name further criteria (D.A.7).
- Rate the relevance of the criteria proposed below for prioritising tests relevant for ED identification when considering a specific test method (D.B) and name further criteria (D.B.8).

3.3.1 Rating the relevance of criteria proposed for prioritising tests relevant for ED identification when considering an overall set of test methods

The experts rated the proposed criteria (D.A.1–D.A.6) from: 1 = low relevance to 10 = high relevance and results are given in **Table 6**.

Table 6: Criteria and their ranking of relevance considered for an overall set of test methods.

criterion	ranking	relevance score		
		median	min	max
Coverage of relevant diseases/disorders (D.A.1)	1	10	5	10
Coverage of relevant life stages/exposure time windows (D.A.4)	1	10	5	10
Coverage of relevant range of species/disorders (D.A.2)	3	8	1	10
Coverage of relevant molecular targets (D.A.3)	3	8	3	10
Consideration of animal welfare (D.A.5)	5	7	1	10
Consideration of costs (D.A.6)	6	5	1	10

Note: 2 experts did not reply to this question

The coverage of the relevant diseases/disorders as well as relevant life stages/exposure time windows were rated of the highest relevance for developing or enhancing test methods. The lowest concern was given to the consideration of the costs.

Table 7: Criteria named by experts for a set of test methods in addition to those listed in **Table 6** (free text answers to question D.A.7 Would you like to comment on the above criteria or add any other criterion for a set of test methods?)

1. Especially **human relevance** is currently not well established for many models. For example, to establish neurodevelopmental effects sexual dimorphism and species differences should be taken into account, which is currently not the case.
2. **Costs has been given a lower priority** because if cost savings is implemented at the level of testing, the potential that **society will later have to pay higher medical costs** and experience lower quality of life is increased.
3. test simplicity, evaluation model, clinical relevance

4. I assume that consideration of animal welfare with a high score does not down prioritize experimental animal studies as long as they are performed according to animal welfare guidance.
5. As I hope is reflected in my answers, the principal concerns are **vulnerable windows** and **physiological outcomes**.
6. For D.A.3 (coverage of relevant molecular targets): 10 if in vitro, less if in vivo
7. D.A.5 (animal welfare) and D.A.6 (costs) are **highly prioritized** but should **not be on expense of protection of health or environmen**t. C.f. also response under point C.3
8. Consistent with the current state of the art of endocrine science, the default approach to assessing potential EDCs must **include low-dose studies** relative to human exposures and below those dose ranges used for traditional toxicity testing. Assessments should take into account that there may be no detectable threshold below which EDC can be presumed to be safe, and that potency is an inaccurate predictor for toxic effects, due to variations depending on hormonal systems and many other factors.
9. The aforementioned criteria are mixed for both in vitro and in vivo methods and sometimes difficult to clarify in the answers. For instance **animal welfare is a highly relevant** issue, but since the **WHO definition asks for information on intact animals** it is as yet unavoidable. In the future this may be extrapolated from in vitro tests. The development of harmonized TGs under MAD could in part address this issue. The same applies for **costs**, if data from one method are accepted in all OECD Member Countries this will save costs. Moreover, if we would like to include potency in the regulatory decisions of EDs it is essential that all relevant and sensitive endpoints have been addressed/measured.
10. The **sensitivity and integrity of each test method is very important** to identify endocrine disruptors. If there is even a small possibility of false positive results, the evaluation considering the weight of evidence (WoE) needs to be conducted carefully. From the perspective of **animal welfare**, unnecessary additional studies should not be required based on the results of studies with low integrity. Typically, **coverage of relevant molecular targets** in an overall set of test methods would be less of a priority versus coverage of relevant diseases/disorders and species relevance. However for endocrine disruptors **RELEVANT molecular targets** are a necessary component of the ED identification process. It is especially important to understand how **modulation of molecular targets is related to adversity** as the endocrine system is highly adaptive. All the criteria mentioned for endocrine disruptor identification and evaluation are important for prioritization based on the WHO/IPCS definition and the regulatory consequences in addition the **aspect of adversity** is key; all new test methods intended to be used in the **regulatory context should be able to define adversity** in contrast to e.g. fluctuations of responses within the physiological homeostasis.
11. Coverage of relevant data gaps - eg which **MOAs** and associated specific endpoints give rise to **effects in birds that would not be seen in mammals or other taxa**.
12. A test has to be relevant in first place for a known human ailment, irrespective of molecular targets involved. To be used it needs to be relatively **inexpensive**, and to **conform to ethical requirements**. Most of all it must be **relevant in terms of species** (human and/or wildlife) and in terms of **time window**.
13. with regard to relevant molecular targets: **Less well understood MoAs should be considered** as well - regarding relevant range of species: for wildlife: knowledge on the endocrine systems and therefore also on disorders is limited for invertebrates, birds, amphibians, reptiles, plants and microbes - with regard to **relevant life stages/windows of exposure**: tests for early exposure, late effects in life should be considered - **test systems for chronic low-dose exposure** are missing
14. Regarding D.A.1 (coverage of diseases) we consider this criteria **rather not relevant for Environment**. Regarding D.A.6 (costs) the relevance of **costs** with regard to validation exercise and development of a TG are considered
15. A priority test should **address a known biological mechanism** - this is critical for building a deeper understanding of chemical activity, predicting toxic effects, predicting cumulative, cross-species and life-stage effects.

16. Of course **animal welfare** is important in that we need to reduce where possible, and use the least number of animals to get the necessary information. Similarly, **cost** will also be a consideration, but at the end of the day a good test method will save costs.
Non-criteria/general comments:
17. We propose an integrated and comprehensive long-term toxicity/developmental/carcinogenicity bioassay capable of generating information on a broad spectrum of different end-points and relevant hypotheses. **Sprague Dawley rat**, already in use for carcinogenicity bioassays by most organizations including the Ramazzini Institute and the National Toxicology Program, and the Endocrine Disruptor Screening Program of the U.S. Enviromental Protection Agency [EPA], has been demonstrated as an appropriate and relevant model for identifying, extrapolating and predicting several toxic effects in humans. Therefore integrated toxicological tests on Sprague Dawley rat represents a unique opportunity for investigating multiple toxicological end-points at once, sparing animal lives in accordance to the 3Rs (replacement, reduction and refinement). An **integrated study design based on a stepwise process** is described, that complies and expands the state of the art of current guidelines: the Organization for Economic Co-operation and Development (OECD) Guidelines 453, 443 (OECD 2009, 2011), National Toxicology Program (NTP) Guidelines (NTP 2011a, e), and EFSA Guidelines 2013 (EFSA 2013). This strategy will provide data for more comprehensive risk assessments, by including prenatal, lactational, neonatal exposures as well as continued exposure and observation of animals up to at least 30 months (130 weeks) of lifetime post birth. This integrated study design is efficient in that the same generational cohort of rats used for studying long-term toxicity/carcinogenicity can be monitored in satellite parallel experiments designed to measure markers and parameters related to system-specific toxicities, metabolic alterations, and endocrine disturbances.
18. The correlation of endocrine disruptors and humans/wildlife is of high relevance because there are many doubts/gaps to be studied yet, and more information (as much as possible) on **species differences, molecular pathways, life stage susceptibility, and non stress conditions** for proper comparisons are important.
19. The development of MoA/AOP informed in vitro/in silico focussed IATAs with improved human or environment relevant metabolism and improved (human/environmental) relevance should be the main focus of development. This appears essential in order to assess a relevant number of substances in forseeable time, to allow regular re-testing of substances along the continously ongoing scientific progress in understanding ED MoA Testing and Assessment and also to eventually address mixture toxicity. This requires critical characterisation of the performance of actual animal testing reference methods (reproducibility and relevance for target of evaluation human health or environment) in order to correctly use the uncertainty information for the validation of new methods and to define acceptable performance of new approaches. Moreover further work for approaches to define adversity on a cellular level as well as toxicokinetic in vitro to in vivo extrapolation will be essential. The METiCx consortium from the Horizon 2020 call representing 24 scientific institutions is prepared to support these developments. (3x)
20. Test method prioritisation should be **linked with AOP development**
*See also Comments on Prioritisation listed in 4.3.2 (*Table 5*)*

3.3.2 Rating the relevance of criteria proposed for prioritising tests relevant for ED identification when considering a specific test method

The experts rated the proposed criteria (D.B.1–D.B.7) from: 1 = low relevance to 10 = high relevance and results are given in **Table 8**.

Providing significant information on at least one endocrine related perturbation and allowing to distinguish endocrine disrupting effects from general toxicity or indirect effects on the endocrine system are the most important criteria that need to be taken into account while developing/enhancing a specific test method.

Table 8: Criteria and their ranking of relevance considering a specific test method.

criterion	ranking	relevance score		
		median	min	max
The test method is providing significant information on at least one endocrine related perturbation (D.B.1).	1	9.0	4	10
The test method allows to distinguish endocrine disrupting effects from general toxicity or indirect effects on the endocrine system (D.B.2).	1	9.0	4	10
Additional endpoints that could improve the diagnostic value or sensitivity of the test method (D.B.3).	3	8.0	4	10
Additional endpoints incorporated for test method enhancement would not hamper the practical feasibility of the test (D.B.4).	3	8.0	1	10
The test method is technically feasible, could be validated or is already under validation (D.B.5).	3	8.0	1	10
The outcome of the test will be accepted for regulatory decision making without requiring in most cases further additional testing (D.B.6).	3	8.0	1	10
The test could be performed in high throughput (D.B.7).	7	6.0	2	10

Note: 3 experts did not reply to this question

Table 9: Criteria named by experts for a specific test method in addition to those listed in **Table 8** (free text answers to question D.B.8 Would you like to comment on the above criteria or add any other criterion for a specific test method?)

1. compatibility between users, **standardization**
2. More than one endpoint is usually better to understand what is happening in the entire endocrine system and generally, one single in vivo study can supply it when also performed with high throughput techniques as well. Then, a **combined study with multiple endpoints** that characterize the substance mode of action could be helpful for regulatory purposes.
3. The test needs to be relevant, sensitive and specific. Then it needs to be cheap, ethically acceptable, and at least moderate throughput.
4. Tests having the most impact- that is those that can **provide the most meaningful information with the most confidence** in the conclusion (either negative or positive)- should be prioritized the highest for future development. This means the biological underpinning supporting the tests utility for informing the question should exist. In vitro HTS will only provide limited information concerning the MIE or early key events. Long term effects/adversity as a consequence of activation of the MIE need to be established and characterized.
5. In reference to question **D.B.1.** Although it depends on the case, the identification of endocrine disruptors should not be decided from the data of the single study. Measurement of a single endocrine related perturbation should not identify a chemical as an endocrine disruptor rather it could indicate a potential to interact with the endocrine system. This is a key aspect of the EPA Endocrine Screening Program and the reason for a tiered testing strategy and battery of tests to evaluate a potential for interaction and adversity. It is also important that any test methodology designed to identify endocrine disruptors be able to differentiate between adversity due to endocrine disruption and perturbations of the endocrine system that are secondary to alternative modes of action. A directly mediated effect is a critical aspect of the WHO definition for identify endocrine disruption.
6. **D.B.2.** In order to prove that an effect is directly mediated by a certain endocrine effects of the chemical (and not by a secondary mechanism provoked by systemic toxicity or other pathways), rescue experiments might also be conducted with the co-treatment of the relevant hormone, in case not the whole AOP is known for the substance. E.g. chemicals causing reduced thyroid hormone production leading to impaired behavior might be co-treated with exogenous thyroid hormones.
7. **D.B.2.** the **differentiation between systemic toxicity and endocrine effects** is desirable but we must accept especially **invertebrate tests** where this is not always possible until endocrine specific endpoints have been developed and validated. **e.g. in the mollusc tests** under Development!
8. **Ad D.B.2;** Establishing endocrine disrupting effects should not be the primary aim, if this is not the most relevant MOA. However, when endocrine disrupting effects are suspected, appropriate test methods should be available to verify that. HT is secondary to high predictivity.
9. **D.B.3** Additional Endpoints: augmentation of existing assays with relevant in vivo biomarkers
10. **D.B.3**. **Hormonal feedback loops** might also be tested, when molecular biomarkers assessed in order to investigate the ED relevance of a certain chemical/concentration. As it is possible that an organism can "balance out" an ED effect up to a certain concentration due to e.g. a negative autoregulatory loop, in order to predict higher tier effects, these feedback effects should be taken into account. This can be tested only in in vivo systems.
11. **D.B.6** High Throughput: Important for in vitro methods
12. Explanation to **DB7 (regulatory acceptance)**: The focus should be on the development of in vitro/in silico IATAs, this may require parallel or sequential testing approaches. Also medium throughput should be acceptable. see also DA7.
13. **D.B.7** Difficult to answer because the meaning of regulatory decision making spans from very strict regulation (e.g. a ban) to less strict regulation (e.g. risk reduction or consumer advise).
14. **D.B.7**: Important for in vivo, currently not realistic for in vitro Prioritization: C.f. response under point C.3

15. **D.B.7** Of course it is preferable to have methods that are accepted in regulatory settings and that require no further testing. However, such tests would often imply high costs and a lot of testing in, different strains/species/taxa. Therefore robust screening methods that would reliably predict (and not underpredict) a potential ED mode of actions and potential adverse effects is considered highly relevant.
16. In reference to question **D.B.7**. one test cannot identify an endocrine disruptor. While it is highly relevant (relevance=10) new tests be accepted for regulatory decision making and even more ideal that confident conclusions can be drawn from a test outcome (either negative or positive), it is unclear how any one positive test can be used for a regulatory decision related to endocrine disruption without requiring additional test data to link mechanistic activity with a disease/disorder outcome
17. **D.B.7** To us this is a quite theoretical criteria. It might be very valuable to have additional tests that give further Information on the mode of Action that on their own do not allow for a regulatory decision making.
Non-criteria/general comments:
18. These criteria are difficult to answer in theory, as their relative importance depends on other characteristics of the test (e.g. over all sensitivity, relevance, animal use).
19. The integrated protocol we propose permits a reduction in terms of animal use of up to 70% if compared with the separate performance of each NTP and OECD protocol currently in use for long-term carcinogenicity and toxicity bioassays and one generation reproductive toxicity studies (NTP 2011a) (NTP 2011e; OECD 2009, 2011).
20. Obesity, diabetes and metabolic syndrome are complicated diseases with a multifactorial etiology. It will be difficult to have a unique test that will be accepted for regulatory decision.
21. Endpoints selected include biochemical and physiological endpoints which allow calibration of in vivo test results to lower tier assays more applicable to chemicals screening and prediction of effects in non-model species.
22. Metabolism needs to be incorporated, epigenetic mechanisms need to be elucidated and added. Increased focus needed on IATA development. To be honest these questions above are not helpful.
23. This exercise is not useful. We need a battery of tests to provide the necessary information.

3.4 How well are relevant pathways covered?

In section E of the survey, experts expressed their opinion on the addition of novel test methods (TM) or new endpoints to existing tests (EP) in order to cover additional pathways or better cover the pathways already included in the OECD CF. The test methods and endpoints addressed in this section were taken from the OECD DRP 178 (OECD, 2012).

Table 10 gives an overview on the overall ranking of novel test methods and endpoints across the various pathways, whereas chapters 0 - 3.4.7 cover the specific pathways.

JRC used the following methodology to calculate the overall ranking of the TMs and EPs indicated by the experts:

(Equation 5) $Ratio = \frac{very\ relevant + relevant + relevant\ but\ practical\ problems}{low\ relevance + not\ relevant}$,

identifying the relevance rating by the experts; and

(Equation 6) $Ratio = \frac{relevant\ but\ practical\ problems}{very\ relevant + relevant + relevant\ but\ practical\ problems} \times 100\ [\%]$,

determining the % of answers identifying practical problems for a relevant TM or EP.

Table 10: Overview on the overall ranking of novel test methods (TM) and endpoints (EP) across the various pathways (E.A – E.G) regarding the relevance and practical problems. Ranking is based on ratios determined according to Equations 5 and 6 above.

Pathway	**Endpoint / Test Method**	**Relevance Ranking (Eq5)**	**Practical problems [%] (Eq6)**
HPA	(EP) adrenal steroid synthesis (in vivo)	24.0	21
HPG gestagenic	(EP) reduced fertility in fish (TG229; MEOGRT)	22.0	18
PPAR	(TM) Adipocyte differentiation in cultured pre-adipocyte cells	21.0	19
HPT	(TM) neurite extension assay	14.0	29
HPT	(TM) neural progenitor cell proliferation assay	14.0	29
HPA	(EP) Stress response (in vivo)	12.5	44
Retinoid	(TM) RXR transactivation assay	12.5	12
HPG gestagenic	(TM) Progesterone receptor (PR) transactivation test	11.5	17
HPG estrogenic	(EP) Gonad histopathology in chronically exposed amphibians (TG 231 AMA; included in LAGDA)	10.5	14
HPT	(TM) XETA (Xenopus Embryonic Thyroid Signaling Assay)	9.5	11
HPT	(TM) thyroid peroxidase assay	7.7	9
HPT	(TM) iodine uptake assays	7.7	9

HPT	(TM) Tadpole tail explant resorption assay	7.5	27
retinoid	(TM) RAR transactivation assay	7.0	14
PPAR	(TM) PPARalpha,beta/delta,gamma transactivation assay	7.0	10
Retinoid	(EP) weight gain, increased adipose tissue mass, increased lipid accumulation, reduced retinoid levels in vivo (TG 415, 416, 443, fish and amphibians)	6.8	15
HPT	(TM) dendritic arborization assay	6.0	42
HPT	(EP) TH production in thyroid gland explants	6.0	33
HPA	(EP) corticotropin (ACTH) release (in vivo)	5.8	22
HPT	(TM) T4 binding protein displacement assay	5.8	13
HPG gestagenic	(TM/EP) assessment in exposed oocytes and sperm ex vivo or in oocytes/sperm derived from adult fish from TG 229 and MEOGRT	5.7	29
HPT	(TM) T-screen assay	5.7	18
HPA	(TM) adrenal steroid synthesis (in vitro), e.g. modified TG 456	5.3	19
Vitamin D	(TM) VDR transactivation assay	5.0	0
HPG gestagenic	(TM) membrane PR binding assay	4.8	21
HPG estrogenic	(EP) GnRH neuron development in brain of chronically exposed fish (MEOGRT)	4.7	14
Vitamin D	(TM) Vitamin D hydroxylase assay (in vivo)	4.3	15
PPAR	(EP) Weight gain in chronically exposed animals (TG 415, 416, 443, LAGDA)	4.2	5
HPG androgenic	(EP) behavioural assessments in any in vivo study	4.2	40
Retinoid	(EP) EROD induction in in vivo assays	4.0	5
Vitamin D	(EP) RIA or EIA for serum vitamin D levels (could potentially be applied to any in vivo exposure assay)	3.8	13
Vitamin D	(EP) reduced bone length in juvenile rodent (TG 416, 443)	3.2	31
HPA	(TM) GR transactivation test (in vitro)	3.2	21
retinoid	(TM) AhR transactivation assay	3.2	16
HPT	(TM) TR reporter assays	3.2	16
Retinoid	(EP) CYP1A mRNA or protein quantification in in vivo assays	2.9	10
PPAR	(TM) Peroxisome proliferation assay	2.8	24
HPT	(TM) AhR reporter assays	2.7	19

HPT	(TM) CAR reporter assays	2.3	14
Vitamin D	(EP) Brain size measurements in rodent offspring	1.9	46
somatotropic	(EP) Fetal birth weight and length in rodent multigeneration tests (TG 416, 443)	1.7	4
Vitamin D	(TM) AhR transactivation assay	1.6	0
Vitamin D	(EP) EROD activity assay (biomarker, could potentially be applied to any in vivo exposure assay)	1.3	38
somatotropic	(EP) analyses of hepatic GR mRNA levels in fish/mammals in vivo assays	1.1	24
somatotropic	(EP) Analyses of hepatic IGF-1 mRNA levels in fish/mammal in vivo assays	1.1	14
somatotropic	(TM) TR and GR transactivation assays	1.0	25
somatotropic	(EP) Growth evaluation in fish assays (MEOGRT)	1.0	20
HPT	(TM) EMSA, DNA pull-down assay	1.0	50

TM = test method; EP = endpoint; HPA = hypothalamus-pituitary-gonad axis; HPG = hypothalamus-pituitary-adrenocortical axis; HPT = hypothalamus-pituitary-thyroid axis; PPAR = peroxisome proliferator-activated receptor.

The experts expressed the opinions given in **Table 11** applying to all pathways.

Table 11: General comments relevant to all pathways

1. A good way to determine the appropriateness/need of novel assays is to **build relevant AOPs and identify likely key events**.
2. At the moment the main focus should be to develop and assess **in vitro/in silico** focussed **IATAs** for **EATS and non-genotoxic carcinogenicity**. As soon as IATAs have been established for these fields, the experience and resources shall be used to extend to other fields. See other comments to this inquiry.
3. Chronic exposure studies are of high relevance
4. A key element currently missing from the above approach is the ability to take into account metabolism and bioavailability in vitro, which may be addressed through the addition of exogenous metabolising systems and the improvement of **metabolism/bioavailability** prediction models (per OECD DRP 97 on the use of metabolising systems for in vitro testing of endocrine disruptors; and Jacobs, M.N., Laws, S.C., Willett, K., Schmieder, P., Odum, J., Bovee, T.F. 2013. In vitro metabolism and bioavailability tests for endocrine active substances: What is needed next for regulatory purposes? ALTEX. 30(3):331-51). This aspect needs to be addressed urgently **in relation to all axes** covered by this survey.
5. general comment for the survey - the **relevance for in vitro and in vivo can differ depending on the application of the data** - this may influence votes depending on the viewpoint of the respondent.
6. Microarrays
7. Consider OECD DRP 178, EFSA Opinion on Endocrine active substances (2013), and Kortenkamp report

3.4.1 Hypothalamus-pituitary-adrenocortical (HPA) axis (E.A)

In section E.A of the survey, experts were asked to:

- judge the addition of two test methods tests and three endpoints (listed in **Figure 5** and **Table 12**) to address effects on the HPA pathway (E.A.1)) .
- name any other TM or EP (E.A.2, **Table 13**)
- give general comments on the assessment of the HPA pathway (E.A.3, **Table 14**)

The replies were ranked with the methodology described above.

Table 12: Overview on the ranking of novel test methods and endpoints for the HPA pathway (E.A)

Endpoint / Test methods	**Relevance Ranking (Eq 5)**	**Practical Problems [%](Eq 6)**
(EP) adrenal steroid synthesis (in vivo)	24.0	21
(EP) Stress response (in vivo)	12.5	44
(EP) corticotropin (ACTH) release (in vivo)	5.8	22
(TM) adrenal steroid synthesis (in vitro), e.g. modified TG 456	5.3	19
(TM) GR transactivation test (in vitro)	3.2	21

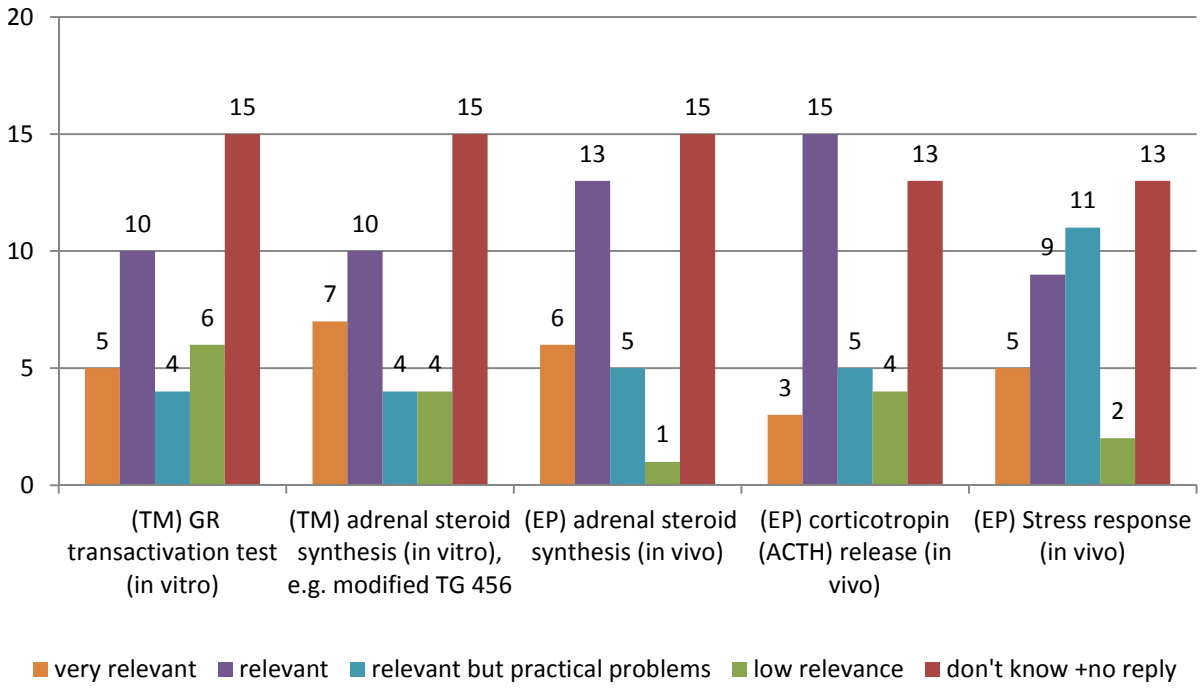

Figure 5: Replies of experts to question E.A.1: How would you judge the addition of below mentioned tests and endpoints to address effects on the HPA pathway?

Table 13: Expert suggestions for any other HPA relevant new test method or endpoint (free text replies to question E.A.2 Would you like to suggest any other HPA relevant novel test? If yes, please specify the test or relevant endpoints below including a reference)

General
1. Microarrays
2. A good way to determine the appropriateness/need of novel assays is to build HPA-relevant AOPs and identify likely key events.
Test method / endpoint proposals Humans/mammals
3. Exposure from prenatal life until 130 weeks of age; multiple WOS i.e. prenatal neonatal prepubertal pubertal adults (parous and virgin for females rodents)
4. human primary adrenal cell culture models [e.g. published for testic. cells in Hofer et al., J Clin Endo Metab 2014]
5. GR binding assays. (Commercial e.g. Polar Screen) Mineralocortocoid receptor reporter assays. (Any commercial) Human stress response in Human biomonitoring or clinical study.
6. CRH receptor agonist and antagonist assays
7. Aryl hydrocarbon Receptor (AhR) Agonist/antagonist assays
8. Steroid/corticosteroid biosynthesis assays (e.g. 11ß-hydroxysteroid dehydrogenase type 2 activity in human cells)
9. There was a SPSF on fish proposed by UK in the OECD program, which is relevant to the HPA. There is no specific SPSF on mammals at this stage. Such tests may need to be further developed.
Test method / endpoint proposals Wildlife
10. More work should be done on the value of cortisol as a stress marker; particularly in fish, this can be measured non-invasively (i.e. in the water, see numerous publications by Alexander Pickering Scott) at the back of other TGs without any major modifications. In this way we add value to the test without using extra animals.
11. Development and validation of a Test Guideline for chemicals that disrupt the hypothalamicpituitary-adrenal (interrenal) axis in vertebrates (fish)

Table 14: Experts general comments on assessing the HPA pathway (free text replies to question E.A.3 Do you have any general comment on assessing the HPA pathway in the screening for endocrine disrupting substances?)

General
1. Chronic exposure studies are of high relevance
2. functional models should be used
3. The development of IATAs for the HPA axis should be of intermediate priority. In vitro focussed approaches need to be explored for developing practical testing approaches. At OECD level much more work was already carried out for the field of EATS. In order to progress this and achieve comprehensive, regulatory useful testing and assessment approaches in shorter terms, in the moment the main focus should be to develop and assess in vitro/in silico focussed IATAs for EATS and non-genotoxic carcinogenicity. As soon as IATAs have been established for these fields, the experience and resources shall be used to extend to other fields. See other

	comments to this inquiry.
4.	Yes, it is a very important pathway, affecting directly and indirectly pretty much the whole spectrum of the endocrine system. We have the obligation to look at this in more detail in terms of test method development and validation
5.	Information and test aimed at the HPA pathway are of high relevance and importance because they will help to distinguish chemical specific effects from general stress response. Carefully designing the experimental protocol is critical for the method development for this pathway.
6.	Here, importance needs to be given to tests addressing the establishment of the HPA axis and the ED modulation of set-points.
7.	Primary resource focus needs to be ensuring adequate test coverage for EATS, other related endocrine endpoints can be addressed next.
8.	We suggest that efforts to assess the effect of substances on the HPA axis should focus on building mechanistic understanding of its pathways, and that as such the development of assays probing specific, defined biological events will offer more progress than development of further tests for apical endpoints in animals. In the absence of an improved understanding of the underlying biological mechanisms the complexity of the HPA pathway, its interconnections with other endocrine axes, its susceptibility to a range of environmental factors, and its homeostatic responsiveness will continue to make it difficult to draw clear conclusions from apical studies, and particularly to separate out effects on this pathway from effects on other pathways and non-endocrine mediated effects. Given that there already exists a broad complement of animal studies likely to catch HPA related adverse effects, priority should be given to the development and validation of mechanistic assays, without which further apical assays are unlikely to add much useful information. Data from mechanistic assays will be of immediate use for informing on endocrine mode of action in individual substance assessments, but can also be used to substantiate and further develop adverse outcome pathways for HPA axis disruption, in turn informing prioritization of tests probing additional key events. Initially these assays may be a combination of in vivo and in vitro methods but as AOPs are more fully described, the aim should be creation of an in vitro test battery capable of predicting apical consequences through integrated testing approaches. This will help shift regulatory programs from needing repeatedly to measure effects of individual substances in multiple species, towards scenarios where adverse effects can be predicted for humans and a variety of wildlife species, and ultimately, mixture and additive effects can begin to be tackled. A key element currently missing from the above approach is the ability to take into account metabolism and bioavailability in vitro, which may be addressed through the addition of exogenous metabolising systems and the improvement of metabolism/bioavailability prediction models (per OECD DRP 97 on the use of metabolising systems for in vitro testing of endocrine disruptors; and Jacobs, M.N., Laws, S.C., Willett, K., Schmieder, P., Odum, J., Bovee, T.F. 2013. In vitro metabolism and bioavailability tests for endocrine active substances: What is needed next for regulatory purposes? ALTEX. 30(3):331-51). This aspect needs to be addressed urgently in relation to all axes covered by this survey.
9.	The HPA axis is central to homeostatic functions of all vertebrate species, including metabolism, growth, immune system function, and reproduction. The HPA axis can be affected by many types of environmental factors, including light, temperature and diet, as well as by intrinsic circadian rhythms. This centrality and complexity is often used as a rationale for testing chemical effects in vivo - as the complexity and interrelated factors co-exist in the intact animal. However, this also creates enormous difficulty in interpreting experimental results and unraveling cause and effect relationships from incidental, secondary, or unrelated effects. Understanding chemical effects on central biological systems such as endocrine and hormone systems is therefore particularly suited to Integrated Approaches to Testing and Assessment (IATA) based on Adverse Outcome Pathways and related information. Therefore, for developing an chemical effective assessment approach, it is critical to focus on assays that address defined, specific biological mechanisms. Initially, these assays will likely be a combination of in vitro and in vivo methods - until the pathways are described well enough to allow prediction of outcome from upstream events sufficiently to answer the regulatory question at hand. Several of these mechanistic assays are likely to address more than one pathway (e.g. AhR, PPARa). A priority should be building and informing endocrine-related pathways, so that both toxicologists and regulators can better understand the biological landscape they are attempting to assess. In addition, a focus on mechanistic understanding

and pathway building would lead regulatory programs from needing to repeatedly measure effects toward the ability to predict effects in humans and wildlife. In addition, there is already a broad spectrum of tests that address apical endpoints that would catch possible adverse effects; priority should be on tests that provide mechanistic information and inform AOPs. Ideally, an assessment program would be built of a battery of mechanistic tests, largely in vitro, that will predict downstream apical effects. A major currently missing element of this approach is the ability to take into account metabolism, by improving prediction models and adding exogenous systems (see OECD DRP 97: DETAILED REVIEW PAPER ON THE USE OF METABOLISING SYSTEMS FOR IN VITRO TESTING OF ENDOCRINE DISRUPTORS; Jacobs, M.N., Laws, S.C., Willett, K., Schmieder, P., Odum, J., Bovee, T.F. 2013. In vitro metabolism and bioavailability tests for endocrine active substances: What is needed next for regulatory purposes? ALTEX. 30(3):331-51).
Humans/mammals
10. In the event that a histopathological finding in adrenal is observed, it is difficult to evaluate whether it causes the adrenal dysfunction or not. It may be an adaptive change to maintain the homeostasis but not an adverse effect. To understand the relationship of histopathological findings and functional effects is very difficult. For example transportation of animals from test cages to necropsy room and method of sacrifice (i.e. cervical dislocation vs. CO2) is known to influence stress response signaling. Though these should be randomly distributed across treatment groups the increased variability adds complexity to the data interpretation
11. The corticosteroid metabolism in animals differs in part from the humans, animal testing is not always appropriate, therefore it would be desirable to have data from experiments with human cell lines and from Human biomonitoring studies or clinical studies if available.
Wildlife
12. This is an important area for development because we now know that the hormone response of some wildlife to stress is being damaged by environmental pollutants
13. UK will be reviewing the HPA axis for fish at the OECD.

3.4.2 Hypothalamus-pituitary-gonad (HPG) axis (E.B)

In section E.B of the survey, experts were asked to:

- judge the addition of test methods tests and endpoints (listed in **Figure 6** and **Table 15**) to address effects on the estrogen signaling pathway (E.B.1)) .
- name any other TM or EP (E.B.2) for the estrogen signaling pathway (**Table 16**)
- judge the addition of test methods tests and endpoints (listed in **Figure 7** and **Table 15**) to address effects on the androgen signaling pathway (E.B.3)) .
- name any other TM or EP (E.B.4) for the androgen signaling pathway (**Table 17**)
- judge the addition of test methods tests and endpoints (listed in **Figure 8** and **Table 15**) to address effects on the gestagenic signaling pathway (E.B.5)) .
- name any other TM or EP (E.B.6) for the gestagenic signaling pathway (**Table 18**)
- express general comments on the assessment of the HPG pathway (E.B.7, **Table 19**)

The replies to E.B.1, E.B.3, and E.B.5 were ranked with the methodology described above (Equation 5) and the results are given in **Table 15**.

Table 15: Overview on the ranking of novel test methods and endpoints for the HPG pathway (E.B)

Pathway	Endpoint or Test Method	Relevance Ranking (Eq 5)	Practical Problems [%](Eq 6)
HPG gestagenic	(EP) reduced fertility in fish (TG229; MEOGRT)	22.0	18
HPG gestagenic	(TM) Progesterone receptor (PR) transactivation test	11.5	17
HPG estrogenic	(EP) Gonad histopathology in chronically exposed amphibians (TG 231 AMA; included in LAGDA)	10.5	14
HPG gestagenic	(TM/EP) assessment in exposed oocytes and sperm ex vivo or in oocytes/sperm derived from adult fish from TG 229 and MEOGRT	5.7	29
HPG gestagenic	(TM) membrane PR binding assay	4.8	21
HPG estrogenic	(EP) GnRH neuron development in brain of chronically exposed fish (MEOGRT)	4.7	14
HPG androgenic	(EP) behavioural assessments in any in vivo study	4.2	40

Figures 6-8 display the replies of the experts to questions E.B.1, E.B.3, and E.B.5.

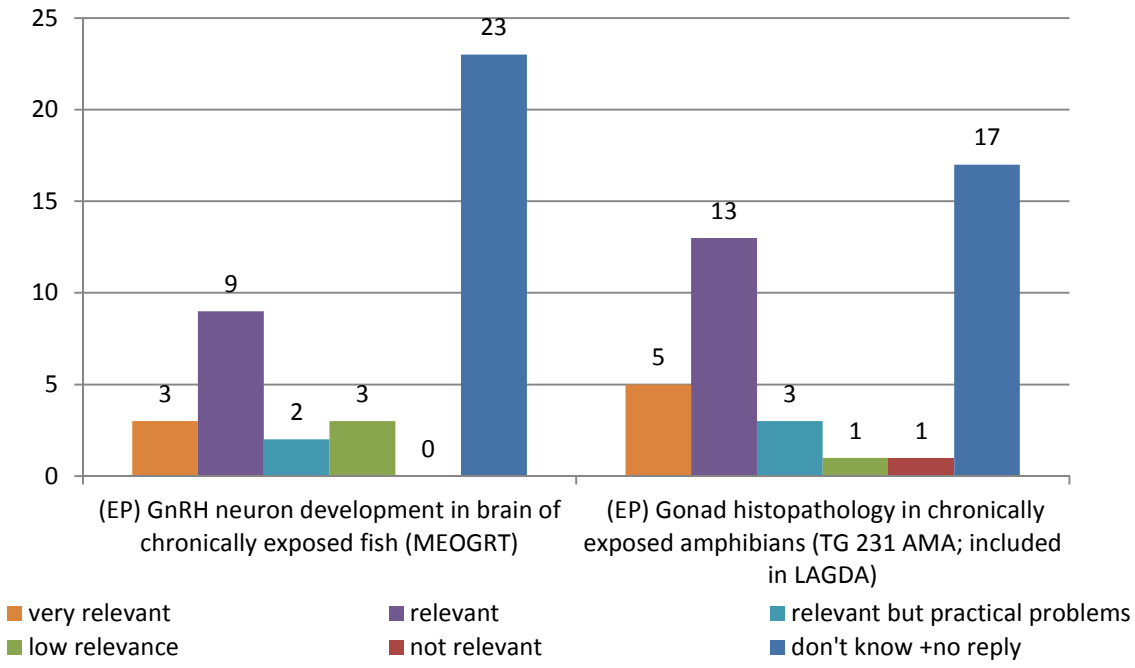

Figure 6: Replies of experts to question E.B.1 How would you judge the addition of below mentioned tests and endpoints to address effects on the estrogen signaling pathway?

Table 16: Expert suggestions for any other estrogen signaling pathway relevant new test method or endpoint (free text replies to question E.B.2 Would you like to suggest any other novel test relevant to the estrogen signaling pathway?)

General
1. Exposure from prenatal life until 130 weeks of age; multiple WOS i.e. prenatal neonatal prepuberal puberal adults (parous and virgin for females rodents)
2. Since estrogen signalling in vivo mostly acts through non-classical genomic pathways, albeit via classical receptors, then in vitro tests need to be included which specifically address these pathways and, for example, can target SERM effects as well as classical endpoints. The listed reference is not yet sensitive enough but could be made so (Mol Cell Endocrinol. 2007. 276:45-54. A novel molecular assay to discriminate transcriptional effects caused by xenoestrogens. Koohi MK, Walther N, Ivell R.)
3. To our understanding behavioural assessment is missing.
4. inclusion of metabolism: see Jacobs et al 2013 ALTEX
Test method / endpoint proposals Human/Mammalian
5. human primary testicular (and ovarian) mixed cell culture models [Hofer et al., J Clin Endo Metab 2014]
6. human membrane ER activity; The endocrine mechanism of action is heavily covered by existing in vitro and in vivo assay, therefore, priorities for development of new assays lie elsewhere.
Test method / endpoint proposals Wildlife
7. The regulation of HPG axis in fish models used for testing (i.e. medaka, zebrafish and fathead minnow) has no resemblance with the regulation of the HPG axis in the vast majority of fish (>95% of species) that have an annual reproductive cycle; we used these models because they are convenient, not representative and as such we do not protect the environment.

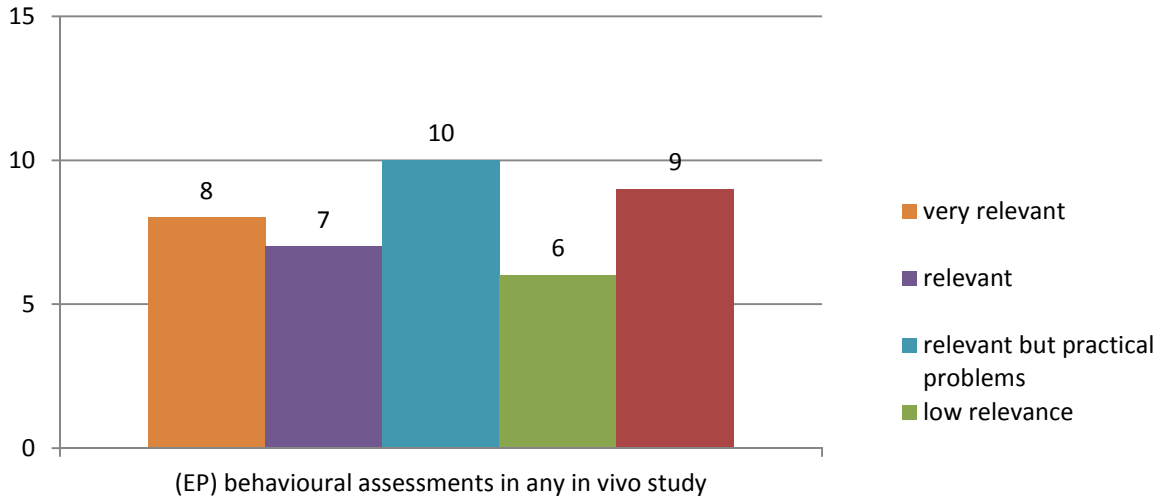

Figure 7: Replies of experts to question E.B.3 How would you judge the addition of below mentioned tests and endpoints to address effects on the androgen signaling pathway?

Table 17: Expert suggestions for any other androgen signalling pathway relevant new test method or endpoint (free text replies to question E.B.4 Would you like to suggest any other novel test relevant to the androgen signaling pathway?)

General
1. As above, systems targeting non-classical pathways are of great relevance here
2. inclusion of metabolism see Jacobs et al ALTEX 2013
3. mechanistic tests for anti-androgenic activity; probably need toxicogenomic studies to develop AOP and identify relevant biomarkers for additional assays.
Test method / endpoint proposals Human/Mammalian
4. Sperm aneuploydia Epididymal Sperm Aneuploidies in Three Strains of Rats (X. R. Lowe,1 J. M. de Stoppelaar,2 J. Bishop,3 M. Cassel,1 B. Hoebee,2 D. Moore II,1 and A. J. Wyrobek1*.Environmental and Molecular Mutagenesis 31:125-132) (1998) Detected by Multicolor Fluorescence In Situ Hybridization
5. PSA production in human, prostate derived cell line LNCAP: Lorenzetti S, Marcoccia D, Narciso L, Mantovani A. (2010) Cell viability and PSA secretion assays in LNCaP cells: a tiered in vitro approach to screen chemicals with a prostate-mediated effect on male reproduction within the ReProTect project. Reprod Toxicol. 30(1):25-35.; Smeriglio A, Trombetta D, Marcoccia D, Mantovani A, Lorenzetti S. Intracellular distribution and biological effects of plant bioactives in a sex steroid-sensitive model of human prostate adenocarcinoma. Anticancer Agents in Medicinal Chemistry 2014 14:1386-96
6. see above, mixed human primary testicular cell culture model (Leydig and Sertoli cells) [Hofer D et al. J Clin Endo Metab 2014]
7. Assessment of testosterone levels in male foetuses in TG 414
Test method / endpoint proposals Wildlife
8. The one assay that has not gained regulatory status as yet (is only a guidance document) is

that is highly informative and relevant is the spiggin assay (in the form of the androgenised female stickleback screen, AFSS. This must be converted into a TG asap; The recently developed spiggin medaka transgenic model can be incorporated to be used as a screen before the AFSS.

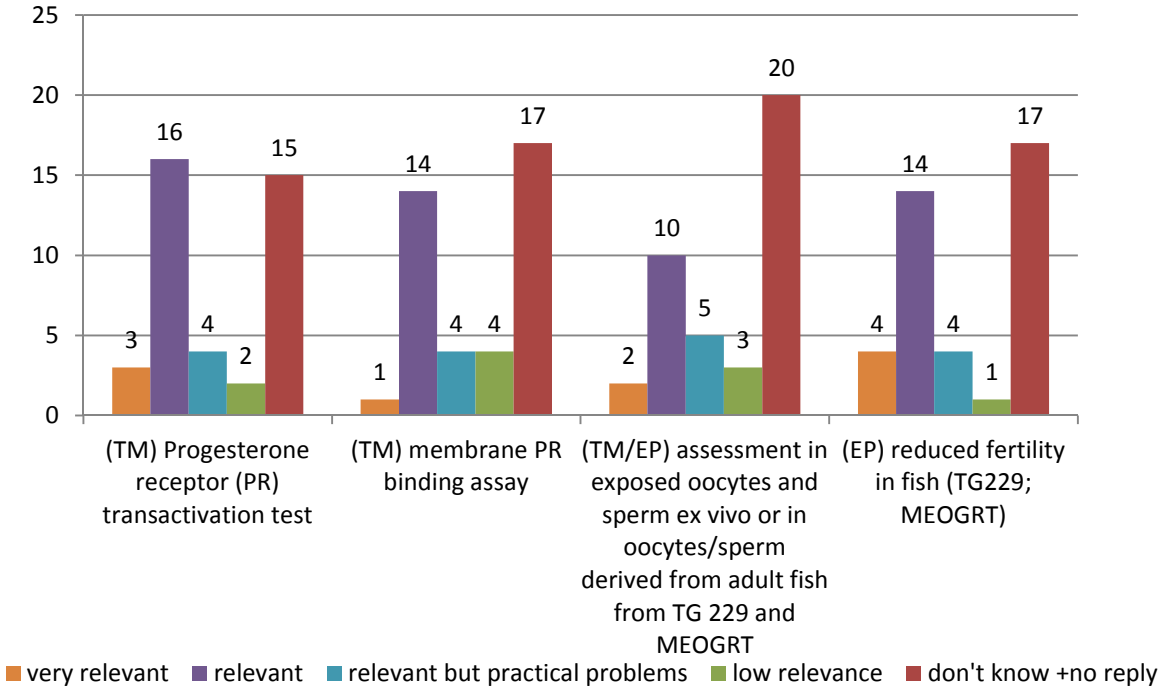

Figure 8: Replies of experts to question E.B.5 How would you judge the addition of below mentioned tests and endpoints to address effects on the gestagenic signaling pathway?

Table 18: Expert suggestions for any other gestagenic signaling pathway relevant new test method or endpoint (free text answers to question E.B.6 Would you like to suggest any other gestagenic signaling pathway relevant novel test?)

General
1. Ex vivo investigation of calcium signalling in human sperm could also be relevant: Schiffer, C. et al. Direct action of endocrine disrupting chemicals on human sperm. EMBO Rep. (2014). doi:10.15252/embr.201438869
2. The relevance of the listed assays is unknown and reduced fertility is non-specific. It is unlikely that gestagenic effects would be missed given adequate coverage of E and A pathways.
Test method / endpoint proposals Wildlife
3. The problem with TG229 is that it requires the fish to be in an optimal reproductive state before the test begins; effects then are not easily assessed within the 21d window; we need more time to observe the whole spectrum of effects and unravel the mechanism

Table 19: Expert general comments on assessing the HPG pathway (free text answers to the question E.B.7 Do you have any general comment on assessing the HPG pathway in the screening for endocrine disrupting substances?)

General
1. Clearly, the PR is relevant for many endocrine processes, but the added value of a single PR transactivation test (or any other single NR) is limited.
2. Gene expression in human primary endocrine cell culture models and whole blood RNA expression covers a large range of relevant genes as "liquid biopsies" - pathway/biomarker identification and evaluation is ongoing
3. In silico/in vitro focussed IATAs for the HPG axis, including improved metabolism and human/environmental relevance should be a high priority field for development. OECD progress is already available in this field. See earlier comments and references in the final remarks. The METiCx consortium is prepared to support these developments.
4. Assesment of oxytocin and AMH (anti- Mullerian hormone) should also be considered. **Oxytocin** (OXT) (Greek, "quick birth") is a mammalian hormone that also acts as a neurotransmitter in the brain. In addition to its well-known peripheral hormonal functions (i.e., induction of labour and milk ejection), OXT acts as an important neuronal messenger within the brain regulating social and emotional behaviours in a wide variety of animal species including humans(Lee et al 2009). Oxytocin plays an organisational role in the central nervous system and the oxytocin neural system is also thought to be involved in the underlying mechanisms that guide the development of social behaviours (e.g. maternal behaviour), reproduction and stress responses (Mogi et al 2014). References: Lee HJ, Macbeth AH, Pagani JH, Young WS. Oxytocin: the great facilitator of life. Prog Neurobiol. 2009;88(2):127-151. Mogi K, Ooyama R, Nagasawa M, Kikusui T Effects of neonatal oxytocin manipulation on development of social behaviors in mice. Physiol Behav. 2014 Jun 22;133:68-75. doi: 10.1016/j.physbeh.2014.05.010. Epub 2014 May 21. Since the serum levels of **AMH** has been shown to correlate with follicle pool in rats (Yeh et al. 2007) it could have the potential to be used as an endpoint in toxicological tests when assessing exhaustion of follicle reserve due to chemical insult. Indeed, it has been used as a biomarker to assess the degree of ovarian damage caused by exposure to the chemotherapeutic agent cisplatin (Yeh et al. 2006). Cisplatin is a very potent chemical and it must be assumed that most endocrine disrupting chemicals exerts much more subtle effects on the follicle pool. If the AMH-measurement is to be a useful biomarker of effects on ovarian follicle pool of endocrine disrupters, it could be argued that it must be able to detect the effects caused by less potent chemicals than Cisplatin. References: Yeh J, Kim B, Liang YJ, Peresie J (2006) Mullerian inhibiting substance as a novel biomarker of cisplatin induced ovarian damage. Biochem Biophys Res Commun 348:337-344 Yeh J, Kim B, Peresie J, Liang YJ, Arroyo A (2007) Serum and ovarian Mullerian inhibiting substance, and their decline in reproductive aging. Fertil Steril 87:1227-1230 98
5. There are more molecular pathways involved in the HPG axis. One example is activin, inhibin, and follistatin. Nitrobenzene is a known chemical targeting inhibins. The development of endpoints/methods for these additional pathways should be taken into account.
6. None of the above mentioned tests address the establishment of the HPG axis during early pregnancy, or at puberty, i.e. during the dynamic phases of endocrine establishment.
7. higher relevance within next years
8. Please see response to E.A.3: comments re. the advantages of prioritising development of mechanistic assays to provide mode of action data and elucidation of AOPs also apply to assessment of the HPG pathway. Numerous apical animal tests capable of catching adverse effects on reproductive systems already exist, but mechanistic assays for non-nuclear-receptor mediated modes of action are what is lacking.
9. Similar to the HPA axis, it is critical to focus on assays that address defined, specific biological mechanisms. Initially, these assays will likely be a combination of in vitro and in vivo methods - until the pathways are described well enough to allow prediction of outcome from upstream events sufficiently to answer the regulatory question at hand. Several of these mechanistic assays are likely to address more than one pathway (e.g. AhR, PPARa). A priority should be building and informing endocrine-related pathways, so that both toxicologists and regulators

can better understand the biological landscape they are attempting to assess. In addition, a focus on mechanistic understanding and pathway building would lead regulatory programs from needing to repeatedly measure effects toward the ability to predict effects in humans and wildlife. In addition, there is already a broad spectrum of tests that address apical endpoints that would catch possible adverse effects; priority should be on tests that provide mechanistic information and inform AOPs. Ideally, an assessment program would be built of a battery of mechanistic tests, largely in vitro, that will predict downstream apical effects. A major currently missing element of this approach is the ability to take into account metabolism, by improving prediction models and adding exogenous systems (see OECD DRP 97: DETAILED REVIEW PAPER ON THE USE OF METABOLISING SYSTEMS FOR IN VITRO TESTING OF ENDOCRINE DISRUPTORS; Jacobs, M.N., Laws, S.C., Willett, K., Schmieder, P., Odum, J., Bovee, T.F. 2013. In vitro metabolism and bioavailability tests for endocrine active substances: What is needed next for regulatory purposes? ALTEX. 30(3):331-51).
Human/Mammalian
10. As a general comment it is important to understand that many hormones e.g. FSH, LH, testosterone, are secreted in pulses in rodents. Thus, when measuring these hormones in serum of rodents a high variability is to be expected and the group size needs to be high enough (20 to have a 80% chance to see a 25% change) to make sure to cover a real effect (see e.g. Vidal et al 2013, reproductive system and mammary gland, in Toxicologic pathology (edited by Sahota, Popp, Hardisty, Gopinath, CRC).
11. Behavioral assessments are complex and can be highly variable.
12. The relevance of findings to human adversity is difficult to characterize and they are only likely to pick up compounds with profound effects which would be identified by other test parameters. Before developing new test methods a review of the relevance of the existing methods (retrospective analysis of data etc) would be beneficial.
Wildlife
13. Fish models are not representative of real species with annual reproduction and as such regulation. We must face this!

3.4.3 Somatotropic axis (E.C)

In section E.C of the survey, experts were asked to:

- judge the addition of test methods tests and endpoints (listed in **Figure 9** and **Table 20)** to address effects on the somatotropic signalling pathway (E.C.1) .
- name any other TM or EP (E.C.2) for the somatotropic signalling pathway (**Table 21**)
- express general comments on the assessment of the somatotropic pathway (E.C.3, **Table 22**)

The replies were ranked with the methodology described above and the results are given in **Table 20)**.

Table 20: Overview on the ranking of novel test methods and endpoints for the somatotropic pathway (E.C)

Endpoints or Test methods	Relevance Ranking (Eq 5)	Practical Problems [%](Eq 6)
(EP) Fetal birth weight and length in rodent multigeneration tests	1.7	4

(TG 416, 443)		
(EP) analyses of hepatic GR mRNA levels in fish/mammals in vivo assays	1.1	24
(EP) Analyses of hepatic IGF-1 mRNA levels in fish/mammal in vivo assays	1.1	14
(TM) TR and GR transactivation assays	1.0	25
(EP) Growth evaluation in fish assays (MEOGRT)	1.0	20

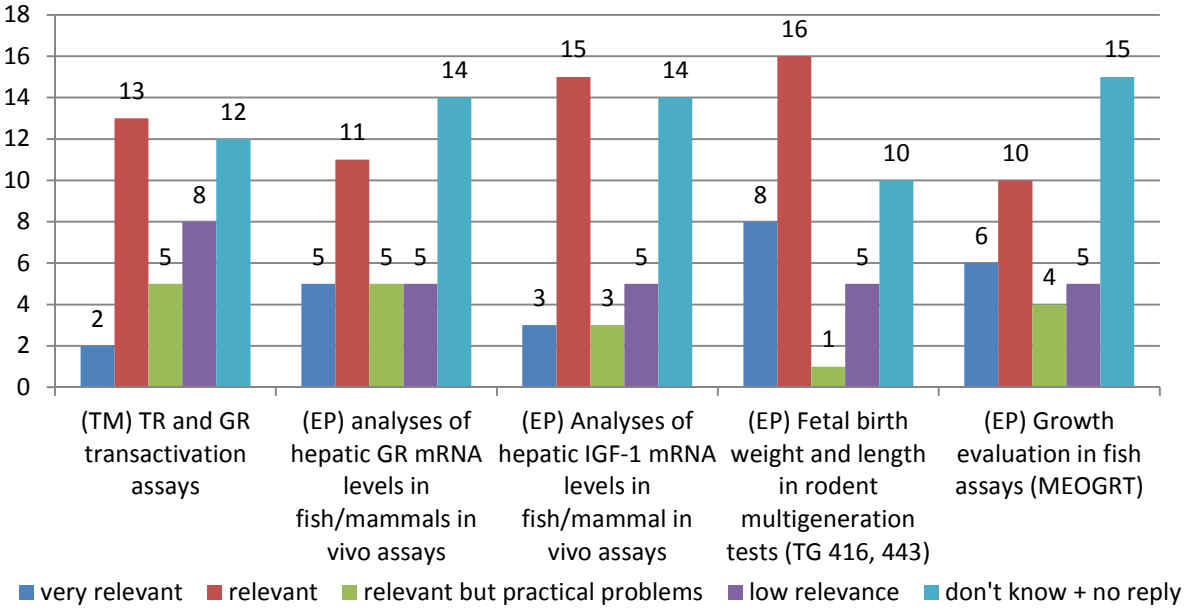

Figure 9: Replies of experts to question E.C.1 How would you judge the addition of below mentioned tests and endpoints to address effects on the somatotropic pathway?

Table 21: Expert suggestions for any other somatotropic pathway relevant new test method or endpoint (free text answers to the question E.C.2 Would you like to suggest any other novel test relevant for the somatotropic axis? If yes, please specify the test or relevant endpoints below including a reference.)

JRC Note: no specific test named but references were given:

Kortenkamp et al., State of the Art Assessment of Endocrine Disrupters, 2011 http://ec.europa.eu/environment/chemicals/endocrine/pdf/sota_edc_final_report.pdf

EFSA Scientific Opinion on the hazard assessment of endocrine disruptors http://www.efsa.europa.eu/de/search/doc/3132.pdf

Table 22: Expert general comments on assessing the somatotropic pathway (free text answers to the question E.C.3 Do you have any general comment on assessing the somatotropic axis in the screening for endocrine disrupting substances?)

General
1. Today IATAs for this axis should be of lower priority, see earlier comments. However where there is specific interest and resources are available, the OECD DRP 178 provides a starting point for further developments.
2. Chemicals interfering with pathways of thyroid, corticosteroids etc may disrupt the somatotropic axis. Methods and endpoints to be developed should indicate the specific effects on the somatotropic axis and should be used for the regulatory identification of EDCs.
3. In in vivo studies, it is hard to distinguish the secondary (indirect) effects derived from the deterioration of physical condition and the primary (direct) effects on the HPG axis. The identification of endocrine disruptors should be conducted in combination with some in vitro and in vivo studies but not a single study. For example, it is well known severely reduced body weight gain or renal toxicity frequently causes the secondary effects such as hormone imbalance by the deterioration of physical condition. Since growth alteration is not necessarily endocrine specific, addition of this endpoint is not relevant for assessing the somatotropic axis. Assessment of this axis is of low relevance. The impact of EDCs on this axis is unknown and the promiscuous nature of chemicals means that if the chemical in question is capable of causing ED then it will be detected during the assessment of other more relevant axes.
4. Dynamic tests are required which will mimic more the influence of EDcs on the pubertal and prepubertal growth spurts.
5. As effects on the somatotropic axis are likely to occur via interactions with estrogen, androgen, thyroid or corticosteroid signaling, tests aimed at this axis per se are not a priority. Building AOPs for these other pathways should help identify whether there are additional key events that need addressing.
6. Is unlikely this system will be missed given coverage of E, A, T and corticosteroid systems, plus lack of chemical MoA for this pathway suggests this pathway is a low priority for assay development. Potential missing Key Events could be identified by building the relevant AOPs for this pathway.
Wildlife
7. MEOGRT: In order to differentiate endocrine effects from systemic toxicity, more somatotropic pathway rel. endpoints need to be addressed in the hypothesized adverse outcome pathway (lower tiered e.g. molecular endpoints)

3.4.4 Retinoid signalling pathway (E.D)

In section E.D of the survey, experts were asked to:

- judge the addition of test methods tests and endpoints (listed in **Figure 10** and **Table 23**) to address effects on the retinoid signalling pathway (E.D.1) .
- name any other TM or EP (E.D.2) for the retinoid signalling pathway (**Table 24**)
- express general comments on the assessment of the retinoid pathway (E.D.3, **Table 25**)

The replies were ranked with the methodology described above and the results are given in **Table 23**.

Table 23: Overview on the ranking of novel test methods and endpoints for the retinoid signaling pathway (E.D)

Endpoints or Test methods	Relevance Ranking (Eq 5)	Practical Problems [%](Eq 6)
(TM) RXR transactivation assay	12.5	12
(TM) RAR transactivation assay	7.0	14
(EP) weight gain, increased adipose tissue mass, increased lipid accumulation, reduced retinoid levels in vivo (TG 415, 416, 443, fish and amphibians)	6.8	15
(EP) EROD induction in in vivo assays	4.0	5
(TM) AhR transactivation assay	3.2	16
(EP) CYP1A mRNA or protein quantification in in vivo assays	2.9	10

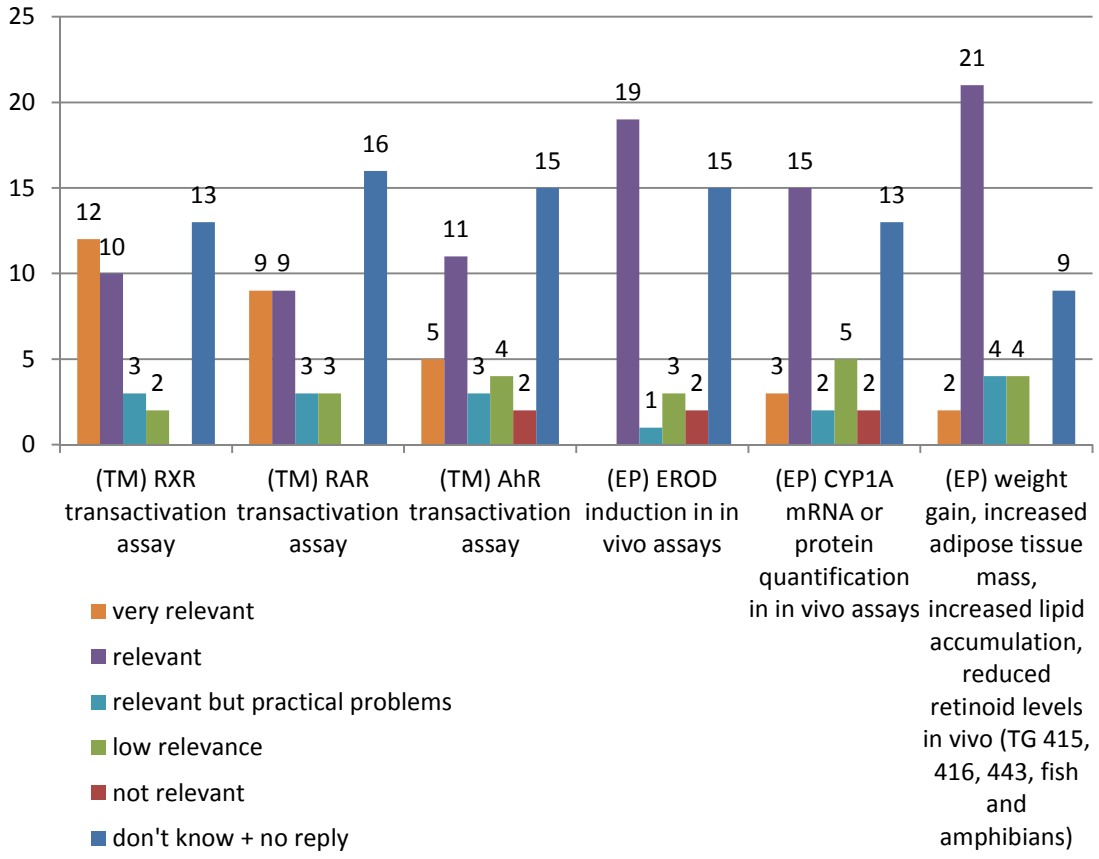

Figure 10: Expert replies to question E.D.1 How would you judge the addition of below mentioned tests and endpoints to address effects on the retinoid signalling pathway?

Table 24: Expert suggestions for any other retinoid signalling pathway relevant new test method or endpoint (free text answers to question E.D.2 Would you like to suggest any other novel test relevant for the retinoid signalling pathway?)

General
1. This is already being explored in a new DRP on retinoic acid signalling, and is being led by Sweden. Advisable to wait for outcome.
2. DRP 178 is a starting point.
Test method / endpoint proposals Human/Mammalian
3. droplet formation in 3T3-L1 cells, PPARgamma transactivation (Pereira-Fernandez et al., 2013
4. Adipocyte differentiation in vitro where mechanism can be evaluated (OECD DRP 178).
Test method / endpoint proposals Wildlife
5. We need an invertebrate model there; they are by far more sensitive and informative

Table 25: Expert general comments on assessing the retinoid signalling pathway (free text answers to question E.D.3 Do you have any general comment on assessing the retinoid signalling pathway in the screening for endocrine disrupting substances?)

General
1. Sweden has initiated a SPSF for the retinoid system within the OECD
2. IATAs for the Retinoid Acid axis should be of intermediate priority. A detailed review paper for testing and assessment in this field is planned at OECD WNT by SE. At OECD level much more work was already carried out for the field of EATS. In order to progress this and achieve comprehensive, regulatory useful testing and assessment approaches in shorter terms, in the moment the main focus should be to develop and assess in vitro/in silico focussed IATAs for EATS and non-genotoxic carcinogenicity. As soon as IATAs have been established for these fields, the experience and resources shall be used to extend to other fields. See other comments to this inquiry.
3. Retinoic acid concentration in vivo in serum is very low and quite difficult to measure. Careful concentration on the correct marker is necessary before proposing something like this.
4. RXR transactivation; The ligand specificity of RXRs is not so high and they can be activated by many natural components such as oleic acid, one of the major fatty acid in the mammalian body, in addition to 9-cis RA, DHA (Lengqvist et. al, Molecular & Cellular Proteomics 3.7 692 (2004)). Therefore, it is easily expected that many positive result would be obtained by this test method. It may be difficult to predict overall outcome in vivo by the "positive" compound in the presence of 9-cis RA and abundant natural fatty acid.
5. AhR transactivation; "The mechanism in which retinoid levels is reduced by AhR agonists is not fully understood" (ENV/JM/MONO (2012) 23). Their primary target is AhR, but not RAR/RXR, suggesting indirect mechanism. An effect on retinoid pool would accompany other toxic effects by AhR. AhR affects so many biological functions. It is difficult to know whether an effect on retinoid level is an independent effect or a secondary effect by other toxicity. EROD induction, CYP1A mRNA; These are markers for AhR agonist activity (ENV/JM/MONO (2012) 23).
6. Weight gain, increased adipose: These are not specific marker for RAR/RXR signaling. "Maintenance of lipid homeostasis in the whole organism is complex and changes in lipid would not definitively indicate the involvement of RXR" (ENV/JM/MONO (2012) 23). Systemic toxicity can secondary change these parameters. In addition, many of these endpoints can be affected by ligands for nuclear receptors such as PPAR and FXR, and natural ligands for PPAR and FXR are not hormones. Thus, positive result for these endpoints might include effects other than RAR/RXR pathway. Many of endpoints listed in the table above are lacking specificity for RAR/RXR signaling pathway. In addition, a possible in vivo effect by a compound with RAR/RAR agonist/antagonist would be hard to be predicted in the presence of homeostasis (e.g. by

natural agonists for RXR). These test results would be useful for understanding MOA of the chemicals that exerts in vivo effect such as developmental abnormalities, but less useful for screening EDs.
7. Please see response to E.A.3: comments re. the advantages of prioritising development of mechanistic assays to provide mode of action data and elucidation of AOPs also apply to assessment of the retinoid signaling pathway. As aspects of the retinoid signaling pathway, RXR signaling in particular, have such a central role in numerous endocrine processes, assays providing mechanistic data on this pathway could be considered of especially high relevance and prioritised for further development/standardisation. DRP 178 suggests RAR, RXR and AhR reporter assays could be used to discern anchoring molecular events triggering assessment along multiple adverse outcome pathways, and that RXR reporter assays should have a prominent role in any endocrine screening program. Apical in vivo measurements such as altered lipid or retinoid levels are unlikely to support firm conclusions re. disruption of retinoid signaling in the absence of data from mechanistic assays as above, as DRP 178 also indicates.
Human/Mammalian
8. see above, use of human liver cell cultures
Wildlife
9. retinoids might have environmental relevance as it has been shown that some cyanobacteria are able to produce and release retinoid-like compounds into the environment at concentrations equivalent to those causing e.g. teratogenicity in zebrafish (A. Jonas et al. Aquatic Toxicology 155 (2014) 283-290). Furthermore, the crosstalk between the retinoic and thyroid signaling pathways underlines also the environmental relevance of the retinoid signaling pathway for wild life species.

3.4.5 Hypothalamus-pituitary-thyroid (HPT) axis (thyroid signalling pathway)(E.E)

In section E.E of the survey, experts were asked to:

- judge the addition of test methods tests and endpoints (listed in **Figure 11** and **Table 26** to address effects on the HPT pathway (E.E.1) .
- name any other TM or EP (E.E.2) for the retinoid signalling pathway (**Table 27**)
- express general comments on the assessment of the retinoid pathway (E.E.3, **Table 28**)

The replies were ranked with the methodology described above and the results are given in **Table 26**.

Table 26: Overview on the ranking of novel test methods and endpoints for the HPT pathway (E.E)

Endpoints or Test methods	**Relevance Ranking (Eq 5)**	**Practical Problems [%](Eq 6)**
(TM) neurite extension assay	14.0	29
(TM) neural progenitor cell proliferation assay	14.0	29
(TM) XETA (Xenopus Embryonic Thyroid Signaling Assay)	9.5	11
(TM) thyroid peroxidase assay	7.7	9
(TM) iodine uptake assays	7.7	9

(TM) Tadpole tail explant resorption assay	7.5	27
(TM) dendritic arborization assay	6.0	42
(EP) TH production in thyroid gland explants	6.0	33
(TM) T4 binding protein displacement assay	5.8	13
(TM) T-screen assay	5.7	18
(TM) TR reporter assays	3.2	16
(TM) AhR reporter assays	2.7	19
(TM) CAR reporter assays	2.3	14
(TM) EMSA, DNA pull-down assay	1.0	50

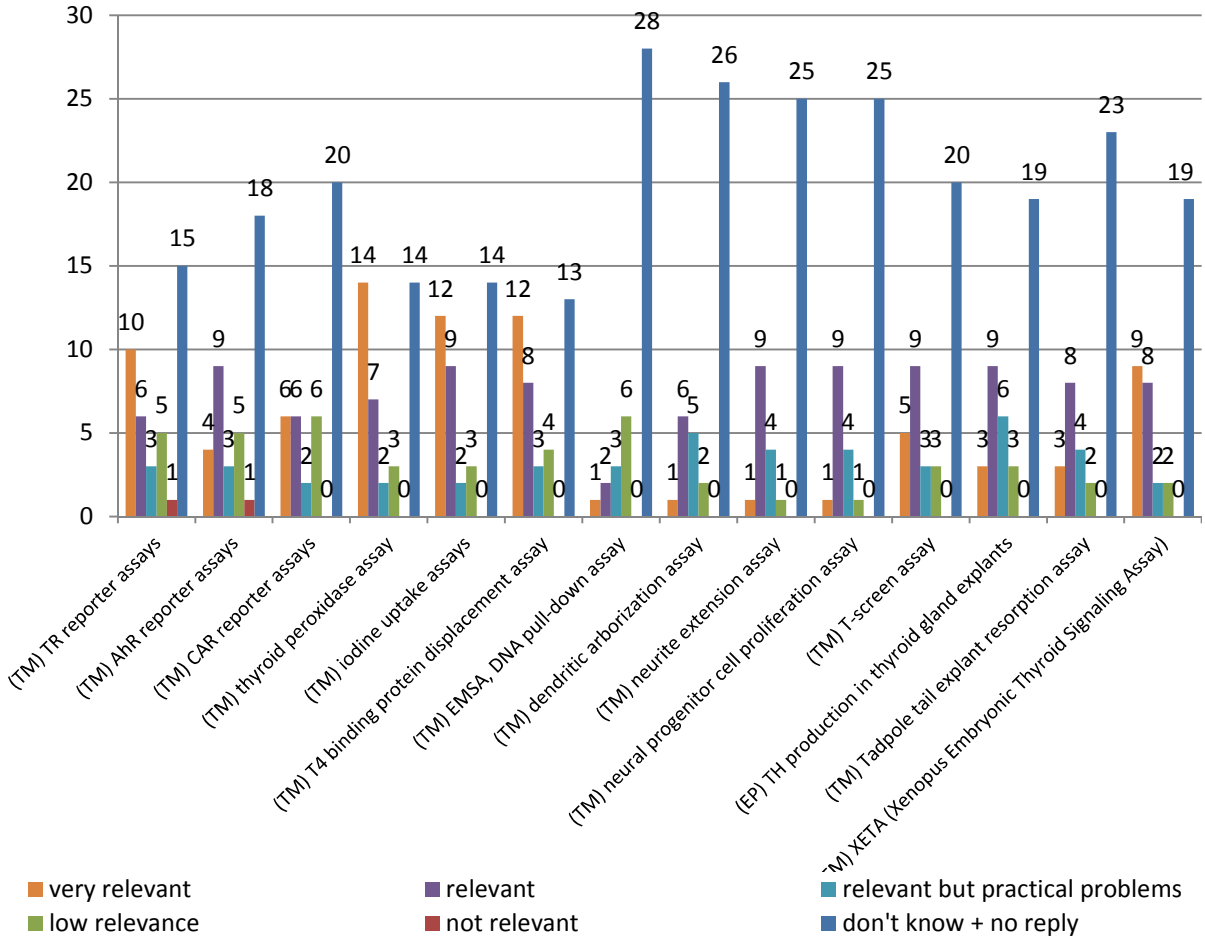

Figure 11: Expert replies to question E.E.1 How would you judge the addition of below mentioned tests and endpoints to address effects on the HPT pathway?

Table 27: Expert suggestions for any other HPT pathway relevant new test method or endpoint (free text answers to question E.E.2 Would you like to suggest any other novel test relevant for the HPT axis?)

General
1. This is the holy grail in endocrinology test method development; we need more models than amphibians!
2. See OECD Thyroid scoping document DRP 207 (*comment received 4x*)
Test method / endpoint proposals Human/Mammalian
3. Stably transfected human TR reporter assays (same for other reporter assays); TRH release assay, TRH receptor activation assay in cell line; T3, T4 deiodinization assays; TH membrane transporter activity assay. See OECD DRP 207: NEW SCOPING DOCUMENT ON IN VITRO AND EX VIVO ASSAYS FOR THE IDENTIFICATION OF MODULATORS OF THYROID HORMONE SIGNALLING.
Test method / endpoint proposals Wildlife
4. Similarly to XETA, alternative fish embryonic assays with high throughput potential and integrative capacity for thyroid disruptors with multiple mode of action are available (Fetter et al. Reprod Toxicol. 2015 May 8. p: S0890-6238(15)00067-2 and Thienpont et al Environ. Sci. Technol. 2011, 45, 7525-7532).

Table 28: Expert general comments on assessing the HPT axis (free text answers to question E.E.3 Do you have any general comment on assessing the HPT axis in the screening for endocrine disrupting substances?)

General
1. general comment for the survey - the relevance for in vitro and in vivo can differ depending on the application of the data - this may influence votes depending on the viewpoint of the respondent.
2. Only few compounds seem to interact with the TR, despite QSAR-based predictions. More relevant seems production of thyroid hormones and interactions with transport/metabolism.
3. The development of IATAs for the HPT axis should be of high priority. A great amount of OECD work is available defining the priority needs in the field. See reference in final remarks and earlier comments. The METiCx consortium is prepared to support this work.
4. The development/standardisation of biologically relevant in vitro tests as well as IATA development is highly needed.
5. No other than that the complexity of this system requires an open approach; I would start from massive parallel sequencing to fish out candidate biomarkers in existing animal models
6. Most of the assays are in vitro assays assessing one specific molecular initiating action of thyroid disruption. For compounds with unknown or multiple mode of action on the thyroid signaling pathway (such as phenyltiourea which is a thyroidperoxidase and a deiodinase inhibitor) higher tier effects might be underestimated.
7. A well developed pre validation plan has been developed by the METiCx consortium
8. Please see response to E.A.3: comments also apply to assessment of the HPT axis. In addition to building on the thyroid scoping document and taking forward the assays therein identified as in a high state of readiness for Test Guideline development, current work on thyroid AOPs within the OECD programme should be used to provide a framework for guiding prioritisation.
9. This is a critical pathway that has immediate need for AOP development and screening-level assay development. Previous comments (i.e. E.A.3) also apply.
10. We need screens and tests which include endpoints of thyroid hormone action.

Human/Mammalian
11. In the event that the effects on thyroid gland are observed in in vivo studies using rodent (e.g. rats, mice), the possibility that those are secondary effects via liver toxicity is considered to be comparatively high. Many chemicals are known as a hepatic microsomal enzyme inducer, and induce UGT, which enhances metabolism of T4 by conjugation and biliary excretion of the conjugated hormone. Induction of hepatic microsomal enzyme results in a decrease in T3 and T4 half-life. As a result of the feedback function for decreased circulating thyroidal hormones on the hypothalamic-pituitary-thyroid axis, the pituitary gland enhances the release of TSH. Therefore, it is anticipated the validated test methods to verify the secondary effects via liver toxicity are necessary. Addition of thyroid hormone measurements in dams and pups would be problematic. First, in pups, insufficient information is available to fully understand inherent variability, etc. While thyroid hormones in adult males typically have fairly low variability, it is expected one would see greater variability in the pups, and likely in the females as well. It should be remembered that lactating females are in a slightly hypothyroid condition. This could result in a decreased potential for detecting effects due to variability, etc? Also, the females will be receiving the test material for differing durations due to differences in co-housing duration prior to confirmed pregnancy. In addition, as written, if the TGs are being used for non-oral studies, where administration may stop on GD19, the animals will be without test substance exposure for 2 weeks prior to blood collection, which will likely diminish any ability to detect test substance-related effects. If thyroid hormones are to be measured, it would be prudent to understand these issues prior to inclusion in the test guideline. An alternative is to measure thyroid hormones in the P1 males, which will not have as many variables to confound data interpretation, and would still provide information on whether a typical pattern of thyroid hormone modulation is occurring. The other concern is that MOST, i.e., >90%, of compounds that alter liver weight will cause transient effects on thyroid hormones that do not necessarily manifest in long-term thyroid hormone modulation leading to adverse thyroid effects. Given that, there are large concerns over how this data would be interpreted and used for regulatory decision making
Wildlife
12. I marked the TR assay of low relevance because of the very high specificity of the TR LBD for T3 (and TRIAC). Current knowledge suggests that the TR is not a primary target - but more chemicals affects I- uptake, organification and distribution/ metabolism.
13. Regarding XETA, only those substances can be assessed that act through the thyroid hormone receptor. However, given that this is an in vivo system, mechanisms leading to altered thyroid hormone level, might also be assessed

3.4.6 Vitamin D Signalling Pathway (E.F)

In section E.F of the survey, experts were asked to:

- judge the addition of test methods tests and endpoints (listed in **Figure 12** and **Table 29** to address effects on the Vitamin D signalling pathway (E.F.1) .
- name any other TM or EP (E.F.2) for the Vitamin D signalling pathway (**Table 30**)
- express general comments on the assessment of the Vitamin D signalling pathway (E.F.3, **Table 31**)

The replies were ranked with the methodology described above and the results are given in **Table 26**.

Table 29: Overview on the ranking of novel test methods and endpoints for the on the Vitamin D signalling pathway (E.F)

Endpoints or Test methods	Relevance Ranking	Practical Problems

	(Eq 5)	[%](Eq 6)
(TM) VDR transactivation assay	5.0	0
(TM) Vitamin D hydroxylase assay (in vivo)	4.3	15
(EP) RIA or EIA for serum vitamin D levels (could potentially be applied to any in vivo exposure assay)	3.8	13
(EP) reduced bone length in juvenile rodent (TG 416, 443)	3.2	31
(EP) Brain size measurements in rodent offspring	1.9	46
(TM) AhR transactivation assay	1.6	0
(EP) EROD activity assay (biomarker, could potentially be applied to any in vivo exposure assay)	1.3	38

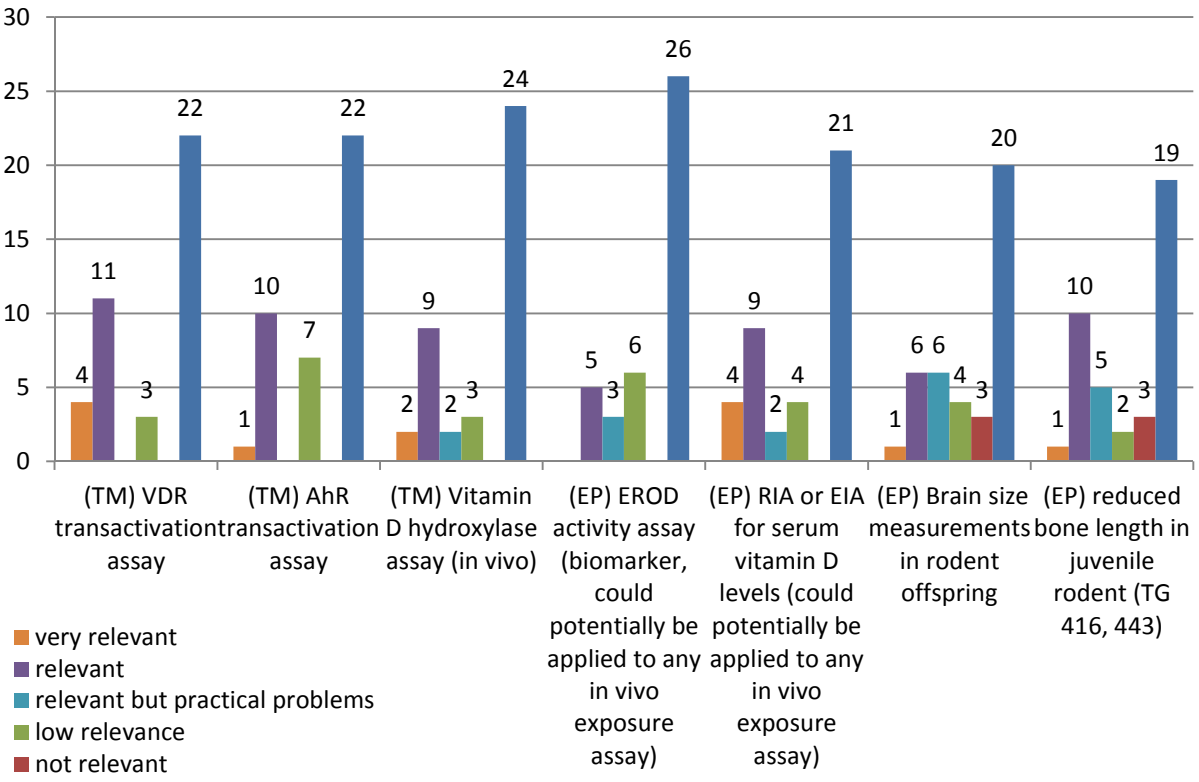

Figure 12: Expert replies to question E.F.1 How would you judge the addition of below mentioned tests and endpoints to address effects on the Vitamin D signaling pathway?

Table 30: Expert suggestions for any other Vitamin D signaling pathway relevant new test method or endpoint (free text answers to question E.F.2 Would you like to suggest any other novel test relevant for the Vitamin D signalling pathway?)

General
1. Vitamin D binding protein and other vitamin D associated genetics, NGS

Table 31: Expert general comments on assessing the Vitamin D signalling pathway (free text answers to question E.F.3 Do you have any general comment on assessing the Vitamin D signalling pathway in the screening for endocrine disrupting substances?)

General
1. very relevant, hereditary and environmental factors under development [Trummer O et al. J Clin Endo Metab 2012, Saternus R et al, Endocrinology 2015]
2. Today the development of IATAs for this axis should be of lower priority. However where there is specific interest and resources are available, the OECD DRP 178 provides a starting point for further developments.
3. not as urgent as EATS
4. Please see response to E.A.3: comments also apply to assessment of the Vitamin D signalling pathway.
5. As with the above axes, it is critical to focus on assays that address defined, specific biological mechanisms. Initially, these assays will likely be a combination of in vitro and in vivo methods - until the pathways are described well enough to allow prediction of outcome from upstream events sufficiently to answer the regulatory question at hand. Several of these mechanistic assays are likely to address more than one pathway (e.g. AhR, PPARa). A priority should be building and informing endocrine-related pathways, so that both toxicologists and regulators can better understand the biological landscape they are attempting to assess. In addition, a focus on mechanistic understanding and pathway building would lead regulatory programs from needing to repeatedly measure effects toward the ability to predict effects in humans and wildlife. In addition, there is already a broad spectrum of tests that address apical endpoints that would catch possible adverse effects; priority should be on tests that provide mechanistic information and inform AOPs. Ideally, an assessment program would be built of a battery of mechanistic tests, largely in vitro, that will predict downstream apical effects. A major currently missing element of this approach is the ability to take into account metabolism, by improving prediction models and adding exogenous systems (see OECD DRP 97: DETAILED REVIEW PAPER ON THE USE OF METABOLISING SYSTEMS FOR IN VITRO TESTING OF ENDOCRINE DISRUPTORS; Jacobs, M.N., Laws, S.C., Willett, K., Schmieder, P., Odum, J., Bovee, T.F. 2013. In vitro metabolism and bioavailability tests for endocrine active substances: What is needed next for regulatory purposes? ALTEX. 30(3):331-51).
Human/Mammalian
6. Measurement of vitamin D in serum is not a good biomarker due to its low half-life in serum. The correct biomarker for assessing vitamin D status is 25-hydroxy vitamin D3 (see relevant peer-reviews e.g. Ross et al 2011, Dietary Reference Intakes for Calcium and Vitamin D, ISBN 978-0-309-16394-1). Further RIA or EIA methods do not deliver accurate measurements, LC-MS is much better.

3.4.7 PPAR Signalling Pathway (E.G)

In section E.G of the survey, experts were asked to:

- judge the addition of test methods tests and endpoints (listed in **Figure 13** and **Table 32** to address effects on the PPAR signalling pathway (E.G.1) .

- name any other TM or EP (E.G.2) for the PPAR signalling pathway (**Table 33**)
- express general comments on the assessment of the PPAR signalling pathway (E.G.3, **Table 34**)

The replies were ranked with the methodology described above and the results are given in **Table 32**.

Table 32: Overview on the ranking of novel test methods and endpoints for the on the PPAR signaling pathway (E.G)

Endpoints or Test methods	Relevance Ranking (Eq 5)	Practical Problems [%](Eq 6)
(TM) Adipocyte differentiation in cultured pre-adipocyte cells	21.0	19
(TM) PPARalpha,beta/delta,gamma transactivation assay	7.0	10
(EP) Weight gain in chronically exposed animals (TG 415, 416, 443, LAGDA)	4.2	5
(TM) Peroxisome proliferation assay	2.8	24

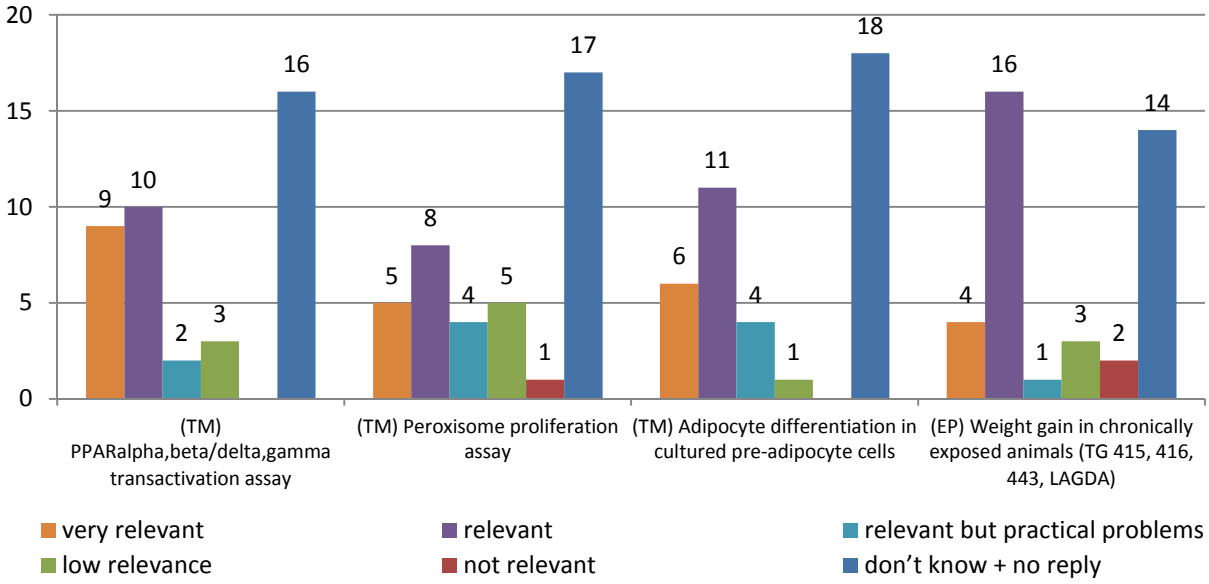

Figure 13: Expert replies to question E.G.1 How would you judge the addition of below mentioned tests and endpoints to address effects on the PPAR signaling pathway?

Table 33: Expert suggestions for any other PPAR signaling pathway relevant new test method or endpoint (free text answers to question E.G.2 Would you like to suggest any other novel test relevant for the PPAR signalling pathway?)

General
1. Weight gain in animals may not significantly change, however they may have an increase in adiposity. Fat weight should be measured.

Table 34: Expert general comments on assessing the PPAR signalling pathway (free text answers to question E.G.3 Do you have any general comment on assessing the PPAR signalling pathway in the screening for endocrine disrupting substances?)

General
1. Today the development of IATAs for this axis should be of lower priority. However where there is specific interest and ressources are available, the OECD DRP 178 provides a starting point for further developments.
2. A species comparative approach is being taken in the VMG-NA. Important for metabolic syndrome, lipid metabolism etc. but lower priority compared to EATS Generally, this survey is fragmented and piecemeal in its approach, should instead be approaching the testing needs as part of an IATA approach, then one can suggest priorities for test method development in a more substantive and integrated way.
3. Please see response to E.A.3: comments also apply to assessment of the PPAR signalling pathway. The potential anchoring role of PPAR events in a number of AOPs (e.g. connections with HPG axis, and interaction with RXR) suggest assays probing PPAR mechanisms would be useful to prioritise. Plus, as noted in DRP 178, species differences are of significance with this axis so it will be essential to understand underlying mechanisms before results of animal studies can be interpreted more usefully. Again, there are various PPAR AOPs under development within the OECD programme and these should be used to guide prioritisation of relevant assays for further work.
4. This is a critical pathway that has immediate need for AOP development and screening-level assay development. Previous comments (i.e. E.A.3 and E.D.3) also apply.
Human/Mammalian
5. The relevance of certain endpoints following PPAR activation are not relevant to humans. It will be important to differentiate human relevant effects from those that have been deemed to be non-human relevant. Activation of PPARs have also been shown to have beneficial effects in humans. This is not to say activation of PPARs by environmental chemicals is beneficial, but it highlights the importance of associating activation with adversity. In general assessment of this axis is of low relevance. The impact of EDCs on this axis is unknown and the promiscuous nature of chemicals will lead to the detection of ED in other pathways if the chemical in question is capable of causing ED.

3.4.8 Potential use of epigenetic tests within the OECD endocrine disruptor testing conceptual framework

This part F.E of the survey was included under the questions on the OECD Conceptual Framework in the survey section F. However, it should have been addressed in section E together with the other new approaches that might be relevant for inclusion in the OCED CF. Therefore in the report we address it here below.

In Annex 1 of the Detailed Review Paper (OECD DRP, No. 178, 2012) it is concluded that the evidence "is highly-suggestive of a role for epigenomic dysregulation mediating the effects of exposures to endocrine disruptors". Therefore experts were asked about their views on including epigenetic endpoints in the OECD CF. Possible endpoints that were considered in the survey to be included in the tests are listed below (for more details see the Detailed Review Paper):

- DNA modifications (cytosine methylation)
- miRNA and RNA expression studies
- studies of chromatin components and structure
- multivariate/systems analysis to identify key regulatory factors
- luminometric methylation analysis (LUMA) for global methylation analyses

An overview of TMs that could potentially be adapted for epigenomic studies of effects of endocrine disruptors, was presented in the survey and is shown in **Table 35** below based on the overview in the OECD DRP (2012).

In this section F.E of the survey, experts were asked to:

- rate the relevance of updating tests in the OECD CF to include epigenetic endpoints/tests (**Figure 14**)
- provide any comments regarding the inclusion of epigenetic tests for the identification of endocrine disruptors (**Table 36**).

Table 35: OECD TGs that could potentially be adapted for epigenomic studies of effects of endocrine disruptors (Table 7 of Annex 1 of the detailed review paper OECD No 178, OECD 2012)

Type of study	*Test Guidelines (TG)*	*Description*
		• Zebrafish embryo epigenetic toxicity assay
General exposure studies	• TG 451 • TG 452 • TG 453	• Carcinogenicity Studies • Chronic Toxicity Studies • Combined Chronic Toxicity/Carcinogenicity Studies
Post-mitotic cell studies	• TG 424	• Neurotoxicity Study in Rodents
Prenatal effects	• TG 414 • TG 426	• Prenatal Development Toxicity Study • Developmental Neurotoxicity Study
Reproductive effects	• TGs 415, 416 • TG 421 • TG 422 • TG 443	• One and Two-Generation Reproduction Toxicity • Reproduction/Developmental Toxicity Screening Test • Combined Repeated Dose Toxicity Study with the Reproduction/Developmental Toxicity Screening Test • Extended One-Generation Reproductive Toxicity Study
Potentially relevant tests to be used in combination	• TG 473	• *In vitro* Mammalian Chromosome Aberration Test

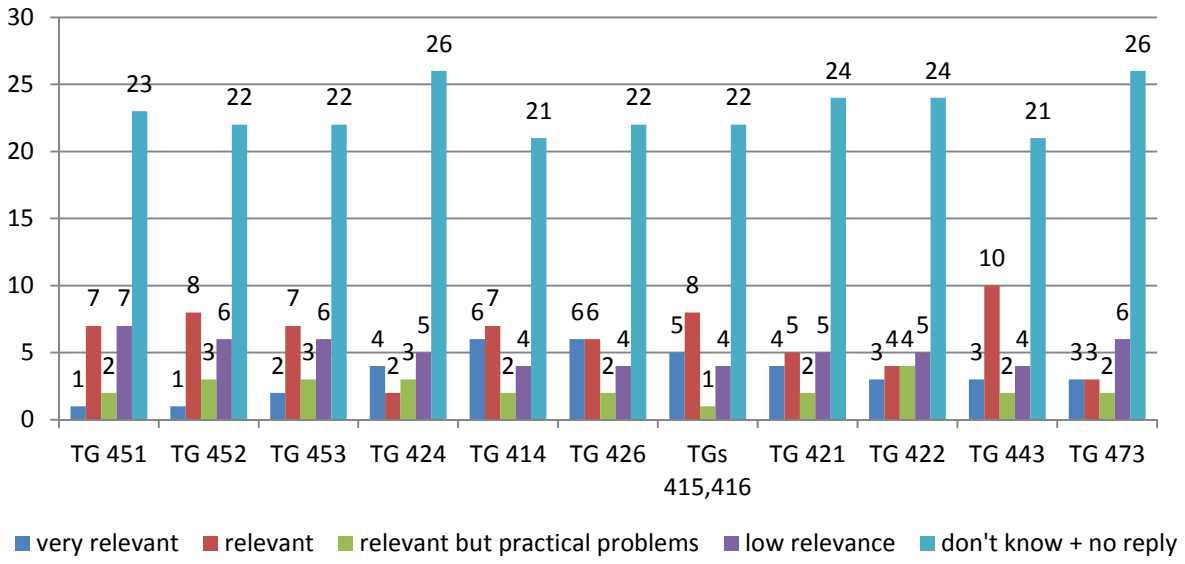

Figure 14 Expert replies to the question "Please rate the relevance of updating tests in the OECD CF (tests further specified in table above) to include epigenetic endpoints/tests." (F.E.1). *Note: The percentage of experts not replying or answering "don't know" was 53-65%.*

Table 36: Expert comments on the inclusion of epigenetic tests (replies in free text fields to question "Add here any other comments related to including epigenetic tests for the identification of endocrine disruptors" (F.E.2)).

1.	Whether epigenetics play a role in environmental chemical induced endocrine disease is still an open question. Therefore development of test methods for epigenetics in premature.
2.	This is still an area of research. Also epigenetic changes are not specific to endocrine disrupters
3.	For sure, epigenetic changes should be taken into account in future testing strategies, especially in context of developmental neurotoxicological effects. However, currently knowledge on normal, physiological epigenetic variation is largely lacking, let alone effects of EDC-induced epigenetic changes as well as transgenerational effects. Therefore, I'm not sure if we are ready yet to implement epigenetic endpoints in ED identification.
4.	miRNA and mRNA whole blood gene expression for candidate genes
5.	Epigenetic MoA is a field repeatedly indicated by OECD working groups, it appears to be a high priority issue from the view of scientific and policy awareness. Nevertheless regulatory applicable IATAs inclusive for this may not be achievable in short term but rather intermediate or long term, be it in vitro or combined in vivo approaches. A structured, explorative approach is necessary focussing on the long term development and assessment of in vitro focussed approaches. The METiCx consortium is prepared to support these developments.
6.	Please take note of the recommenations in Table 7 of Greally & Jacobs, 2013 Ref. Greally JM, Jacobs MN.In Vitro and In Vivo Testing Methods of Epigenomic Endpoints for Evaluating Endocrine Disruptors, ALTEX. 2013;30(4):445-71.
7.	Not yet sufficiently familar with these tests to comment.
8.	We would support the recommendations as outlined in Greally & Jacobs, Altex 2014: Greally & Jacobs. In vitro and in vivo Testing Methods of epigenetic Endpoints for evaluating endocrine Disruptors, ALTEX (2014), 7/13.

9. Where are the fish tests that are so much easier to develop focusing on epigenetic effects? The FET (TG246) is an amazing platform for this and is not even considered an animal test!
10. A major practical problem in the rodent tests is to identify the target organ/tissue for epigenetic modifications, which may be distinct from the organ/tissue showing the apical effect (e.g. epigenetic effects in the hypothalamus may underlie obesogenic effects), and such models are therefore not suitable for screening of epigenetic effects. In general, there is insufficient knowledge on the mechanistic role of epigenetic modifications in the pathway between initiation and apical effect, but ED dependent epigenetic effects can be identified in dedicated models (cell culture models, zebrafish embryos, viable yellow agouti (Avy) mouse) and associated with an altered phenotype (e.g. respectively differentiation, development, coat color). The zebrafish embryo epigenetic toxicity assay given in the first row of the above copied Table 7 of Annex 1 of the detailed review paper OECD No 178 is missing in the questionaire F.E.1 table .
11. The field of epigenetics is still advancing and needs to advance even further before informed introduction of epigenetic endpoints should be considered for addition into existing guideline studies.
12. Specific methylation targets rather than global methylation assays would be preferred here, once sensitive gene regions have been identified.
13. Extended one generation or two generation tests (TG 415/416, 443) might give the possibility to monitor both somatic epigenetic modifications, and more evident, inherited epimutations linked to endocrine disruption. Assays including only one generation might be useful to gain indications for epigenetic modifications linked to endocrine disruption or provide supporting information. However, these tests might not give sufficient evidence for an epigenetic endocrine disruption mechanism. Epigenetic modifications of the nervous system might be detected the best during development of individuals (TG 414, TG 426), rather than in adult individuals (TG 424). Development might represent the more sensitive window for endocrine disruption whereas in the adult endocrine disruption might be compensated. Non-mammalian tests should be taken into account as well (e.g. zebrafish embryos). Consider this a general view, as we are not experts in epigenetics.
14. Fundamental research is required prior to (1) evaluating the uncertainty in the ERA that is caused by epigenetic effects and (2) including epigenetic effects in environmental risk assessment procedures (OECD 2011f, Vandegehuchte & Janssen 2011, Head et al. 2012).
15. but this needs a great deal of exploratory work first, and a workshop is being held in November 2015 to start to explore this..
16. We do see value in tests probing epigenetic processes, to the extent that such tests could provide mechanistic data on the biological processes underlying endocrine disruption. Addition of epigenetic endpoints to existing studies will be useful if this information is used to inform AOPs. We support the recommendation of DRP 178 that an important goal will be the development of in vitro and short term assays for epigenetic changes predictive of adverse outcomes. Given the amount of further fundamental research needed however these are not so high a priority for resource allocation as other aspects, in particular metabolism. In addition to Table 7 above, DRP 178 indicates that adding epigenetic endpoints to relevant in vitro assays will also in principle be feasible - this should also be considered in any further discussion.
17. Since the molecular mechanism of epigenetic transmission are as yet not well described, the focus should be on first elucidated mechanism(s) - in vitro or in vivo - and designing diagnostic endpoints before considering modification of any of these animal-ingestive test methods.
18. I do not have any experience with the potentiall use of epigenetic tests
19. An updated 2 generation test would be useful to identify epigenetic effects

3.5 Views on current OECD Conceptual Framework

In this section, experts were asked to express their opinion on current tests/screens included in the OECD Conceptual Framework (CF), existing tests that are not included in the OECD CF and the need for the development of new ones. This section is divided into 5 subsections, specific for:

1. *In vitro* tests
2. *In vivo* mammalian studies
3. Non-mammalian vertebrate *in vivo* studies
4. Invertebrate studies

Table 37 gives an overview on the overall ranking regarding relevance of mammalian and non-mammalian vertebrate test methods of the OECD conceptual framework regarding their diagnostic value to detect endocrine related perturbations. More detail on all individual tests is provided in sections 3.5.1 - 3.5.4 showing the results for survey sections F.A.- F.D.

JRC used the following equation to calculate the overall ranking of the TMs indicated by the experts:

(Equation 7) $$ratio = \frac{\#very\ relevant + \#relevant}{\#low + \#no\ relevance}$$

Table 37: Overview on the overall ranking of test methods (TM) included in the OECD Conceptual Framework (E.A – E.G) based on Equation 7 above. (* In some cases for the non-mammalian tests, there were no answers stating "no or low relevance", so they should be considered of high relevance even if they have no numerical score.)

Test method	**Ranking (Eq7)**
Amphibian metamorphosis assay (AMA, TG 231)	*
Fish sexual development test (FSDT, TG 234)	*
Medaka extended one-generation reproduction test (MEOGRT)	*
Extended one-generation reproductive toxicity study (TG 443)	29.0
Larval amphibian growth and development assay (LAGDA)	20.0
Xenopus embryonic thyroid signalling assay (XETA)	20.0
Fish short term reproduction assay (FSTRA, TG 229)	17.0
Developmental neurotoxicity (TG 426)	8.7
Pubertal development and thyroid function assay in intact juvenile/peripubertal male rats (OCSPP 890.1500)	7.7
Detection of endocrine active substances, acting through estrogen receptors, using transgenic cyp19a1b-GFP zebrafish embryos (EASZY)	7.5

Pubertal development and thyroid function assay in intact juvenile/peripubertal female rats (OCSPP 890.1450)	7.3
Androgenised female stickleback screen (AFSS, GD 140)	5.3
21-day fish assay (TG 230)	4.7
Two-generation reproduction toxicity study (TG 416)	4.6
Prenatal developmental toxicity study (TG 414)	4.4
Fish reproduction partial lifecycle test (no validation ongoing)	3.0
Uterotrophic bioassay in rodents (TG 440)	2.8
Avian reproduction test (TG 206)	2.8
Avian two-generation test (ATGT)	2.8
Combined 28 day reproductive screening test (TG 421+422)	2.7
Hershberger bioassay in rats (TG 441)	2.6
Combined chronic toxicity / carcinogenicity study (TG 453)	2.0
One-generation reproduction toxicity study (TG 415)	1.8
Fish life cycle toxicity test (FLCTT) (no validation ongoing)	1.6
Repeated dose 28 day study (TG 407)	1.6
Repeated dose 90 day study (TG 408)	1.4

* In some cases for the non-mammalian tests, there were no answers stating "no or low relevance", so they should be considered of high relevance even if they have no numerical score.

3.5.1 *In vitro* methods (F.A)

In section F.A of the survey, experts were asked to:

- comment on the technical performance, weaknesses, limitations, regulatory acceptance etc. of *in vitro* test methods in the OECD CF (F.A.1 and F.A.2, **Table 38**)
- name any other relevant *in vitro* TM not included in the current OECD CF (F.A.3, **Table 39**)
- propose the development of new *in vitro* tests (F.A.4, **Table 40**)

Table 38: Overview of comments on the technical performance, weaknesses, limitations, regulatory acceptance of in vitro TMs (free text comments to the question "Please add here your comment on any current OECD CF *in vitro* test (F.A.2)).

General comments
1. need to combine with metabolism.
2. Test materials of in vitro assays (e.g. cell lines) should be open to the public and affordable.

3. The in vitro tests are only useful for identification but at the moment not useful for quantitative risk assessment mainly because only the nominal concentration is given whereas the actual concentration should be assessed esp. in the tests where cells are exposed to the test substance (see also Blaauboer BJ, Boekelheide K, Clewell HJ, Daneshian M, Dingemans MML, Goldberg AM, et al. The use of biomarkers of toxicity for integrating in vitro hazard estimates into risk assessment for humans. ALTEX. 2012;29(4):411-25.
4. Steroid binding and transactivation assays all make the mistake of assuming that EDs with disruptive effects are all likely to act directly as steroid receptor ligands. This is most likely not true, as in the case of phthalates which are anti-androgenic not by any interaction with the androgen receptor, but by disrupting Leydig cell development within a critical androgen-dependent time-window, Steroidogenesis assays using secondary cell lines represent steady state adult-like situations which are relatively insensitive compared to those cell types during their differentiation and development in vivo. This is especially true for secondary Leydig cell lines, which in many cases do not even indicate the full complement of steroidogenic enzymes, and do not reflect events during Leydig cell differentiation.
5. OECD has recently approved a human recombinant ER binding assay that should replace the rat cytosol test or any other non-human mammalian ER test. Note also that TG 455 has been updated to include anti-estrogenic activity.
ER binding assay *Note by JRC: since July 2015 OECD TG 493*
6. low sensitivity
7. tests on binding/transactivation should be integrated with functional assays using human-relevant biomarkers (such as the PSA assay). These assays would be the next step, downstream to receptor interactions, in building Adverese Outcome Pathways for endocrine disrupters
8. Actual tests used with estrogen receptors are based in the classical model of ERs acting as transcription factors binding to ERE in the DNA. Now we know that ERs act outside the nucleus activating other signaling pathways. It is already accepted by endocrinologist that extranuclear initiated signaling is important in multiple endocrine systems. Moreover, it has been demonstrated using genetically modified mice that both ERs when outside the nucleus mediate low dose actions of environmental estrogens such as bisphenol-A. Therefore, it is necessary to incorporate new tests related to ER activation outside the nucleus.
AR binding assay
9. low sensitivity
10. AR binding assay - rat cytosol from castrated rats may not be specific for other species
11. need to combine with metabolism.
12. despite promising results, these were dropped by the lead country (US) in pre validation
Estrogen receptor transactivation (TG 455)
13. very useful, high sensitivity
14. need to combine with metabolism.
15. Actual tests used with estrogen receptors are based in the classical model of ERs acting as transcription factors binding to ERE in the DNA. Now we know that ERs act outside the nucleus activating other signaling pathways. It is already accepted by endocrinologist that extranuclear initiated signaling is important in multiple endocrine systems. Moreover, it has been demonstrated using genetically modified mice that both ERs when outside the nucleus mediate low dose actions of environmental estrogens such as bisphenol-A. Therefore, it is necessary to incorporate new tests related to ER activation outside the nucleus.
Androgen receptor transactivation *Note by JRC: since July 2016 OECD TG 458*
16. very useful, high sensitivity

Steroidogenesis assay (TG 456)
17. very useful, high sensitivity
18. Current T456 guideline only describes the assessment of T and E2 production in H295R. In recent publications, often metabolic profiling of hormone production is performed. This is a valuable addition to the current test guideline and could be considered for implementation.
19. Valuable test, presently E2 and testosterone is included in the OECD TG. Could be improved by including several other steroid hormones as well.
20. Steroidogenesis shows quite high variability but is still very usable. Validation of the method for hormones other than Estradiol and Testosterone (e.g. progesterone) could be great.
Aromatase assay
21. very useful, high sensitivity
22. The steroidogenesis assay is also able to measure aromatase activity.
23. has not yet been specified in the OECD CF and should be included if the OECD test guideline is available or under development
24. Aromatase assay is limited as it only detects inhibition of the enzyme; however increased aromatase activity is equally important information to have as increased activity is associated with effects on puberty and reproductive functioning. To some extent this test is redundant as the steroidogenesis assay provides more relevant functional information.
MCF-7 cell proliferation assay
25. very useful as the most integrative estrogen sensitive in vitro assay, but results need to be interpreted in connection with ER transactivation
26. This is not a specific assay for ER activation. Still valuable for screening for effects of breast cancer cell proliferation
27. The project for MCF-7 cell proliferation assay was stopped. This assay should be deleted.
28. MCF-7 cell proliferation assay is not specific for estrogenic/anti-estrogenic chemicals - the relevance of the data for identifying EDs is questionable.

Table 39: Overview of suggestions for the inclusion of other existing *in vitro* TMs into the OECD CF (free text comments to the question "Apart from what was addressed above, are you aware of any existing additional *in vitro* test that could add value and should be further considered? (F.A.3)).

1. there are a LOT of in vitro tests not mentioned here that are upstream or downstream of the targeted biology.
2. CYP17 activity assay, nuclear receptor coregulator assays
3. AhR assay PPARgamma PPARalfa 3T3 adipocyte differentiation assay
4. We agree with the OECD press release made after the last EDTA meeting (Oct. 2014) that there is a particular need for development of in vitro tests investigating key events of the AOP for thyroid hormone disruption/perturbations of the thyroid system. For example one promising key event which could be covered by development of an OECD TG seems to be an in vitro assay for thyroid peroxidase activity (TPO).
5. We are developing a Leydig cell differentiation test, which is using relatively undifferentiated Leydig cells and differentiating these in vitro, under the influence of various factors, including EDs.
6. This is well developed in the METICX consortium proposal.
7. In line with comments under Section E on the importance of mechanistic approaches, it is

generally necessary for the Conceptual Framework to be updated to refer to a greater variety of in vitro tests, including methods relevant beyond the EA modalities currently covered. Although additional in vitro methods stand in varying states of validation and standardisation, many are already capable of adding valuable insights, used appropriately as part of weight of evidence analyses. As the Conceptual Framework is not restricted to Test Guideline methods or intended to act as a prescriptive testing strategy, it will be more useful if it takes a broader and more inclusive approach. The thyroid scoping document identifies a number of in vitro assays as in an advanced state of development, these are obvious candidates to be included. Assays addressing endocrine and androgen mechanisms of action in the ToxCast (http://www.epa.gov/ncct/toxcast/) and Tox21 (http://www.epa.gov/ncct/Tox21/) programs should also be included. Focus on the addition of in vitro assays targeting molecular events playing a role in more than one pathway, (e.g. RXR and AhR reporter assays) could also be considered. These would provide information relevant to a variety of potential adverse outcomes, which would be helpful in individual assessments, but their use and further development would also help progress the characterisation of endocrine AOPs in a systematic manner. In vitro assessment of metabolism is an area needing urgent further work, but it should already be mentioned in the Conceptual Framework, as previous analysis (OECD DRP 97, Jacobs et al. ALTEX. 30(3):331-51) has concluded that some of the necessary tools for incorporating the effects of metabolism and bioavailability are already available. The DRP specifically comments that although ideal solutions for simultaneously testing endocrine potential and metabolism are a way off, this should not prevent the use of the best combination of tests already available.
8. Assays addressing endocrine and androgen mechanisms of action in the ToxCast (http://www.epa.gov/ncct/toxcast/) and Tox21 (http://www.epa.gov/ncct/Tox21/) programs should be included.

Table 40: Overview of proposals for the development of new *in vitro* TMs (free text comments to the question "Apart from what was addressed above, do you see a need for developing new *in vitro* tests to better assess endocrine disrupting substances? Will there be any added value from such new tests from a regulatory perspective?" (F.A.4)).

1. Yes and Yes
2. Not necessarily new tests, but rather smart use/combination of testing would enhance applicability for regulatory purposes. More effort should be placed on smart combination of test results together with implementation of PBPK/QIVIVE strategies.
3. I wish to stress the major importance of developing in vitro/in vivo assays relevant to metabolic syndrome and diabetes, See above section E for relevant examples (e.g., PPAR signalling)
4. development of human in vitro tests from primary human testicular mixed cell culture
5. The focus of further work should be the development of in vitro/in silico focussed IATAs with improved metabolism relevant for human and/or environment. The current CF informs on potentially relevant assays and also the OECD GD 150 provides a basis, but does not intend to represent an IATA. See earlier comments including to DA7.
6. There is a great need for developing assays based on human induced pluripotent stem cells based on e.g. High Content Imaging. Such an assay can predict effect on developmental toxicity
7. There is a need for development of new in vitro tests to get information on deseases/disorders with potential endocrine MoA like diabetes or obesitas. The biological relevance shoul be high.
8. C.f. the response above (F.A.3) In general, in order to address which in vitro tests to focus development of, it might be feasible to consider development of AOPs as the first step, and target MIEs and KEs with development of promising in vitro tests to assure mechanistic relevance and predictivity. Furthermore, we find it of very high priority to work on inclusion of metabolic capacity in all in vitro systems.
9. it urges tests related to extranuclearly initiated actions of steroid receptors.
10. The problem with in vitro tests is that they are giving many false positives; there's no organic chemical in the literature that doesn't have at least one reference as an EDC from in vitro tests; this is very disappointing in terms of the 3Rs but endocrine disruption is unfortunately a field that requires relevant in vivo data for proper regulation; if not we are risking over-regulating or

missing important activities.
11. Yes. The aforementioned assays were only for a limited number of pathways. Many pathways like PPAR, RXR, etc have not yet been included. As the regulatory identification of endocrine disruptors is not limited to EATS pathways, the addition of assays for other pathways is highly relevant.
12. Whatever test is used or developed a causal link between the parameter evaluated and the long term adverse effect must be established. Understanding the AOP/MOA is critical before pertinent tests can be validated.
13. Leydig cells are the key intermediate in most ED effects leading to androgen-dependent aspects of TDS and cryptorchidism, which together represent one of the most frequent and obvious consequences of ED exposure in the human population. As mentioned above, Leydig cell development is not addressed by any of the existing tests (including secondary, immortalized Leydig cell lines), even though these cells are known in vivo to be major targets of phthalate exposure.
14. - in vitro transactivation assays for orphan receptors, i.e. estrogen related receptors (EER) may be of relevance (e.g. rather strong binding of BPA and its metabolites).
15. IATA development.
16. Yes, in line with previous responses, development of in vitro tests in general should be a priority for better assessment of endocrine disrupting substances. Mechanistic data from such tests informs on the existence of an endocrine mode of action, which is essential from a regulatory perspective within the current EU legal framework and under each of the options being suggested for horizontal criteria on identification of endocrine disruptors. An improved battery of in vitro tests will help address mixtures and cumulative effects which is also very important from a regulatory perspective -in vitro methods capable of rapid/low cost turnover will be essential for this. Concerning human health effects, in vitro methods offer improved scope for ensuring that data is human relevant, counteracting the limitations inherent in animal studies arising through species differences. Considering environmental effects, understanding of underlying biology obtainable from in vitro tests will enable accurate assessments of the relevance of a particular result to multiple wildlife species/taxa.
17. Assays that address thyroid mechanisms of action (as described in the OECD Thyroid Scoping Document 207) and metabolism (described in OECD DRP 97, Jacobs et al. ALTEX. 30(3):331-51) of componds should be a high prioirity for development (as suggested in DRP 178). There is definite added value of additional in vitro tests as these tests provide essential mechanistic information regarding potential endocrine mode of action - a requirement under current EU legislation and likely to be critical for defining ED chemicals under future EU policy. The benefits of prioritizing in vitro mechanistic assays include: increased assessment through-put for screening existing and new chemicals as well as mixtures; the ability to include tests based on human biology (and to account for species differences); the eventual possibility to move away from apical animal tests to more predictive Integrated Approaches to Testing and Assessment based on upstream events along a series of AOPs (see OECD Guidance, Templates and knowledgebases for creating, assessing and using AOPs: http://www.oecd.org/chemicalsafety/testing/adverse-outcome-pathways-molecular-screening-and-toxicogenomics.htm). As mentioned above, a critical aspect of relying on in vitro methods is to account for metabolism. Options for including metabolism in in vitro assessments have been described in OECD DRP 97 and by Jacobs et al. ALTEX. 30(3):331-51.
18. Yes - to prioritise substances for more in depth testing. New and improved in vitro tests are also needed for getting better regulatory acceptance of in vitro studies for ED identification in the long run (important for reducing animal testing and in cases where in vivo tests are not feasible/not allowed as in cosmetics regulation). Also, we need to focus on in vitro test for mechanisms currently not covered by existing in vitro tests.

JRC Summary of the comments in free text fields of Questions F.A regarding *in vitro* TMs

The most important aspect mentioned for *in vitro* tests for the identification of EDs, was their use in smart combination in a Weight of Evidence approach. *In vitro* tests should ideally be combined in a mechanistically based framework addressing specific key events in relevant AOPs. The current OECD CF tests focus on estrogenic and androgenic effects, this should be widened to other relevant pathways. Another important aspect was to consider metabolism when using in vitro tests.

3.5.2 Mammalian *in vivo* test methods (F.B)

In section F.B of the survey, experts were asked to:

- rate the mammalian in vivo studies of the OECD CF regarding their relevance considering their diagnostic value to detect endocrine related perturbations (F.B.1, **Figure 15**)
- comment on the diagnostic value to detect endocrine related perturbations (F.B.2 and F.B.3, **Table 41**)
- comment on the TMs concerning their ability to distinguish endocrine effects from general toxicity and indirect effects on the endocrine system (F.B.4 and F.B.5, **Table 41**)
- propose relevant enhancements of mammalian in vivo TMs (F.B.6 and F.B.7, **Table 41**)
- mention additional existing in vivo TMs for consideration (F.B.8)
- propose the development of new mammalian in vivo studies to better identify endocrine disrupting substances considering the added value from regulatory perspective (F.B.9, **Table 42**).

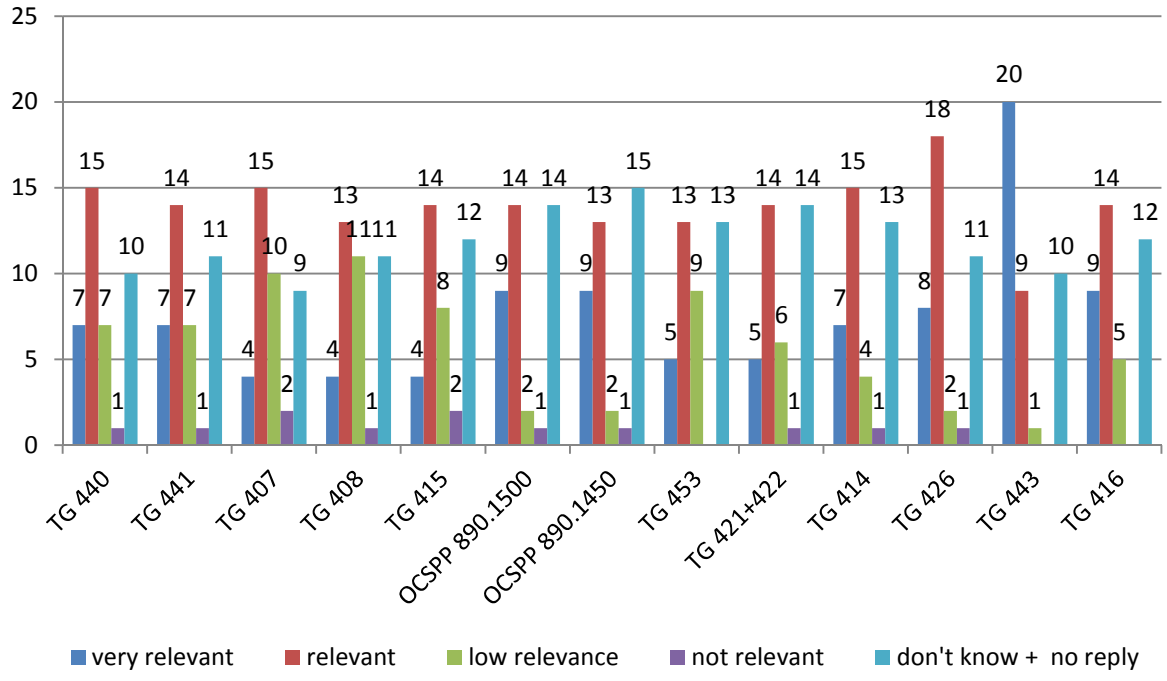

Figure 15: Replies of experts to question F.B.1: Please rate the mammalian *in vivo* studies below regarding their relevance considering their diagnostic value to detect endocrine related perturbations in the test alone or in combination with other tests. (*Note:* The percentage of experts not replying or answering "don't know" was 25-38%.)

Table 41: Expert comments on mammalian *in vivo* studies regarding (1) their diagnostic value to detect endocrine related perturbations in the TM alone or in combination with other tests (F.B.2/3), (2) regarding the TM ability to distinguish endocrine related effects from general toxicity or indirect effects on the endocrine system (F.B.4/5), (3) proposing enhancements to the TMs (F.B.6/7).

General comments on mammalian in vivo test methods
• diagnostic value to detect endocrine related perturbations in this test alone or in combination with other tests
1. Note we had difficulty with the question because it was not clear whether diagnoses of MoA was meant or diagnoses of adverse effects or both. We have answered the questions as diagnoses MoA. If adverse effects should be included as well the high relevance of the mechanistic studies should be changed into relevant instead of high relevance 2. The level 4 and 5 studies in the current OECD Conceptual Framework are of highest diagnostic value, as they are able to define adversity, potency and the leading health effect of a substance on the whole organism. The data generated from TGs 440, 441 and the male and female pubertal development assays are heavily influenced by the general state of the animal. It is, therefore, sometimes difficult to interpret the data and to differentiate the effects due to general toxicity from those due to ED 3. General comment: usually a combination of tests needed for diagnostic purposes - TG 443 being the most diagnostic. But of course the TG 440 and Tg 441 in intact animals will also be diagnostic 4. We note that you use the wording perturbations, not MoA and adverse effects, respectively. This might confuse the reader and the responses. In the tables above (F.B.1 and F.B.2), the question cannot be answered in general but depend on case by case evaluations of substances and effects. 5. Most acute or semi-acute studies carried out in adult rodents are of little value, since they are relatively insensitive, and only address more reprotox issues rather than ED-related endpoints. 6. No single test alone can be used to get a diagnostic for ED.
• allow to distinguish endocrine disrupting effects from general toxicity or indirect effects on the endocrine system
7. All tests may provide, albeit with different accuracy, NOAEL for endocrine effects and NOAEL for other toxic effects. Endocrine disruption may be identified by individual effects (thyroid changes) or y recognized effect patterns (e.g., those resulting from aromatase inhibition) 8. Depends on dosing and effects 9. The question is unclear as it is in many cases difficult to distinguish between effect induced due to endocrine disruption or other modes of action due to the lack of assessment of endpoints for mode of action. 10. In many cases it is not possible to answer clear yes or no to this question (F.B.4). In some cases the assays will allow to distinguish, in other cases they will not. It depends on the substance and which other MoAs/AOPs it acts through. 11. The answer depends upon dosages tested. 12. Assays like Uterotrophic, Hershberger tests etc are not designed to detect general toxicity. It is possible to use assays like repeated dose, reproductive toxicity tests, etc to assess chemical specific effects. This is often done by using a weight of evidence approach to take into account toxicity results together with mechanistic information. 13. To distinguish whether an effect is direct or secondary a set of in vivo studies is necessary covering a broad range of durations (from single to long-term) and consequently a set of studies. It appears this is only possible by expert judgment in a weight-of-evidence approach on a case-by-case basis. As a general rule effects seen at doses which produce excessive toxicity (e.g. markedly reduced body weight gain or severe histopathological lesions) are not useful to assess a primary effect on the endocrine system. 14. difficult to distinguish general tox from ED with only one test...

15. As summarised in OECD GD150 with respect to TGs 415 and 416, plus 421 and 422: as all the endpoints are apical, it is difficult to discern mechanism of action from [these tests] alone. Information on mechanism of action needs to be obtained from in vitro EATS assays or in vivo lower tier tests (p. 343, p. 496). Though data from apical studies can and should be used where possible to inform AOP development until pathways are better characterised, this is a key reason why they are not a priority for future work on identifying endocrine disruptors.

- proposals for enhancing them

16. General comments: All methods should be upgraded with new ED-endpoints once they are validated for one in vivo test and suitable to include in other TG
17. Enhancing may be useful, where sufficient data available for validation with regard to reliability and relevance, but not a high priority. See response to DA7.
18. see Greally and Jacobs ALTEX 2013 and Annex to DRP 178.Enhancing may be useful, where sufficient data available for validation with regard to reliability and relevance. Enhancing may be valid, but experimental work needed first.
19. Generally, we support the addition of further mechanistic endpoints to the above indicated existing animal studies as an interim measure until pathways are better understood; this information should be used to inform AOPs and identify key events that can be targeted by new or existing in vitro methods.

Uterotrophic bioassay in rodents (TG 440)

- diagnostic value to detect endocrine related perturbations in this test alone or in combination with other tests

20. quite insensitive, the ER AR transactivation assays are mostly providing the information needed
21. provide a limited and "narrow" information. Adding more enpoints (e,g,, gene expression) could improve the value of these assays.
22. not sensitive
23. In view of costs and animal welfare this will only be asked in some rare cases. Moreover, it does not provide information on apical endpoints, it only provides information on ER MoA
24. Detection of estrogen antagonists is not currently validated
25. this assay is relevant to screen endocrine active substances but if performed on ovariectomized females, other data are needed to decide if the substance is an endocrine disruptor as the animal is not intact.
26. TG 440 and 441 are mechanistic assays that can be replaced by in vitro assessments (publications on this have been submitted by US EPA and NIH and should be available soon)

- allow to distinguish endocrine disrupting effects from general toxicity or indirect effects on the endocrine system

-

- proposals for enhancing them

27. The current uterotrophic and Hershberger provide a limited and "narrow" information. Adding more enpoints (e,g,, gene expression) could improve the value of these assays.

Hershberger bioassay in rats (TG 441)

- diagnostic value to detect endocrine related perturbations in this test alone or in combination with other tests

28. quite insensitive, the ER AR transactivation assays are mostly providing the information needed
29. provide a limited and "narrow" information. Adding more enpoints (e,g,, gene expression) could improve the value of these assays.
30. not sensitive
31. This test is relevant but has ethical issues with surgical castration; it would be nice to be replaced
32. See previous but then for AR
33. Liver enzyme inducers can confound results due to an increase in testosterone metabolism. Potent estrogens can increase seminal vesicle weights.
34. this assay is relevant to screen endocrine active substances but as it is performed on castrated

males, other data are needed to decide if the substance is an endocrine disruptor as the animal is not intact. 35. TG 440 and 441 are mechanistic assays that can be replaced by in vitro assessments (publications on this have been submitted by US EPA and NIH and should be available soon)
• allow to distinguish endocrine disrupting effects from general toxicity or indirect effects on the endocrine system
-
• proposals for enhancing them
36. The current uterotrophic and Hershberger provide a limited and "narrow" information. Adding more enpoints (e,g,, gene expression) could improve the value of these assays. 37. A sensitive, quali and quantitative tool to evaluate endocrine disruptors are the morphometric analysis of the uterus, the epididymis and testes (could be applied to any in vivo exposure assays). In the uterus, is possible to evaluate the endometrium thickness, diameter and number of glands, which are also indicative endpoints related to endocrine disruptors activity (i.e. estrogenicity). In males, sperm parameters, testicular and epididymal histo-morphometry are very important, informative and could also be applied to any in vivo exposure assays. Plasma hormone levels (estradiol, FSH, LH and testosterone) are very relevant and indicative information on steroidogenesis and HPG axis in vivo evidence, applied to any assay as well. A suggestion for evaluation of these hormones also in adult female animals is important, but the comparison of groups can only be between the rats that are in the same cycling period (e.g estrous). Testosterone can also be measured in female animals even at very low levels when using the radioimmunoassay as the most sensitive and quantifying method. More than one time point of collection on the same treatment dose would also reduce the bias of hormonal variation, increase the statistical significance and confirm a pattern. Alternatively to the standard Hershberger test that adopts surgical male castration, Ashby et al. (2004) and Tinwell et al. (2007) proposed an assay lasting 10 days in intact immature weanling rats (submitted to chemical castration with testosterone propionate or flutamide) thus avoiding the surgical procedure. With the entire male reproductive system intact, this assay can be used for the detection of compounds with anti/androgenic activities. Specifically on the pubertal development and thyroid function assay in intact juvenile/peripubertal female rats (OCSPP 890.1450), the period of estrous cycling evaluation should be higher, especially because, when the females are immature, there are differences between species first day of vaginal opening and ovary maturation. Then, the current 20 days of observation is not enough to proper characterize the days of cycling period into the categories proposed. A minimum of more 20 days would be necessary to better evaluate at least 3-4 complete cycles and establish a correct pattern for this endpoint. Guerra MT, de Toledo FC, Kempinas Wde G. In utero and lactational exposure to fenvalerate disrupts reproductive function in female rats. Reprod Toxicol 2011; 32:298-303; Weibel ER: Principles and methods for the morphometric study of the lung and other organs. Lab Invest 1963, 12(1):131-155; Leblond CP, Clermont Y: Spermiogenesis of rat, mouse, hamster and guinea pig as revealed by the periodic acid-fuchsin sulfurous acid technique. Am J Anat 1952, 90(2):167-215; Robb GW, Amann RP, Killian GJ: Daily sperm production and epididymal sperm reserves of pubertal and adult rats. J Reprod Fertil 1978, 54(1):103-107; Ashby J, Lefevre PA, Tinwell H, Odum J, Owens W. Testosterone-stimulated weanlings as an alternative to castrated male rats in the Hershberger anti-androgen assay. Regul Toxicol Pharmacol 2004; 39:229-38; Tinwell H, Friry-Santini C, Rouquie D, Belluco S, Elies L, Pallen C, et al. Evaluation of the antiandrogenic effects of flutamide, DDE, and linuron in the weanling rat assay using organ weight, histopathological, and proteomic approaches. Toxicol Sci 2007; 100:54-65.
Repeated dose 28 day study (TG 407)
• diagnostic value to detect endocrine related perturbations in this test alone or in combination with other tests
38. at best only giving very indirect information about endocrine disruption 39. exposure time frame misses important developmental windows, too low number of animals to evaluate hormone levels (power analysis should be applied to address this question), protocol does not require that females are sacrificed in same stage of estrous leading to high variability of endpoints 40. have severe limitations due to lack of (mandatory) endocrine specific endpoints and to lack of

exposure during sensitive windows of development.
41. not sensitive
42. endocrine-relevant endpoints added to TG 407 have been shown to be relatively insensitive and therefore of limited utility

- allow to distinguish endocrine disrupting effects from general toxicity or indirect effects on the endocrine system

43. TG 407 and 408 can normally not distinguish indirect effects from endocrine disrupting effects, do not cover windows of susceptibility
44. TG 407, 408, 416 lack of endocrine specific endpoints
45. Does not differentiate

- proposals for enhancing them

46. There are inconsistencies between TG 407 (Repeated Dose 28-Day Oral Toxicity Study in Rodents) and TG 408 (Repeated Dose 90-Day Oral Toxicity in Rodents) since the updating of TG 407, and these include omissions in blood and tissue examinations in TG 408. However for some classes of substances, such as food flavourings, in some jurisdictions (e.g. Europe) TG 408 is required, even though in some cases, with the new added endpoints, the updated TG 407 has more relevant tissue and blood measurements. Regulatory authorities would find it very helpful to have similar endpoints included in the 90 day studies also. Can the ED relevant endpoints and tissues be amended in TG 408 as for TG 407, without going through a full validation process? Would a retrospective data analysis help such a process? There are other examples of lower level tests where there need to be more consistent revisions in line with updated TGs, and similarly there may also such examples in the ecotoxicity TG series.Both the 28 and the 90-day studies (OECD TG 407 and OECD TG 408, respectively) are included in level 4 of the OECD Conceptual Framework, however, only OECD TG 407 has been updated and validated for the detection of endocrine disrupters. To increase the ability for detection of ED effects in OECD TG 408 it is recommended to include as mandatory similar endpoints as in OECD TG 407, incl. the current optional ED related endpoints. Ref: http://www.cend.dk/rapporter/EDtestingstrategy.pdf A feasibility study for minor enhancements of TG 414 (Prenatal Developmental Toxicity Study) with ED-relevant endpoints have been proposed by DK and is included on the OECD Workplan. Assessment of testosterone levels and anogenital distance (AGD) in male foetuses as well as an enhancement with some additional text giving guidance on evaluation of abnormalities of male and female genitalia would be relevant.
47. In our view, the optional ED related endpoints in the repeated dose 28-day study TG407 (including weight of uterus and ovaries, oestrus cyclicity, histopathologic changes in mammary glands and pituitary and circulating levels of T3, t4 and TSH) should become mandatory. Under REACH they could possibly become mandatory if requested as a default or in cases where there are any available information indicative of endocrine activity of the substance, even though these parameters are not mandatory (needed as a default in all cases) according to the OECD TG. It could be considered whether the provisions of REACH make it impossible to request these parameters as a default (i.e. mandatory) in all dossier evaluation cases under REACH

Repeated dose 90 day study (TG 408)

- diagnostic value to detect endocrine related perturbations in this test alone or in combination with other tests

48. at best only giving very indirect information about endocrine disruption
49. have severe limitations due to lack of (mandatory) endocrine specific endpoints and to lack of exposure during sensitive windows of development.
50. not sensitive
51. Relevant for REACH etc although more ED related parameters could/should be included to increase the sensitivity,
52. this assay should be updated by adding the same endocrine sensitive endpoints included in the updated version of the repeated dose 28 day study (TG407)
53. the current TGs 408, 415, 453, combined repro, and 416 do not contain endocrine-specific endpoints and therefore do not inform ED mechanisms of action.

- allow to distinguish endocrine disrupting effects from general toxicity or indirect effects on the endocrine system

54. TG 407 and 408 can normally not distinguish indirect effects from endocrine disrupting effects, do not cover windows of susceptibility
55. TG 407, 408, 416 lack of endocrine specific endpoints
56. Does not differentiate
57. do not contain endocrine-specific endpoints and therefore do not inform ED mechanisms of action
58. without the addition of the endocrine sensitive endpoints, the assay will not give sufficient information to distinguish general effects from endocrine-disrupting effects
59. do not contain endocrine-specific endpoints and therefore do not inform ED mechanisms of action

- proposals for enhancing them

60. There are inconsistencies between TG 407 (Repeated Dose 28-Day Oral Toxicity Study in Rodents) and TG 408 (Repeated Dose 90-Day Oral Toxicity in Rodents) since the updating of TG 407, and these include omissions in blood and tissue examinations in TG 408. However for some classes of substances, such as food flavourings, in some jurisdictions (e.g. Europe) TG 408 is required, even though in some cases, with the new added endpoints, the updated TG 407 has more relevant tissue and blood measurements. Regulatory authorities would find it very helpful to have similar endpoints included in the 90 day studies also. Can the ED relevant endpoints and tissues be amended in TG 408 as for TG 407, without going through a full validation process? Would a retrospective data analysis help such a process? There are other examples of lower level tests where there need to be more consistent revisions in line with updated TGs, and similarly there may also such examples in the ecotoxicity TG series.Both the 28 and the 90-day studies (OECD TG 407 and OECD TG 408, respectively) are included in level 4 of the OECD Conceptual Framework, however, only OECD TG 407 has been updated and validated for the detection of endocrine disrupters. To increase the ability for detection of ED effects in OECD TG 408 it is recommended to include as mandatory similar endpoints as in OECD TG 407, incl. the current optional ED related endpoints. Ref: http://www.cend.dk/rapporter/EDtestingstrategy.pdf A feasibility study for minor enhancements of TG 414 (Prenatal Developmental Toxicity Study) with ED-relevant endpoints have been proposed by DK and is included on the OECD Workplan. Assessment of testosterone levels and anogenital distance (AGD) in male foetuses as well as an enhancement with some additional text giving guidance on evaluation of abnormalities of male and female genitalia would be relevant.
61. TG 408: The repeated dose 90-day study TG 408 should in our view be updated to include as mandatory similar endpoints as in TG 407, including the currently optional ED related endpoints mentioned above.
62. Repeated dose toxicity has been proposed to update for enhancement.
63. Senstive, ED-specific endpoints could be added to the indicated assays if doing so does not affect the stastical power of the other endpoints. This information should be used to inform AOPs and identify key events that could be queried by new or existing in vitro methods.

One-generation reproduction toxicity study (TG 415)

- diagnostic value to detect endocrine related perturbations in this test alone or in combination with other tests

64. sensitivity of endpoints is not adequate
65. Outdated method that does not or hardly measure essential mechanistic parameters, nor does it measure all relevant apical endpoints
66. the current TGs 408, 415, 453, combined repro, and 416 do not contain endocrine-specific endpoints and therefore do not inform ED mechanisms of action

- allow to distinguish endocrine disrupting effects from general toxicity or indirect effects on the endocrine system

67. Does not differentiate
68. do not contain endocrine-specific endpoints and therefore do not inform ED mechanisms of action; As summarised in OECD GD150 with respect to TGs 415 and 416, plus 421 and 422: as all the endpoints are apical, it is difficult to discern mechanism of action from [these tests] alone
69. do not contain endocrine-specific endpoints and therefore do not inform ED mechanisms of action

- proposals for enhancing them

70. TG 415 and TG 416: In our view, they should be deleted from the OECD TGs after the adoption of TG 443.

71. should be deleted, has become redundant and we have better methods available
72. For TG443, TG 415 and TG 416, these might be improved by addition of tail-bleed sampling to assess INSL3 as a new parameter to measure the dynamics of Leydig cell differentiation and impacts on the HPG axis. (Insulin-like factor 3 as a monitor of endocrine disruption. Anand-Ivell R, Ivell R. Reproduction. 2014.147:R87-95)

Pubertal development and thyroid function assay in intact juvenile/peripubertal male rats (OCSPP 890.1500)

- diagnostic value to detect endocrine related perturbations in this test alone or in combination with other tests

73. could have a enhanced relevance if they become juvenile toxicity tests assessing effects in the critical phase of post-natal development (see Maranghi F, Mantovani A. Targeted toxicological testing to investigate the role of endocrine disrupters in puberty disorders. Reproductive Toxicology 2012, 33(3):290-6; Fucic A., Mantovani A. Puberty dysregulation and increased risk of disease in adult life: possible modes of action. Reproductive Toxicology 2014; 44C:15-22)
74. not all relevant exposure time windows are included, hormone measurement have wide physiological variability and n=15 may or may not provide enough power to determine differences
75. not sensitive
76. Relevant, for EATS pathways
77. Many factors can affect age at preputial separation for example body weight/composition. Effects that decrease bodyweight gain can effect this endpoint in a non-endocrine mediated pathway
78. as the animals are exposed during a very critical period of their development sensitive to endocrine disruption, this assay is certainly useful and could be updated by adding some additional endpoints not yet covered (immune system, …)

- allow to distinguish endocrine disrupting effects from general toxicity or indirect effects on the endocrine system

79. May not be able to differentiate depending on other effects in the study (i.e. decreased weight gain)

- proposals for enhancing them

80. The pubertal development assays (males/females) could have a enhanced relevance if they become juvenile toxicity tests assessing effects in the critical phase of post-natal development (see Maranghi F, Mantovani A. Targeted toxicological testing to investigate the role of endocrine disrupters in puberty disorders. Reproductive Toxicology 2012, 33(3):290-6; Fucic A., Mantovani A. Puberty dysregulation and increased risk of disease in adult life: possible modes of action. Reproductive Toxicology 2014; 44C:15-22)

Pubertal development and thyroid function assay in intact juvenile/peripubertal female rats (OCSPP 890.1450)

- diagnostic value to detect endocrine related perturbations in this test alone or in combination with other tests

81. could have a enhanced relevance if they become juvenile toxicity tests assessing effects in the critical phase of post-natal development (see Maranghi F, Mantovani A. Targeted toxicological testing to investigate the role of endocrine disrupters in puberty disorders. Reproductive Toxicology 2012, 33(3):290-6; Fucic A., Mantovani A. Puberty dysregulation and increased risk of disease in adult life: possible modes of action. Reproductive Toxicology 2014; 44C:15-22)
82. not all relevant exposure time windows are included, hormone measurement have wide physiological variability and n=15 may or may not provide enough power to determine differences
83. not sensitive
84. Relevant, for EATS pathways
85. Many factors can affect age at vaginal opening, for example body weight/composition. Effects that decrease bodyweight gain can affect this endpoint in a non-endocrine mediated pathway.
86. as the animals are exposed during a very critical period of their development sensitive to endocrine disruption, this assay is certainly useful and could be updated by adding some additional endpoints not yet covered (immune system, …)

- allow to distinguish endocrine disrupting effects from general toxicity or indirect effects on the endocrine system

87. May not be able to differentiate depending on other effects in the study (i.e. decreased weight gain)

- proposals for enhancing them

88. The pubertal development assays (males/females) could have a enhanced relevance if they become juvenile toxicity tests assessing effects in the critical phase of post-natal development (see Maranghi F, Mantovani A. Targeted toxicological testing to investigate the role of endocrine disrupters in puberty disorders. Reproductive Toxicology 2012, 33(3):290-6; Fucic A., Mantovani A. Puberty dysregulation and increased risk of disease in adult life: possible modes of action. Reproductive Toxicology 2014; 44C:15-22)
89. A sensitive, quali and quantitative tool to evaluate endocrine disruptors are the morphometric analysis of the uterus, the epididymis and testes (could be applied to any in vivo exposure assays). In the uterus, is possible to evaluate the endometrium thickness, diameter and number of glands, which are also indicative endpoints related to endocrine disruptors activity (i.e. estrogenicity). In males, sperm parameters, testicular and epididymal histo-morphometry are very important, informative and could also be applied to any in vivo exposure assays. Plasma hormone levels (estradiol, FSH, LH and testosterone) are very relevant and indicative information on steroidogenesis and HPG axis in vivo evidence, applied to any assay as well. A suggestion for evaluation of these hormones also in adult female animals is important, but the comparison of groups can only be between the rats that are in the same cycling period (e.g estrous). Testosterone can also be measured in female animals even at very low levels when using the radioimmunoassay as the most sensitive and quantifying method. More than one time point of collection on the same treatment dose would also reduce the bias of hormonal variation, increase the statistical significance and confirm a pattern. Alternatively to the standard Hershberger test that adopts surgical male castration, Ashby et al. (2004) and Tinwell et al. (2007) proposed an assay lasting 10 days in intact immature weanling rats (submitted to chemical castration with testosterone propionate or flutamide) thus avoiding the surgical procedure. With the entire male reproductive system intact, this assay can be used for the detection of compounds with anti/androgenic activities. Specifically on the pubertal development and thyroid function assay in intact juvenile/peripubertal female rats (OCSPP 890.1450), the period of estrous cycling evaluation should be higher, especially because, when the females are immature, there are differences between species first day of vaginal opening and ovary maturation. Then, the current 20 days of observation is not enough to proper characterize the days of cycling period into the categories proposed. A minimum of more 20 days would be necessary to better evaluate at least 3-4 complete cycles and establish a correct pattern for this endpoint. Guerra MT, de Toledo FC, Kempinas Wde G. In utero and lactational exposure to fenvalerate disrupts reproductive function in female rats. Reprod Toxicol 2011; 32:298-303; Weibel ER: Principles and methods for the morphometric study of the lung and other organs. Lab Invest 1963, 12(1):131-155; Leblond CP, Clermont Y: Spermiogenesis of rat, mouse, hamster and guinea pig as revealed by the periodic acid-fuchsin sulfurous acid technique. Am J Anat 1952, 90(2):167-215; Robb GW, Amann RP, Killian GJ: Daily sperm production and epididymal sperm reserves of pubertal and adult rats. J Reprod Fertil 1978, 54(1):103-107; Ashby J, Lefevre PA, Tinwell H, Odum J, Owens W. Testosterone-stimulated weanlings as an alternative to castrated male rats in the Hershberger anti-androgen assay. Regul Toxicol Pharmacol 2004; 39:229-38; Tinwell H, Friry-Santini C, Rouquie D, Belluco S, Elies L, Pallen C, et al. Evaluation of the antiandrogenic effects of flutamide, DDE, and linuron in the weanling rat assay using organ weight, histopathological, and proteomic approaches. Toxicol Sci 2007; 100:54-65.

Combined chronic toxicity / carcinogenicity study (TG 453)

- diagnostic value to detect endocrine related perturbations in this test alone or in combination with other tests

90. The carcinogenicity endpoint is not diagnostic for most hormonal cancers via a hormonal mechanism, such as breast, testis, prostate cancer
91. the minimal number of animals (rats) for a 2-year combined chronic toxicity and carcinogenicity bioassay for a three-dose study (plus the control) performed according to the NTP protocol is 440 (50 animals per dose group of both sexes, plus 20 animals per sex used as sentinel animals); for the OECD TG 453 protocol the minimal number is 480; for the OECD TG443 the minimal number is 1200 (Schiffelers et al. 2015) and for the NTP Modified One- Generation Study 3680 animals.

Only few endpoints are examined through these experimental plans and further multiple distinct experiments are requested to allow a broad-based toxicologic evaluation and more global realistic risk assessment. The protocol we proposed reduces/spares animals and resources because the main long-term bioassay is integrated into several subsets of toxicity testing using different schedules of treatment through planning of interim kills and satellite groups in specific WOS. 92. sensitivity of endpoints not sufficient, no prenatal exposure 93. Relevant, although the sensitivity is highly dependent on the strain and species and very limited MoA included. Another disadvantage of these tests is that they only measure the effects from let us say pubetry to early old age. Early life exposure is not covered. 94. the current TGs 408, 415, 453, combined repro, and 416 do not contain endocrine-specific endpoints and therefore do not inform ED mechanisms of action.
• allow to distinguish endocrine disrupting effects from general toxicity or indirect effects on the endocrine system
95. Does not differentiate 96. do not contain endocrine-specific endpoints and therefore do not inform ED mechanisms of action
• proposals for enhancing them
97. the minimal number of animals (rats) for a 2-year combined chronic toxicity and carcinogenicity bioassay for a three-dose study (plus the control) performed according to the NTP protocol is 440 (50 animals per dose group of both sexes, plus 20 animals per sex used as sentinel animals); for the OECD TG 453 protocol the minimal number is 480; for the OECD TG443 the minimal number is 1200 (Schiffelers et al. 2015) and for the NTP Modified One- Generation Study 3680 animals. Only few endpoints are examined through these experimental plans and further multiple distinct experiments are requested to allow a broad-based toxicologic evaluation and more global realistic risk assessment. The protocol we proposed reduces/spares animals and resources because the main long-term bioassay is integrated into several subsets of toxicity testing using different schedules of treatment through planning of interim kills and satellite groups in specific WOS
Combined 28 day reproductive screening test (TG 421+422) *Note by JRC: these TGs have been updated in 2016 to include some endocrine disruptor relevant endpoints*
• diagnostic value to detect endocrine related perturbations in this test alone or in combination with other tests
98. the minimal number of animals (rats) for a 2-year combined chronic toxicity and carcinogenicity bioassay for a three-dose study (plus the control) performed according to the NTP protocol is 440 (50 animals per dose group of both sexes, plus 20 animals per sex used as sentinel animals); for the OECD TG 453 protocol the minimal number is 480; for the OECD TG443 the minimal number is 1200 (Schiffelers et al. 2015) and for the NTP Modified One- Generation Study 3680 animals. Only few endpoints are examined through these experimental plans and further multiple distinct experiments are requested to allow a broad-based toxicologic evaluation and more global realistic risk assessment. The protocol we proposed reduces/spares animals and resources because the main long-term bioassay is integrated into several subsets of toxicity testing using different schedules of treatment through planning of interim kills and satellite groups in specific WOS. 99. exposure period, low number of animals 100. The relevance has increased after the recent update, although It is stressed that in view of the low number of animals and the limited number of parameters the predictive power is still poor (only detects very potent substances).But as long as there is no regulatory obligation to conduct/provide in vitro screening, this method will provide at least some information on potential EDs. It is stressed that a negative outcome does not indicate the substance is not an ED (the same applies for TG 407 and 408)
• allow to distinguish endocrine disrupting effects from general toxicity or indirect effects on the endocrine system
101. Does not differentiate 102. do not contain endocrine-specific endpoints and therefore do not inform ED mechanisms of action; As summarised in OECD GD150 with respect to TGs 415 and 416, plus 421 and 422: as all the endpoints are apical, it is difficult to discern mechanism of action from [these tests] alone
• proposals for enhancing them

103. TG 421/422: has just been enhanced with endpoints of ED relevance (adoption and release of the revised TGs are foreseen in the autumn) 104. Senstive, ED-specific endpoints could be added to the indicated assays if doing so does not affect the stastical power of the other endpoints. This information should be used to inform AOPs and identify key events that could be queried by new or existing in vitro methods.
Prenatal developmental toxicity study (TG 414)
• diagnostic value to detect endocrine related perturbations in this test alone or in combination with other tests
105. the minimal number of animals (rats) for a 2-year combined chronic toxicity and carcinogenicity bioassay for a three-dose study (plus the control) performed according to the NTP protocol is 440 (50 animals per dose group of both sexes, plus 20 animals per sex used as sentinel animals); for the OECD TG 453 protocol the minimal number is 480; for the OECD TG443 the minimal number is 1200 (Schiffelers et al. 2015) and for the NTP Modified One- Generation Study 3680 animals. Only few endpoints are examined through these experimental plans and further multiple distinct experiments are requested to allow a broad-based toxicologic evaluation and more global realistic risk assessment. The protocol we proposed reduces/spares animals and resources because the main long-term bioassay is integrated into several subsets of toxicity testing using different schedules of treatment through planning of interim kills and satellite groups in specific WOS. 106. will become relevant if enhanced with ED endpoints and this is already on the workplan of OECD. 107. The study does not provide definitive mechanistic information, it does provide information on some apical endpoints 108. Endpoints are apical in nature and could be impacted by non-endocrine active effects. How will we determine if that effect is specifically endocrine related? Is it important for human health to determine if the effect is specifically endocrine related? 109. as the development of many systems are not yet achieved at birth in the rat, this assay seems us too short to give sufficient information on endocrine disrupting effects
• allow to distinguish endocrine disrupting effects from general toxicity or indirect effects on the endocrine system
110. DNT test is not sensitive and can only provide ancillary info 111. Does not differentiate
• proposals for enhancing them
112. A feasibility study for minor enhancements of TG 414 (Prenatal Developmental Toxicity Study) with ED-relevant endpoints have been proposed by DK and is included on the OECD Workplan. Assessment of testosterone levels and anogenital distance (AGD) in male foetuses as well as an enhancement with some additional text giving guidance on evaluation of abnormalities of male and female genitalia would be relevant. 113. TG 414: Enhancement is on the OECD TGP workplan with Denmark as lead country. 114. TG 414, is currently being updated 115. Senstive, ED-specific endpoints could be added to the indicated assays if doing so does not affect the stastical power of the other endpoints. This information should be used to inform AOPs and identify key events that could be queried by new or existing in vitro methods
Developmental neurotoxicity (TG 426)
• diagnostic value to detect endocrine related perturbations in this test alone or in combination with other tests
116. the minimal number of animals (rats) for a 2-year combined chronic toxicity and carcinogenicity bioassay for a three-dose study (plus the control) performed according to the NTP protocol is 440 (50 animals per dose group of both sexes, plus 20 animals per sex used as sentinel animals); for the OECD TG 453 protocol the minimal number is 480; for the OECD TG443 the minimal number is 1200 (Schiffelers et al. 2015) and for the NTP Modified One- Generation Study 3680 animals. Only few endpoints are examined through these experimental plans and further multiple distinct experiments are requested to allow a broad-based toxicologic evaluation and more global realistic risk assessment. The protocol we proposed reduces/spares animals and

resources because the main long-term bioassay is integrated into several subsets of toxicity testing using different schedules of treatment through planning of interim kills and satellite groups in specific WOS.

117. sensitivity of endpoints is not adequate

118. See previous, no or limited information on MoA but does provide information on specific apical endpoints

- allow to distinguish endocrine disrupting effects from general toxicity or indirect effects on the endocrine system

119. DNT test is not sensitive and can only provide ancillary info

120. Does not differentiate

- proposals for enhancing them

121. TG 426: In our view, this guideline could be enhanced to include some behavioral endpoints of relevance for ED effects, such as mating and nursing behavior, since studies of EDs have shown gender-related effects on behavior. References: Hass et al., OECD Conceptual Framework for Testing and Assessment of Endocrine Disrupters as a basis for regulation of substances with endocrine disrupting properties, TemaNord 2004:555, http://www.diva-portal.org/smash/get/diva2:702046/FULLTEXT01.pdf Hass et al., Information/testing strategy for identification of substances with endocrine disrupting properties, Danish Centre on Endocrine Disrupters, 2013, http://www.cend.dk/rapporter/EDtestingstrategy.pdf

122. Senstive, ED-specific endpoints could be added to the indicated assays if doing so does not affect the stastical power of the other endpoints. This information should be used to inform AOPs and identify key events that could be queried by new or existing in vitro methods.

Extended one-generation reproductive toxicity study (TG 443)

- diagnostic value to detect endocrine related perturbations in this test alone or in combination with other tests

123. study is informative for in vivo antiandrogens if anogenital distance and nipple retention is inorporated

124. the minimal number of animals (rats) for a 2-year combined chronic toxicity and carcinogenicity bioassay for a three-dose study (plus the control) performed according to the NTP protocol is 440 (50 animals per dose group of both sexes, plus 20 animals per sex used as sentinel animals); for the OECD TG 453 protocol the minimal number is 480; for the OECD TG443 the minimal number is 1200 (Schiffelers et al. 2015) and for the NTP Modified One- Generation Study 3680 animals. Only few endpoints are examined through these experimental plans and further multiple distinct experiments are requested to allow a broad-based toxicologic evaluation and more global realistic risk assessment. The protocol we proposed reduces/spares animals and resources because the main long-term bioassay is integrated into several subsets of toxicity testing using different schedules of treatment through planning of interim kills and satellite groups in specific WOS.

125. period of senescence is not evaluated, recommended 10 animals for hormone measurements is too low, number of animals is too low if females not sacrificed in same phase of estrous cycle, too high variability.

126. Highly relevant provided the cohorts 2 and 3 are included

127. Currently TG 443 is the only protocol that contains sensitive, ED-specific endpoints.

128. should have both DNT and immunotox cohorts mandatory, as these endpoints are very important.

- allow to distinguish endocrine disrupting effects from general toxicity or indirect effects on the endocrine system

129. During maternal exposure to substances that cause haematological effects (e.g. methaemoglobin formation) the parental systemic anaemia could cause reduced number of litters, post implantation loss, reduced pups weight, retarded development, etc. Then, to distinguish secondary endocrine system impairment, other endocrine disruption endpoints only have to be investigated within the above studies.

130. Does not differentiate

131. AGD

132. is the only protocol that contains sensitive, ED-specific endpoints

133. Currently TG 443 is the only protocol that contains sensitive, ED-specific endpoints

- proposals for enhancing them

134. the minimal number of animals (rats) for a 2-year combined chronic toxicity and carcinogenicity bioassay for a three-dose study (plus the control) performed according to the NTP protocol is 440 (50 animals per dose group of both sexes, plus 20 animals per sex used as sentinel animals); for the OECD TG 453 protocol the minimal number is 480; for the OECD TG443 the minimal number is 1200 (Schiffelers et al. 2015) and for the NTP Modified One- Generation Study 3680 animals. Only few endpoints are examined through these experimental plans and further multiple distinct experiments are requested to allow a broad-based toxicologic evaluation and more global realistic risk assessment. The protocol we proposed reduces/spares animals and resources because the main long-term bioassay is integrated into several subsets of toxicity testing using different schedules of treatment through planning of interim kills and satellite groups in specific WOS.
135. Additional endpoints in extended one-generation tests should become mandatory.
136. TG 443 should include mandatory assessment of mammary gland development (Whole mount). The mammary gland is a target of endocrine active substances and evaluation of this in some protocols might provide information on developmental abnormalities or adverse outcomes not available from other endpoints. The mammary gland whole mount has been proposed as a more powerful method to evaluate organizational effects during development.
137. TG 443: In our view, this guideline should be enhanced to include investigation of mammary gland development, since the mammary gland is a target of endocrine active substances and evaluation of this in some protocols might provide information on developmental abnormalities or adverse outcomes not available from other endpoints. The mammary gland whole mount has been proposed as a more powerful method to evaluate organizational effects during development. We are pleased that this particular investigation has been requested in a couple of relevant recent substance evaluation decisions under REACH and we are of the opinion that the same approach should be used in similar future cases.
138. For TG443, TG 415 and TG 416, these might be improved by addition of tail-bleed sampling to assess INSL3 as a new parameter to measure the dynamics of Leydig cell differentiation and impacts on the HPG axis. (Insulin-like factor 3 as a monitor of endocrine disruption. Anand-Ivell R, Ivell R. Reproduction. 2014.147:R87-95)

Two-generation reproduction toxicity study (TG 416)

- diagnostic value to detect endocrine related perturbations in this test alone or in combination with other tests

139. lack of relevant endocrine endpoints
140. the minimal number of animals (rats) for a 2-year combined chronic toxicity and carcinogenicity bioassay for a three-dose study (plus the control) performed according to the NTP protocol is 440 (50 animals per dose group of both sexes, plus 20 animals per sex used as sentinel animals); for the OECD TG 453 protocol the minimal number is 480; for the OECD TG443 the minimal number is 1200 (Schiffelers et al. 2015) and for the NTP Modified One- Generation Study 3680 animals. Only few endpoints are examined through these experimental plans and further multiple distinct experiments are requested to allow a broad-based toxicologic evaluation and more global realistic risk assessment. The protocol we proposed reduces/spares animals and resources because the main long-term bioassay is integrated into several subsets of toxicity testing using different schedules of treatment through planning of interim kills and satellite groups in specific WOS.
141. low number of animals particularly for hormone measurements
142. Less relevant than TG443, because less ED parameters are included, less apical endpoints and it uses more animals
143. Endpoints are apical in nature and could be impacted by non-endocrine active effects. How will we determine if that effect is specifically endocrine related? Is it important for human health to determine if the effect is specifically endocrine related?
144. the extended one-generation reproductive toxicity study is to be preferred for ED relevant endpoints
145. the current TGs 408, 415, 453, combined repro, and 416 do not contain endocrine-specific endpoints and therefore do not inform ED mechanisms of action.
146. does not contain endocrine-specific endpoints and therefore do not inform ED mechanisms of action.

• allow to distinguish endocrine disrupting effects from general toxicity or indirect effects on the endocrine system
147. TG 407, 408, 416 lack of endocrine specific endpoints 148. During maternal exposure to substances that cause haematological effects (e.g. methaemoglobin formation) the parental systemic anaemia could cause reduced number of litters, post implantation loss, reduced pups weight, retarded development, etc. Then, to distinguish secondary endocrine system impairment, other endocrine disruption endpoints only have to be investigated within the above studies. 149. Does not differentiate 150. do not contain endocrine-specific endpoints and therefore do not inform ED mechanisms of action; As summarised in OECD GD150 with respect to TGs 415 and 416, plus 421 and 422: as all the endpoints are apical, it is difficult to discern mechanism of action from [these tests] alone 151. do not contain endocrine-specific endpoints and therefore do not inform ED mechanisms of action
• proposals for enhancing them
152. TG 415 and TG 416: In our view, they should be deleted from the OECD TGs after the adoption of TG 443. 153. TG 415 and TG 416: In our view, they should be deleted from the OECD TGs after the adoption of TG 443. 154. TG 416, as for the one generation toxicity study, deletion should be considered in view of the fact that new methods have become available (TG 443) 155. For TG443, TG 415 and TG 416, these might be improved by addition of tail-bleed sampling to assess INSL3 as a new parameter to measure the dynamics of Leydig cell differentiation and impacts on the HPG axis. (Insulin-like factor 3 as a monitor of endocrine disruption. Anand-Ivell R, Ivell R. Reproduction. 2014.147:R87-95) 156. TG 416 should be updated wtih endpoints that are in the EOGRTS - at a stroke.

Experts were further asked to mention existing additional mammalian *in vivo* studies that could add value and should be further considered (**F.B.8**). The only proposal received was "The NTP guidelines enhances the sensitivity for some EDs-related endpoints".

Table 42: Expert proposals for the development of new mammalian *in vivo* studies to better identify endocrine disrupting substances (free text comments to question "Apart from what was addressed above, do you see a need for developing new mammalian in vivo tests to better identify endocrine disrupting substances? Will there be added value from such new tests from a regulatory perspective?" (F.B.9)).

1.	No -rather than developing new tests, need to think how to enchance existing ones
2.	A comprehensive juvenile toxicity test would be of paramount importance for hazard charactrerization in the post-natal window, from weaning through to pubertuy and sexual maturation
3.	yes, if standardized
4.	Efforts should be made to overcome the fact that we have no current studies that cover the full Life cycle (exposure in utero - late effects)
5.	See response to DA7.
6.	In general, in order to address which in vivo tests to possibly develop, it might be feasible to consider development of AOPs as the first step to get an overview of the biological understanding, and target MIEs, KEs and AOs with development of promising in vivo tests.
7.	Yes, for instance methods aimed at the metabolic disorders, immune function etc.
8.	Most in vivo tests use rodents, and consequently their relevance can be called into question, besides issues of high variance, hence large and ethically questionable animal usage.

9. New apical mammalian studies would continue to present similar interpretive difficulties as existing tests with respect to distinguishing endocrine-mediated from other effects, unraveling significant effects from homeostatic responses, and translating conclusions across species, where underlying mechanisms continue to be poorly understood. In view of this, and considering resource limitations, developing additional mammalian in vivo tests will contribute less to our ability to identify endocrine disruptors than assays supplying mechanistic information and efforts to arrange this into predictive AOPs.

JRC Summary of the comments in free text fields of Questions F.B regarding mammalian *in vivo* TMs

The replies indicate that most of the available mammalian *in vivo* assays do not contain endocrine-specific endpoints and therefore do not inform ED mechanisms of action. The extended one-generation reproductive toxicity study (TG 443) is to be preferred for ED relevant endpoints. Usually a combination of tests is needed for diagnostic purposes - TG 443 being the most diagnostic, although the Uterotrophic (TG 440) and Hershberger (TG 441) assays in intact animals could also be diagnostic. Others considered that the current Uterotrophic and Hershberger assays were quite insensitive and limited in scope. Some experts indicated that adding more endpoints (e.g. gene expression) could improve the value of these assays. Some considered that *in vitro* ER/AR transactivation assays are mostly providing the information needed and that it might be possible to replace them by *in vitro* assessments referring to recent publications on this from US EPA and NIH.

It was also stated that data generated from TGs 440, 441 and the male and female pubertal development assays are heavily influenced by the general state of the animal, consequently it is sometimes difficult to interpret the data and to differentiate the effects due to general toxicity from those due to ED. The pubertal and thyroid function assays in male and female rats were considered to have issue of non-specificity since many factors could affect age of preputial separation and vaginal opening in males and females respectively. Other experts indicated that the tests could have an enhanced relevance if they become juvenile toxicity tests assessing effects in the critical phase of post-natal development.

The combined carcinogenicity bioassay TG 453 was considered not diagnostic for most hormonal cancers via a hormonal mechanism, such as breast, testis, prostate cancer and early life stage exposure was not included

Both the 28 and the 90-day studies (OECD TG 407 and OECD TG 408, respectively) were considered to have severe limitations due to lack of (mandatory) endocrine specific endpoints and to lack of exposure during sensitive windows of development. Although OECD TG 407 has been updated and validated for the detection of endocrine disrupters (including weight of uterus and ovaries, oestrus cyclicity, histopathological changes in mammary glands and pituitary and circulating levels of T3, T4 and TSH) these endpoints are only optional. Some experts consider that these endpoints should be made mandatory others indicate that the additional endpoints have been shown to be relatively insensitive and therefore of limited utility. However, to increase the ability for detection of ED effects in OECD TG 408 it was recommended to include as mandatory the current optional ED related endpoints from TG 407. The proposal on the OECD workplan from Denmark for a feasibility study for minor enhancements of TG 414 (Prenatal Developmental Toxicity Study) with ED-relevant endpoints was also highlighted.

General comments in relation to proposals for enhancement of existing test guidelines: varied from:

— all TGs should be upgraded with new ED-endpoints once they are validated for one in vivo test,

— enhancing may be useful where sufficient data are available for validation with regard to reliability and relevance, but not a high priority,

— enhancements supported as an interim measure until pathways are better understood

— sensitive ED-specific endpoints could be added to indicated assays if doing so did not affect the statistical power of the other endpoints.

Generated mechanistic information from the enhancements should be used to inform AOPs and identify key events that can be targeted by new or existing *in vitro* methods.

Specific proposals for enhancement of TG 443 included investigation of mammary gland development and addition of tail-bleed sampling to assess INSL3 (as a new parameter to measure the dynamics of Leydig cell differentiation and impacts on the HPG axis).

3.5.3 Non-mammalian vertebrate *in vivo* studies (F.C)

In section F.C of the survey, experts were asked to:

- rate the non-mammalian *in vivo* studies of the OECD CF regarding their relevance considering their diagnostic value to detect endocrine related perturbations (F.C.1, **Figure 16**)
- comment on the diagnostic value to detect endocrine related perturbations (F.C.2 and F.C.3, **Table 43**)
- comment on the TMs concerning their ability to distinguish endocrine effects from general toxicity and indirect effects on the endocrine system (F.C.4 and F.C.5,Table 43)
- propose relevant enhancements of non-mammalian vertebrate *in vivo* TMs (F.C.6 and F.C.7, **Table 43**)
- add any additional comment on the non-mammalian vertebrate *in vivo* TMs within the OECD CF (F.C.8, Table 44)
- mention additional existing non-mammalian vertebrate *in vivo* TMs for consideration (F.C.9, **Table 45**)
- propose the development of new non-mammalian vertebrate *in vivo* studies to better identify endocrine disrupting substances considering the added value from regulatory perspective (F.C.9, **Table 46**).

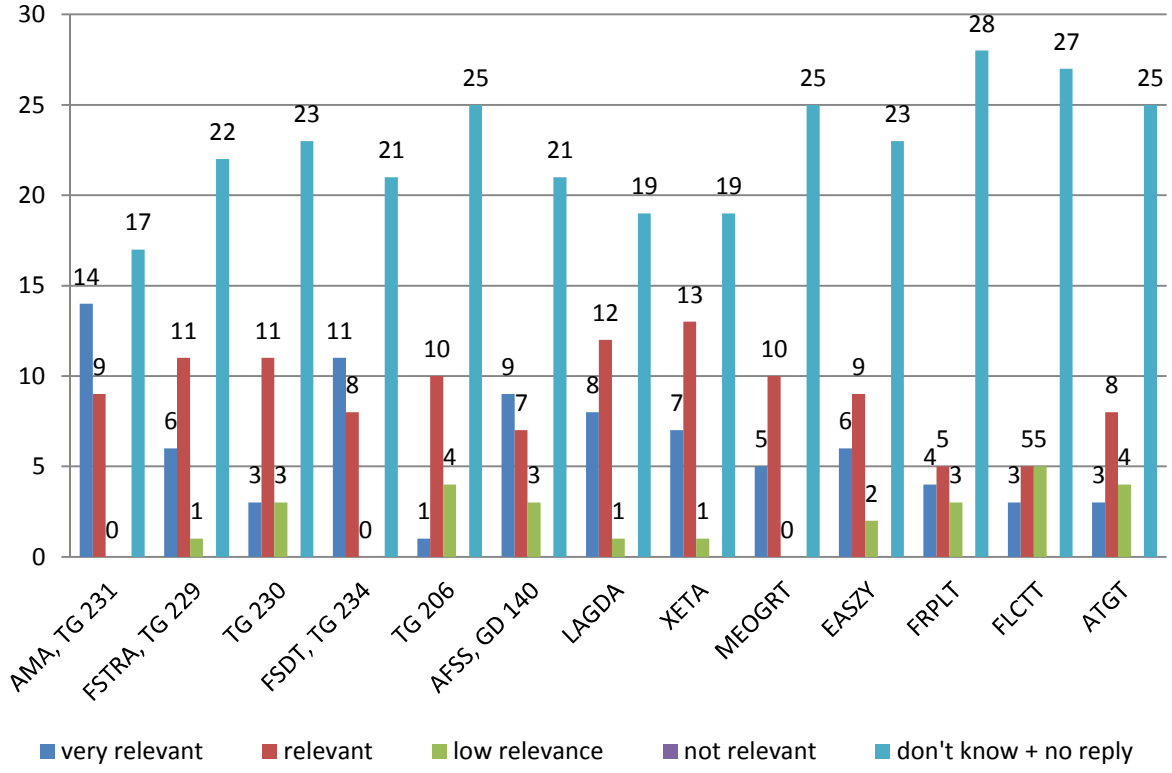

Figure 16: Replies of experts to question F.C.1: Please rate the non-mammalian vertebrate *in vivo* studies below regarding their relevance considering their diagnostic value to detect endocrine related perturbations in the test alone or in combination with other tests. (*Note:* The percentage of experts not replying or answering "don't know" was 42-70 %.)

(Abbreviations: AMA=amphibian metamorphosis assay; FSTRA=fish short term reproduction assay; FSDT=fish sexual development test; AFSS=androgenised female stickleback screen; LAGDA=larval amphibian growth and development assay; XETA=xenopus embryonic thyroid signalling assay; MEOGRT=medaka extended one generation reproduction test; EASZY=detection of endocrine active substances, acting through estrogen receptors, using transgenic cyp19ab-GFP zebrafish embryos; FRPLT=fish reproduction partial life cycle test; FLCTT=fish life cycle toxicity test; ATGT= Avian two generation test.)

Table 43: Expert comments on non-mammalian vertebrate *in vivo* studies regarding (1) their diagnostic value to detect endocrine related perturbations in the TM alone or in combination with other tests (F.C.2/3), (2) regarding the TM ability to distinguish endocrine related effects from general toxicity or indirect effects on the endocrine system (F.C.4/5), (3) proposing enhancements to the TMs (F.C.6/7).

General comments:
• diagnostic value to detect endocrine related perturbations in this test alone or in combination with other tests
1. We note that you use the wording perturbations, not adverse effects and MoA respectively. This might confuse the reader and the responses. Some of the assays provide information regarding MoA, others provide information regarding adverse outcomes. Both are relevant for assessing endocrine disruption. Most definitive /conclusive are assays which are at the highest tier of the OECD Conceptual Framework for endocrine disruption and which provide information about both endocrine activity and adverse effects, like the Fish Sexual Development test (TG 234) or The Enhanced One Generation Reproductive Toxicity Test (TG 443).
2. a) Both fish and amphibian tests could identify endocrine MOAs. These tests, however, focus only on EATS pathways. The diagnostic value of these tests is limited in terms of other

pathways. b) The aforementioned tests may produce false positive or negative results due to factors like concentrations, stress responses etc. The diagnostic value of a single test alone may be limited. Weight of evidence approach for taking all tests into account is important. c) Some tests like TG230 and the EASZY have only a diagnostic value and cannot be used for risk assessment. Other tests like TG229, 234 etc can be used for both MOAs and for adverse effects. Due to the complexity of these assays, there is a need for developing guidance documents for the interpretation of these assay results in the context of identification, classification and risk assessment of EDCs.
3. Ecotoxicity studies are conducted to assess the safety to the environment of the substance under investigation. Therefore apical endpoints related to the protection goal (populations) are most relevant. No test in isolation is capable of identifying a potential endocrine disrupter. Therefore the relevance of a given test is compound specific and depends on the outcome of other mechanistic or apical studies.
4. There is a need for an extended one-generation reproduction test with other species, e.g. Danio rerio.
5. The fish and amphibian tests indicated include endpoints that are fairly specific and sensitive to EDC.
• allow to distinguish endocrine disrupting effects from general toxicity or indirect effects on the endocrine system
6. In many cases it is not possible to answer clearly yes or no to this question. In some cases the assays will allow to distinguish, in other cases they will not. It depends on the substance and which other MoAs/AOPs it acts through.
7. Some modification of the existing test protocols like carefully choosing test concentrations, extra sampling points, extra endpoints, etc may be needed. Stress responses may confound the diagnostic value of the tests. Extra controls or in vitro results may be used to clarify the uncertainties.
8. None of the above mentioned tests in isolation is capable of distinguishing endocrine disrupting effects from general toxicity or indirect effects on the endocrine system.
9. They could with further exploratory work, which would also be applicable to mammalian hazard assessment.
• proposals for enhancing them
10. METiCx consortium proposals.
AMPHIBIAN
Amphibian metamorphosis assay (AMA, TG 231)
• diagnostic value to detect endocrine related perturbations in this test alone or in combination with other tests
11. With exception of thyroid pathology the other endpoints are non-specific growth measurement
12. Both the AMA and XETA provide very good diagnostic input for disruption of the thyroid axis at any level (metabolism/ distribution/ thyroid gland function).
• allow to distinguish endocrine disrupting effects from general toxicity or indirect effects on the endocrine system
13. not designed to detect general toxicity
• proposals for enhancing them
-
Larval amphibian growth and development assay (LAGDA) – Note by JRC: since July 2015

OECD TG 241
• diagnostic value to detect endocrine related perturbations in this test alone or in combination with other tests
14. LAGDA is useful, but misses out the important reproductive stage of the amphibian lifecycle
15. XETA and LAGDA assays are performed in FETAX medium that is lacking iodine. Effects of sodium-iodide symporter inhibitors might be overestimated when iodine is not present in the exposure medium. (Note by JRC: the XETA under validation is no longer using FETAX medium; LAGDA uses water)
16. limited validation of LAGDA and MEOGRTS, see summary record of WNT 2015
• allow to distinguish endocrine disrupting effects from general toxicity or indirect effects on the endocrine system
17. It is possible to use assays like TG234, LAGDA etc to distinguish endocrine MOAs from general toxicity.
• proposals for enhancing them
18. An amphibian model is needed which provides a practical option for conducting full lifecycle tests
Xenopus embryonic thyroid signalling assay (XETA) (ongoing validation)
• diagnostic value to detect endocrine related perturbations in this test alone or in combination with other tests
19. dynamic range appears to be limited
20. Both the AMA and XETA provide very good diagnostic input for disruption of the thyroid axis at any level (metabolism/ distribution/ thyroid gland function).
21. XETA and LAGDA assays are performed in FETAX medium that is lacking iodine. Effects of sodium-iodide symporter inhibitors might be overestimated when iodine is not present in the exposure medium. (Note by JRC: the XETA under validation is no longer using FETAX medium; LAGDA uses water)
• allow to distinguish endocrine disrupting effects from general toxicity or indirect effects on the endocrine system
22. not designed to detect general toxicity
• proposals for enhancing them
23. The XETA test could be enhanced in three prinicpal ways: (i) incorporation of a second fluorescent marker providing more precise readout of thyroid gland activity (eg NIS reporter), (ii) investment in robotisation methods to enable high throughput screening (iii) modulation of test conditions (eg use of low iodide) to optimise detection of certain chemical categories.
FISH
Fish short term reproduction assay (FSTRA, TG 229)
• diagnostic value to detect endocrine related perturbations in this test alone or in combination with other tests
-
• allow to distinguish endocrine disrupting effects from general toxicity or indirect effects on the endocrine system
-
• proposals for enhancing them

24. A combination of TG 229 and TG 234 would cover all relevant life stages and be diagnostic (identifying ED in the environment)
25. TG229 should be enhanced to be used for both screening and testing. The current protocol with three concentrations can only be used as a screening assay. Due to such a design, this assay cannot be used to derive a NOEC for risk assessment. As fecundity is an important reproduction parameter, enhancing this test by an inclusion of 5 test concentrations is important.
21-day fish assay (TG 230)
• diagnostic value to detect endocrine related perturbations in this test alone or in combination with other tests
-
• allow to distinguish endocrine disrupting effects from general toxicity or indirect effects on the endocrine system
26. not designed to detect general toxicity
• proposals for enhancing them
-
Fish sexual development test (FSDT, TG 234)
• diagnostic value to detect endocrine related perturbations in this test alone or in combination with other tests
27. There remains a question regarding the FSDT (OECD TG 234) with regard to how a significant result is determined (methods of statistical analysis when using ratios) and what degree perturbation of sex ratio of offspring is needed before an outcome can be considered an apical effect (although any significant change in sex ratio can indicate endocrine disruption per se).
28. TG 234 is the fish test with population relevance which is cheapest, shortest and uses least animals
29. … there are assays which provide information about both endocrine activity and adverse effects, like the Fish Sexual Development test (TG 234) …. (Note JRC: extracted from reply 1 under general)
• allow to distinguish endocrine disrupting effects from general toxicity or indirect effects on the endocrine system
30. It is possible to use assays like TG234, LAGDA etc to distinguish endocrine MOAs from general toxicity.
• proposals for enhancing them
31. A combination of TG 229 and TG 234 would cover all relevant life stages and be diagnostic (identifying ED in the environment).
Androgenised female stickleback screen (AFSS, GD 140)
• diagnostic value to detect endocrine related perturbations in this test alone or in combination with other tests
32. This should be a full TG as it is superior to the Hershberger assay; the way I view it if spiggin was present in zebrafish or fathead minnow this test should have been a TG; because it is normally expressed only in the stickleback is a GD; this make no scientific sense.
• allow to distinguish endocrine disrupting effects from general toxicity or indirect effects on the endocrine system

33. not designed to detect general toxicity
• proposals for enhancing them
34. The AFSS can be improved in terms of the 3Rs by incorporating the transgenic fish model (spiggin medaka) as a screen
Medaka extended one-generation reproduction test (MEOGRT) – Note by JRC: since July 2015 OECD TG 240
• diagnostic value to detect endocrine related perturbations in this test alone or in combination with other tests
35. limited validation of LAGDA and MEOGRTS, see summary record of WNT 2015
• allow to distinguish endocrine disrupting effects from general toxicity or indirect effects on the endocrine system
-
• proposals for enhancing them
36. as noted under F.C.3 there is a need for an extended one-generation reproduction test with other species, e.g. Danio rerio.
Detection of endocrine active substances, acting through estrogen receptors, using transgenic cyp19a1b-GFP zebrafish embryos (EASZY) *(ongoing validation)*
• diagnostic value to detect endocrine related perturbations in this test alone or in combination with other tests
-
• allow to distinguish endocrine disrupting effects from general toxicity or indirect effects on the endocrine system
37. not designed to detect general toxicity
• proposals for enhancing them
-
Fish reproduction partial lifecycle test (FRPLT)
• no replies
Fish life cycle toxicity test (FLCTT)
• diagnostic value to detect endocrine related perturbations in this test alone or in combination with other tests
-
• allow to distinguish endocrine disrupting effects from general toxicity or indirect effects on the endocrine system
38. TG 206 and FLCTT - lack of endocrine specific endpoints (2x)
• proposals for enhancing them
39. A combination of TG 229 and TG 234 would cover all relevant life stages and be diagnostic (identifying ED in the environment)
BIRDS
Avian reproduction test (TG 206)

• diagnostic value to detect endocrine related perturbations in this test alone or in combination with other tests
40. TG 206 has limited diagnostic value as it is relatively short-term.
• allow to distinguish endocrine disrupting effects from general toxicity or indirect effects on the endocrine system
41. TG 206 and FLCTT - lack of endocrine specific endpoints
42. not designed to distinguish endocrine effects from general toxicity since no specific endocrine endpoints are evaluated in this test.
43. I am not sure if this test allows to clearly distinguish ED from general toxicity.
44. TG 206 currently contains only apical endpoints which do not solely respond to EDs; some of the endpoints are potentially affected by EDs.
45. The two avian tests cannot distinguish ED modes of action.
• proposals for enhancing them
46. I recommend that a research programme be set up with the aim of developing an optimized, targeted approach to protecting birds against endocrine disrupters which leave room for reconsideration and possible adaptation of the current avian tests. This research programme will address the following three questions: 1) Which core characteristics (traits) make birds unique in their potential responses to (certain) endocrine disrupting chemicals. 2) Which MOAs (modes of action) and associated specific endpoints give rise to effects in birds that would not be seen in mammals or other taxa. 3) Which targeted, step-wise in vitro and small scale in vivo protocols should be developed to specifically address these differences between birds and other taxa. The resulting thesis project will provide the following four deliverables; Filling the data gap of fundamental differences between birds and other taxa which would make it necessary to establish the risk of endocrine disruptors to birds as well. b) A comprehensive testing approach that will allow the regulatory process to protect birds from endocrine disruptors with optimal use of resources including minimal animal testing. c) The development of more dedicated, specific mechanism-based in vitro and in vivo assays, to be able to provide no more and no less information than is needed based on the information already available for a compound. d) Reduction of animal use according to the 3Rs (Note JRC: this reply was also given to Avian two-generation test (ATGT)).
Avian two-generation test (ATGT)
• diagnostic value to detect endocrine related perturbations in this test alone or in combination with other tests
47. Avian two-Generation test: It has not been sufficiently established in the validation process to date that this study is sensitive and focused enough to identify unequivocal endocrine effects in birds that would not be revealed in mammalian test systems. In the validation study of the Japanese quail two-generation study there were no substantial responses in Japanese Quail in any of the inter-laboratory tests performed with a known endocrine disrupting compound, vinclozoline (EPA Report, May 2013). Furthermore because of the lack of response it is difficult to differentiate test endpoints which are sensitive and informative from those which are not.
• allow to distinguish endocrine disrupting effects from general toxicity or indirect effects on the endocrine system
48. See TG 206
• proposals for enhancing them
49. See TG 206

Table 44: Expert other comments on non-mammalian vertebrate tests in the OECD CF (F.C.8)

1. The very limited validation of the LAGDA and MEOGRTS were discussed at the last OECD WNT meeting and introduced in the introduction of the TG. Also the difficulties to build up historical control databases for such complex, expensive animal testing methods was discussed, which is further complicated by the fact that there is no OECD agreement yet on a single or limited number of fish species to be used for testing. This reflects the general difficulties with the validation of complex, expensive animal tests and consequently with the reliance on such complex animal tests for regulation.
2. In our view, the current ED CF (Conceptual Framework); GD 150 (Guidance Document On Standardised Test Guidelines For Evaluating Chemicals For Endocrine Disruption) should be supplemented with some general recommendations in testing strategies to consider ED activity/related effects across species. Currently, in vitro screens are relevant for effects in humans and vertebrate wildlife because many of them 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 adverse effects in vivo both in non-mammalian vertebrate species and mammalian species (including wildlife species and humans). A novel paper concludes that when comparing data from fish and rat assays, a high concordance was seen with respect to identifying chemicals that impacted specific endocrine pathways of concern (Ankley and Gray, 2013). Although most chemicals were detected as positive in both the rat and fish assays, eliminating data from one class of vertebrate would weaken the battery. 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 (Ankley and Gray, 2013). Due to the similarities between different vertebrate species it seems to be relevant to consider ED related concerns identified in one species (e.g. in fish or amphibians) as of potential relevance for other species (such as rat or humans); in particular in cases when no or only limited data is available for the latter mentioned species. The current ED Conceptual Framework and Guidance Document do not at the higher levels (from level 3 and onwards) reflect this adequately. Instead the current approach is that "human health concerns" are covered by rat /rodent TGs/data only, whereas environment by those available on fish, amphibian/birds and to some extent invertebrate data only (It is as far as we know currently concluded that the endocrine system of invertebrates - besides perhaps ecdysone system in insects- is not well enough known to be able to establish links between hormone activity and adverse effects indisputably). In our view, this needs further discussion e.g. of the need for supplementing the current Conceptual Framework or GD 150 with some (more) considerations: concerns in e.g. fish for estrogenicity or thyroid effects in amphibians and absence / or clear limitations in availability of relevant data in rodents does in our view give concerns for potential effects in mammalians (e.g. rat and humans) and vice versa. Currently TGs with rodents are typically only regarded as relevant for human health. But, as already mentioned at OECD EDTA meetings and in recent discussions in the REACH Member State Committees concerning substance evaluation cases and nomination proposals for the candidate list, rodent data should be considered in relation to humans health as well as for mammalian wildlife with respect to information regarding endocrine disruption.
3. When evaluating the potential for a substance to act as an ED in the environment an adverse effect with negative consequences on the population must be demonstrated. Several of the above non-mammalian test include endpoints that, while may be useful in a weight-of-evidence approach, do not in and of themselves demonstrate adverse impacts (e.g. VTG induction). Thus, when evaluating adverse population effects, higher priority should be placed on those studies that directly measure fertility, fecundity and/or reproduction. Additionally, in assays which require aqueous exposures care should be taken to ensure environmentally relevant test concentrations.
4. See avian tests above: proposed research programme with the aim of developing an optimized, targeted approach to protection birds against endocrine disrupters which leave room for reconsideration and possible adaptation of the current avian tests.
5. General remark: A number of endocrine modes of action (e.g. effects on the corticosteroid system and on the endocrine control of neural development will most likely not be detected with the available fish screening tests. If aquatic vertebrates shall be protected from all adverse endocrine effects, this is a shortcoming of the current testing framework. Concerning

fish tests: Based on reviews and the evaluation of a number of studies it is concluded that effects on indicative and / or apical endpoints of fish screening tests and the fish full life-cycle test are generally at least as sensitive as effects on fish reproductive behaviour that are evaluated in current standard test (see sections 2.5.1 and 3 of UBA report). Thus, there is a low risk that significant effects of sexual endocrine disruptors on fish are missed, if the assessment of endocrine effects is based on the tiered testing strategy as included in Appendix 7.8-5 of R.7b (ECHA 2008a). Effects on spermatogenesis and oogenesis, i.e. non-standard endpoints, exhibited a particularly high sensitivity to bisphenol A, especially for trout. The LOEC derived for effects of bisphenol A on sperm density and sperm motility in brown trout. - the endpoint is often delayed maturity of sperma, meaing fish can only reproduce later, therefore this is a population relevant endpoint. Similar effects were seen with alcylphenoles (here seasonal spawning fish were affected more than frequent spawners like medaka. Concerning ATGT: From a regulatory point of view, it should be preferred to gain evidence of endocrine disrupting properties in non-mammalian vertebrates from enhanced one-generation tests and make full use of the information that could be obtained from these rather than leaving too many questions open to be answered by a two-generation test.

Table 45: Expert comments regarding additional non-mammalian vertebrate in vivo TMs (free text comments to question " Apart from what was addressed above, are you aware of any existing additional non-mammalian vertebrate in vivo test that could add value and should be further considered?" (F.C.9)).

1. Numerous fluorescent fish embryo based tests are being developed for all the major endocrine axis.
2. The OECD CF did not include fish long term toxicity tests required by REACH and BPR (biocidal product regulation). These tests include fish early life stage test (TG210), fish Short-term Toxicity Test on Embryo and Sac-fry Stages (TG212) and fish juvenile growth test (TG215). These tests may have limited diagnostic value but can indicate the adverse effects, which is important for identification, classification and risk assessment of EDCs. Specifically, TG210 is requested by various pieces of EU legislation including REACH, PPPR, BPR, Regulations for veterinary and human pharmaceuticals, and Regulation for the feed additives. It is suggested that at least TG210 should be included in the OECD CF.
3. See avian tests above: proposed research programme with the aim of developing an optimized, targeted approach to protection birds against endocrine disrupters which leave room for reconsideration and possible adaptation of the current avian tests.

Table 46: Expert proposals for the development of new non-mammalian vertebrate *in vivo* studies to better identify endocrine disrupting substances (free text comments to question " Apart from what was addressed above, do you see a need for developing new non-mammalian vertebrate in vivo tests to better assess endocrine disrupting substances? Will there be added value from such new tests from a regulatory perspective?" (F.C.10)).

1. Efforts should be made to overcome the fact that we have no current studies that cover the full Life cycle (exposure embryo/fetal stage - late effects)
2. Many of these fluorescent fish embryo based tests will provide added value from a regulatory point of view.
3. In general, in order to address which in vivo tests to possibly develop, it might be feasible to consider development of AOPs as the first step, and target MIEs, KEs and AOs with development of promising in vivo tests.
4. Yes. For example, the fish test method for addressing the HPA axis is currently under development. Zebrafish has been used for studying bone disorders and obesity. Such fish models are of high value for regulating EDs.
5. Most of the chronic toxicity and ecotoxicity studies do not match typical exposure scenarios for crop protection chemicals. While in the toxicity studies exposure is kept constant,

exposure of organisms in the field is short-term and maybe repeated. Therefore such test designs could help to elucidate whether unacceptable endocrine disrupting effects could occur under relevant field conditions.
6. See avian tests above: proposed research programme with the aim of developing an optimized, targeted approach to protection birds against endocrine disrupters which leave room for reconsideration and possible adaptation of the current avian tests.
7. It may be necessary to develop new tests with regard to MoAs other than EATS (for example stress response, PPAR, Vitamin D and so on) with further knowledge it may be necessary to develop test for other species (birds, reptiles) or add relevant species to existing TGs
8. …they can be used for exploratory work to identify key biomarkers applicable to current TGs.
9. Please see comments under F.B.9 - also applies here: New apical mammalian studies would continue to present similar interpretive difficulties as existing tests with respect to distinguishing endocrine-mediated from other effects, unraveling significant effects from homeostatic responses, and translating conclusions across species, where underlying mechanisms continue to be poorly understood. In view of this, and considering resource limitations, developing additional mammalian in vivo tests will contribute less to our ability to identify endocrine disruptors than assays supplying mechanistic information and efforts to arrange this into predictive AOPs. (Note JRC: copied from F.B.9)

JRC Summary of the comments in free text fields of Question F.C.2-F.C.10 regarding non-mammalian vertebrate TMs

Birds

There is agreement that neither the avian reproduction assay (OECD TG206) nor the avian two generation assay are diagnostic for ED or can distinguish ED effects from general toxicity effects. Research is proposed which would address ED MoAs specific for birds and not be found with any other assay, e.g. mammalian.

Amphibian

Current amphibian tests (AMA, LAGDA, XETA [ongoing validation]) are diagnostic for "T". With the LAGDA it is possible to distinguish ED effects from general toxicity effects, however, it does not cover reproduction and there is limited experience with the LAGDA.

Fish

There is agreement that current fish tests are diagnostic for EAS. Tests like TG234 or MEOGRT (TG 240) are able to distinguish ED effects from general toxicity effects. Combination of TG229 and TG234 is considered of added value, since this would cover all relevant life stages and be diagnostic (identifying ED in the environment).

3.5.4 Invertebrate test methods (F.D)

In section F.D of the survey, experts were asked to:

- comment on invertebrate tests in the OECD CF regarding e.g. their usefulness for endocrine disruptor identification, technical performance, weaknesses and limitations, regulatory acceptance, etc. (F.D.1/2, **Table 47**)
- mention additional existing invertebrate TMs for consideration (F.D.3, **Table 45**)
- propose the development of new invertebrate studies to better identify endocrine disrupting substances considering the added value from regulatory perspective (F.D.4,**Table 48**).

Table 47: Expert comments on invertebrate studies in the OECD CF (free text comments to question "Do you have any comments on any invertebrate test from the OECD CF (e.g. usefulness of test for endocrine disruptor identification, comments on technical performance, weaknesses and limitations, regulatory acceptance, etc.)? (F.D.1/2)

General comments
1. More research is needed in relation to Development of endocrine specific endpoints in molluscs (and other invertebrates)
2. Most of these assays are useful apical tests, but none have diagnostic value.
3. Due to the limited knowledge in endocrinology of invertebrates, the diagnostic value of the above tests is rather poor and for most tests, hardly known. The current development of mollusc toxicity test, for example did not include any diagnostic endpoints. These tests are often used for detecting adverse effects but not for MOAs . Test methods with diagnostic value should be further developed.
4. The majority of invertebrate tests has the reproductive output as its only ED-relevant endpoint. Without additional endpoints like sex ratio or intersex, and in particular endpoints that can link the reproductive output to a specific MoA, these tests will be difficult to use for ED identification and have only limited diagnostic value. Effects may not be clearly distinguishable from general toxicity. Only the three marked tests *(**Note JRC**: Mysid life cycle toxicity test; Copepod reproduction and development test;Sediment water Chironomid life cycle toxicity test (TG 233))* allow the conclusion of ED activity. The existing invertebrate tests should be enhanced with biomarkers for ED MoAs were possible to be of more use in a regulatory context. Furthermore, in vitro test should be developed for relevant invertebrate MoAs.
5. Invertebrate tests will be needed for regulatory environmental risk assessment (e.g. REACH, pharmaceuticals). None of the tests mentioned above allow to distinguish between systemic toxicity and endocrine related effects. Thus they are not useful at the moment for the identification of EDs. UBA published a scientific report on an Evaluation of uncertainties in the environmental risk assessment of endocrine active substances (submitted separately, especially chapter 5 is of relevance here). That report states that it should be noted that the apical endpoints of such life-cycle tests allow the identification of adverse effects of EDCs, but do not provide causal evidence of an endocrine mode of action (OECD 2006a, 2010a, LeBlanc 2007). In order to unequivocally identify endocrine disruption as the underlying mode of action, specific diagnostic endpoints are required. However, considerable gaps in the knowledge on endocrine system of most invertebrate taxa with the exception of insects, crustaceans (DeFur et al. 1999b, IPCS 2002) and, partly, gastropods hamper the identification and development of appropriate diagnostic endpoints (see sections 2.1.5 and 2.2.6 of the report). As emphasised for example by Soin & Smagghe (2007), more research is needed in most cases to allow a mechanistic understanding of the relationship between the mechanism of action of the substance and the adverse effect in invertebrates. The specificities in invertebrate endocrinology (e.g. the importance of ecdysteroids and terpenoids) are likely to result in specific susceptibilities to endocrine disrupting chemicals (IPCS 2002). In combination with the fact that invertebrates are not adequately represented in REACH

guidance document R.7b, there is a high risk to miss effects of an EDC on certain invertebrate species / taxa (Lafont 2000, Oehlmann & Schulte-Oehlmann 2003, OECD 2006a, Hutchinson 2007), i.e. a considerable uncertainty with regard to the protection of wildlife invertebrates. This is exemplified by the high sensitivity of molluscs to tributyltin and triphenyltin and the associated population declines in many prosobranch species. In addition, molluscs, copepods and echinoderms proved to be particularly sensitive to bisphenol A and 4-tert-octylphenol, i.e. invertebrates possessing vertebrate-type hormones may exhibit a higher sensitivity to substances interacting with vertebrate-type endocrine processes than vertebrates. An environmental risk assessment procedure for potential endocrine disrupters should be based on tests with representatives from the most relevant taxonomic groups, including cnidarians, annelids, crustaceans, insects, molluscs and echinoderms (Hutchinson 2002, Matthiessen 2003, Oehlmann & Schulte-Oehlmann 2003, OECD 2006a).
6. Need to know more about comparative endocrinology
Mollusc partial lifecycle assay & full lifecycle assay & Mollusc Reproductive Toxicity Test (JRC note: in the meantime available: OECD TG 242 *Potamopyrgus antipodarum* Reproduction Test and OECD TG 243 *Lymnaea stagnalis* Reproduction Test) *(same replies to the three tests, merged)*
7. With particular regard to the CF which specifies EATS effects (paragraph A.11 of GD 150) Mollusc assays (repro, partial and full life cycle): Two species are currently included; Potamopyrgus anitpodarum is parthenogenetic, making it unsuitable for assesing effects on the male. Lymnaea stagnalis is hermaphroditic, making assessments on the male difficult, and the species appears less sensitive to potential effects than dioecious species, possibly as they can self fertilize. 8. More research is needed in relation to Development of endocrine specific endpoints in molluscs (and other invertebrates) 9. The mollusc partial life cycle assay is now almost ready to be published as an OECD test guideline.
Chironomid toxicity test (TG 218-219)
10. Chironomid assay (TG 218/219): is not suitable as a test on the EATS pathway since it does not cover reproduction.
Daphnia reproduction test (with male induction) (TG 211) & Daphnia multigeneration assay *(same reply to both tests, merged)*
11. Daphnia reproduction test and multigen assay: is not a suitable representative for the crustacea as while this group appears to be sensitive to vertebrate steroids, this test does not cover effects on the male, as the species is parthenogenetic unless environmentally stressed.
Earthworm reproduction test (TG 222)
12. The species used are hermaphroditic making assessment on male effects difficult 13. In general, they seem to have a significant limitation to distinguish ED effects from general toxicity.
Enchytraeid reproduction test (TG 220)
14. The species used are hermaphroditic making assessment on male effects difficult 15. In general, they seem to have a significant limitation to distinguish ED effects from general toxicity.
Sediment water lumbriculus toxicity test using spiked sediment (TG 225)
16. Lumbriculus test: is essentially a biomass endpoint, so not relevant.
Predatory mite reproduction test in soil (TG 226)
17. Predatory mite test: effects on the male are not assessed as males are not introduced into the test. 18. In general, they seem to have a significant limitation to distinguish ED effects from general

toxicity.
Collembolan reproduction test in soil (TG 232)
19. Collembolan reproduction test: species commonly used (Folsomia candida) is parthenogenetic so that effects on the male are not assessed. 20. In general, they seem to have a significant limitation to distinguish ED effects from general toxicity.
Mysid life cycle toxicity test
-
Copepod reproduction and development test
-
Sediment water Chironomid life cycle toxicity test (TG 233)
-

Table 48: Expert replies regarding additional invertebrate TMs (comments in free text fields to question " Are you aware of any additional invertebrate test already available that would add value? & Do you see any need for developing additional tests on invertebrates? Will there be added value from such tests from a regulatory perspective? (F.D.3 and F.D.4)).

General
1. There is a general need for Development of (non-vertebrate) endocrine specific endpoints in invertebrates
2. Yes, we urgently need a range of invertebrate endocrine mechanistic screening assays, but this will depend on progress being made in understanding invertebrate endocrinology in greater depth.
3. Yes. Endocrinology of invertebrates differ greatly from that of vertebrates. Some pesticides have been specially designed to target the endocrine system of invertebrates, e.g. juvenile hormone, which is unique to invertebrates.
MOA/AOPs
4. Invertebrate tests that allow definitive ED MoA and adversity characterisation would be of a benefit from regulatory perspective.
5. The development of invertebrate tests that allow definitive ED MoA and adversity characterisation may be an interesting goal in terms of testing throughput and 3R policy concern. However this is likely to be a medium term research goal rather than a goal closer to application.
6. Yes but it requires a well-founded insight into the biology of the hormonal systems of the test organisms. In general, in order to address which in vivo tests to possibly develop, it might be feasible to consider development of AOPs as the first step, and target MIEs, KEs and AOs with development of promising in vivo tests.
7. Enhancing the suite of existing test with relevant biomarkers and endpoints would be more efficient.
Bivalves
8. Need to develop a test using bivalves as this group appears to be the most sensitive to vertebrate steroids over a wide range of species.

Gastropods
9. Need to develop a test using a dioecious species of gastropod as this group is also sensitive and currently effects on the male, including mating behaviours, are not covered
Echinoderms
10. Need to develop a test using an echinoderm species as this group appears sensitive to both vertebrate thyriod hormones and sex steroids. 11. Echinoderms are very important for biodiversity in marine and coastal environments and are a phylum with highly specialized characteristics, Yet there are no tests for endocrine disruption in echinoderms
Arthropods
Crustaceans
12. The mysid and copepod tests are relevant for effects in the marine environment, but a dioecious freshwater equivalent test is needed (other than daphnia, in which effects on the male are not addressed).
Hexapoda
13. The collembolan reproduction test using the dioecious species (Folsomia fimitaria) would be a relevant test for effects on the male, but may need further validation at OECD. (F.D.3 – 1)
Nematodes
14. The knowledge for using C. elegans in reprotoxicity testing is available, but standardization of protocols/endpoints is needed.
15. C. elegans reprotoxicity tests could be considered for inclusion. Studies with this nematode show good predictivity for DART compounds, can be used for multi-generational studies and are relatively easy to perform. Moreover, this nematode is very suitable for pathway screening especially when combined with siRNA strategies and/or fluorescent markers. (F.D.3 – 2)

JRC Summary of the comments in free text fields of Question F.D regarding invertebrate TMs

The replies indicate that there is limited knowledge in endocrinology of invertebrates. Current tests on reproduction lack effects on males, since the species used are parthenogenetic; reproduction test with dioecious species are recommended; however, validation of tests would be necessary.

In general, in order to address which *in vivo* tests to possibly develop, it might be feasible to consider development of AOPs as the first step, and target MIEs, KEs and AOs with development of promising *in vivo* tests.

3.6 Experts final remarks in the survey

In the last section (section G) of the survey the experts expressed their final remarks in the free text fields concerning

- new / enhanced test methods that could lead to the greatest contribution to the protection of human health and the environment in a **short-term perspective** (**Table 49**)

- new / enhanced test methods that could lead to the greatest contribution to the protection of human health and the environment in a **long–term perspective** (**Table 50**)
- any other issue related to the questionnaire and providing relevant references (**Table 51**).

Table 49: Expert opinions regarding which new or enhanced TMs would lead to the greatest contribution to the protection of human health and the environment in a short-term perspective (free text comments to question G.1).

1.	Mixture models. In vitro tests for molecular initiating events that perturb thyroid pathways. Same comment for corticosteroids.
2.	Enchanced 416, 443 and chronic/carcinogencity assays
3.	Endocrine mechanistic screening tests for invertebrates, and screening tests for interference with the adrenal/interrenal stress response in fish and other vertebrates
4.	Any test method with more focus on human, mechanistically relevant endpoints as opposed to overt toxicity endpoints in rodents.
5.	Functional in vitro assays using clinically relevant biomarkers; in vitro assays relevant to metabolic synrome; juvenile toxicity roent test
6.	human cell culture models and whole blood gene expression
7.	whole animals, behavioral tests
8.	Enhancement of existing methods
9.	The uterotrophic assay is a short-term screening test sensitive enough to detect potent substances effects at even low concentrations, but of relevance for mammalians, invertebrates and aquatic life. This method could be enhanced with the addition of an anti-estrogenic positive control, uterus morphometry and plasma hormone levels quantification. The inclusion of these endpoints would better characterize not just estrogenic and anti-estrogenic pathways, but also steroidogenesis and HPG axis relation. Another relative short-term assay that could be conducted in the same way as the uterotrophic, but in males is a modified Hershberger assay, using immature rats and androgenic/antiandrogenic positive controls and other 2-3 doses of the test substance only, until the end of one entire period of sperm maturation (52 days). Sperm parameters, testes/epididymis morphometry and hormone analysis would complete the test evaluation in the same animals at once. The extended one-generation assay (with the suggestions of evaluation until F2, multiple time-points of blood and hormone levels collection/quantification and gonads morphometry) is the best and most complete experimental design evaluation for long-term endocrine disruption analysis of all windows of susceptibility, sensitivity and reproductive development. As the current in vivo methods would miss capturing effects triggered during pregnancy but showing up very late in life, an enhanced long term/carcinogenicity study with exposure in utero (merging developmental and long term studies) would be expected to fill this gap. In the pesticides area, long term studies are always required and this enhancement would add values without increasing the number of test animals.
10.	As commented earlier, MoA/AOP informed in silico/in vitro focussed IATA development and assessment with improved metabolism and human/environment relevance for EATS and non-genotoxic carcinogenicity should be the high priority focus. See comment to DA7.
11.	A feasibility study for minor enhancements of TG 414 (Prenatal Developmental Toxicity Study) with ED-relevant endpoints has been proposed by DK and is now on OECD workplan. Need for mode of action data both in vivo and in vitro to validate the AOP approach.
12.	EOGRTS
13.	Biologically relevant in vitro tests in the thyroid signalling pathway, Enhancement of existing test methods with ED-specific endpoints, Invertebrate tests, that allow definite MoA and adversity characterisation MoA/AOP informed in vitro/in silico based IATA development iucluding improved metabolism and non-genotoxic carcinogenicity

14.	ED related enhancement of existing guidelines which are already part of standard info requirements or of relevance for substance evaluation under REACH (e.g. TG414 - which is now on the workplan, TG 407, TG 408, TG 426, TG 443,). Guidance on cross species considerations: In our view, the current ED CF (Conceptual Framework); GD 150 (Guidance Document on Standardised Test Guidelines for Evaluating Chemicals for Endocrine Disruption) should be supplemented with some general recommendations in testing strategies to consider ED activity/related effects across species. The rat should be considered a model for rodents in the environment, not only for human health. Likewise, ED MoA observed in non-rodents vertebrates like fish and amphibians should raise a concern for human health. See also our comment in F.C.8. The following more general issues which are relevant, but not specific for endocrine disruptors, would be valuable to prioritize development of: Inclusion of metabolic capacity in in vitro assays and development of better PBTK modelling Guidance on modelling of Bench Mark Dose Level in existing in vivo testing to put more emphasis on dose/response characterisation Guidance for use of ex-vivo assays and data
15.	We need to add more developmental perspective. mainly prenatal. More behavioral tests are needed
16.	Development of more holistic tests conducted with developmental exposures
17.	The AFSS (GD 140) as a test for androgens/anti-androgens
18.	This is highly dependent on the regulatory framework and the possibility to ask /obtain relevant information. As long as the REACH Annexes are not updated, and it is and remains very difficult, if not impossible, to ask for in vitro screenings methods/information and specific functional apical tests, the update of existing methods seems to be the most efficient manner to get further information. It is acknowledged,however, that these kinds of updates will not be enough to detect developmental neurotoxicity, developmental immunotoxicity, metabolic disorders, intermediate epigenetic effects, etc. The same applies for the environment where the information is also very limited and the possibility to request mechanistic information is limited and the possibility to ask information on various taxa is even more limited. For PPP and biocides although the data requirements are different and more information is required for all substances, still further information on ED mode of action and/or apical endpoints is only triggered. This means that also for PPP it is important that adequate parameters that could trigger an ED mode of action are included in the existing TM. So for the short term updating of existing methods with relevant parameters to detect potential ED mode of action has high priority.
19.	The development of in vitro tests which resemble closest the in vivo situation. Usually these are ex-vivo models which also allow to study species differences and may allow to judge on human relevance The greatest contribution to the protection of human health in the short-term may not be to necessarily add more endpoints or to develop more tests, but to establish clear bounds of interpretation for the test methods and endpoints we are currently using. Some measured endpoints in OECD tests are still debated as to their biological and toxicological significance (e.g. AGD, sperm parameters). If we cannot use the existing approaches to inform decisions, it is unclear how new approaches will help clarify matters.
20.	I believe that addressing the data gap regarding MOAs and associated specific endpoints that give rise to effects in birds that would not be seen in mammals or other taxa wil provide a significant contribution to the protection of birds in short-term and long-term perspective.
21.	At the present time, tests leading to more sensitive detection of gonadal and thyroid disruption in the fetus would have the biggest impact. Bearing in mind that for gonadal disruption, besides known xenoestrogens, the phthalates are currently the class of chemical with biggest known single effect, but are currently without appropriate in vitro tests, then new assays targetting Leydig cell differentiation and development should have a high priority.
22.	Thyroid assays, GR Transactivation test and adrenal steroid synthesis
23.	As mentioned above an extended one-generation test with zebrafish would contribute in a short-term perspective.
24.	incorporation of metabolism into in vitro tests Thyroid hormone pathway in vitro test validation MoA/AOP informed in silico/in vitro focussed IATA development and assessment with improved metabolism and human/environment relevance for EATS and non-genotoxic carcinogenicity should be the high priority focus.

25.	Please see responses to E.A.3, F.A.3 and F.A.4.
26.	From both a short and long-term perspective, methods that inform modes of action (particularly thyroid hormone) and metabolism are the most critical needs for ED assessment. This would allow relatively quick screening of the possible universe of chemicals to identify priorities for further screening. In addition, this information should be used to improve AOPs and Integrated Approaches to Testing and Assessment that would improve the ability to generate useful information for risk assessment in the future.
27.	Screen for thyroid disruption

Table 50: Expert opinions regarding which new or enhanced TMs would lead to the greatest contribution to the protection of human health and the environment in a long-term perspective (free text comments to question G.2).

1.	New in vitro, mechanistic test methods for other pathways than AETS modalities
2.	Invertebrate tests including endocrine specific endpoints are highly needed in both environmental protective and regulatory perspectived
3.	For nearly five decades long-term studies in rodents have been the accepted benchmark for assessing long-term toxicity, and particularly carcinogenicity, of chemicals. With exposures typically lasting about 2 years, long-term bioassays using rats has been a fundamental component of toxicity evaluation for food additives, pesticides, pharmaceuticals, occupational/industrial chemicals, and all manner of product formulations, byproducts, and environmental contaminants. Recently the European Food Safety Authority (EFSA)/ European Union (EU) and the World Health Organization (WHO) pointed out that the current set of internationally accepted test methods capture only a limited part of the chemicals adverse effects, particularly concerning potential endocrine disrupting activities, as current tests are inadequate to detect them. Different testing strategies are needed for identifying biologically significant adverse effects, including low dose effects, consequences from early life-stage exposures, non-monotonic dose response curves, impact of chemical mixtures, or neuro/behavioral effects : these end-points. are often not detectable with current range of toxicity testing. Testing requirements need to be expanded and validated, for both animal testing and alternative in vitro/in vivo methods, in order to generate more adequate and inclusive data relevant to protecting public health. We propose an integrated and comprehensive long-term toxicity/developmental/carcinogenicity bioassay capable of generating information on a broad spectrum of different end-points and relevant hypotheses. Sprague Dawley rat, already in use for carcinogenicity bioassays by most organizations including the Ramazzini Institute and the National Toxicology Program, and the Endocrine Disruptor Screening Program of the U.S. Enviromental Protection Agency [EPA], has been demonstrated as an appropriate and relevant model for identifying, extrapolating and predicting several toxic effects in humans. Therefore integrated toxicological tests on Sprague Dawley rat represents a unique opportunity for investigating multiple toxicological end-points at once, sparing animal lives in accordance to the 3Rs (replacement, reduction and refinement). An integrated study design based on a stepwise process complies and expands the state of the art of current guidelines: the Organization for Economic Co-operation and Development (OECD) Guidelines 453, 443 (OECD 2009, 2011)
4.	Life cycle test with amphibians
5.	Test methods to predict epigenetic (transgenerational) effects.
6.	Epigenetic markers in enocrine disruption testing. In the ecotoxicological field assays for endocrine disruption in echinoderms
7.	animal models
8.	multigenerational studies
9.	Test methods to evaluate metabolic disturbances
10.	As soon as IATAs have been established for the earlier mentioned fields, the experience and resources shall be used to extend to other fields, as Retinoic Acid, HPA axis and others. In long term also epigenetic MoA should be targeted by appropriate in vitro focussed IATAs.

11.	Need for mode of action data both in vivo and in vitro to validate the AOP approach.
12.	Update relevant OECD Test Guidelines to include endocrine disrupting effects taking into account Mode of Action.
13.	Methods tackling deseases involving diabetes and obesitas as well as immune system/stress related disorders in wildlife animals and humans. Also epigenetic effects should be tackled.
14.	Development of TGs to detect MoA and AO of thyroid disrupting substances predictive for effects in humans. More basic longer term research is also needed.
15.	Low-dose developmental exposures in rodents - and translation to primates
16.	HPA and Thyroid axis are way behind and need our attention;
17.	In the long term it is essential that we develop non-animal/in vitro screening batteries to detect potential EDs. If possible this should be included in a high throughput system. In addition further apical test both for human health and the environment needs to be developed. For human health functional tests on for example developmental neurotoxicity, developmental immune toxicity and metabolic syndrome are considered essential for regulation according to the definition of the WNO/IPCS. For environment new and enhanced test methods to be developed should cover the protection of the myriad of species and complex ecosystems. Sensitive species of vertebrates and invertebrates including mammals, birds, amphibians, reptiles, fish, insects, etc should be taken into account. Besides tests, structured testing strategy is of utmost importance.
18.	I believe that addressing the data gap regarding MOAs and associated specific endpoints that give rise to effects in birds that would not be seen in mammals or other taxa will provide a significant contribution to the protection of birds in short-term and long-term perspective.
19.	As in G.1
20.	Somatotropic (TR and GR transactivation assay) and Vitamin D pathways (VDR, AhR transactivation assays and Vitamin D hydroxylase assay) Inclusion of epigenetic studies?? (we are not experts in epigenetics, but the approach sounds promising)
21.	see Jacobs et al ALTEX 2013 and Jacobs and Greally ALTEX 2013. As soon as IATAs have been established for EATS the experience and resources could be extended to other fields, such as Retinoic Acid, HPA axis etc. In long term also epigenetic MoA should be targeted by appropriate in vitro endocrine focussed IATAs.
22.	Please see responses to E.A.3, F.A.3 and F.A.4.
23.	Issues that would be better addressed in the longer term are epigenetic effects (once the mechanisms are better understood), as well as linkages between the different hormone pathways (so that particular adverse outcomes could be better predicted from upstream molecular events).
24.	A comprehensive battery of tests to identify effects on brain development re humans

Table 51: Expert general remarks and relevant references provided (replies in free text fields to question G.3).

1.	The validation of new methods for the study of EDs is a very important issue in which our Institute is involved from many years..
2.	close cooperation with clinical endocrinologists and biobanking
3.	See comment DA7. References: OECD, New Scoping Document on in vitro and ex vico assays for the identification of medulators of thyroid hormone signalling. 2014, Series on Testing and Assessment. Greally, J.M. and M.N. Jacobs, In vitro and in vivo testing methods of epigenomic endpoints for evaluating endocrine disruptors. ALTEX, 2013. 30(4): p. 445-71. Jacobs, M.N., et al., In vitro metabolism and bioavailability tests for endocrine active substances: what is needed next for regulatory purposes? ALTEX, 2013. 30(3): p. 331-51. OECD, The use of metabolising systems for in vitro testing of endocrine disruptors. A detailed review paper for the OECD No.

	97. Paparella, M., et al., Uncertainty of testing methods--what do we (want to) know? ALTEX, 2013. 30(2): p. 131-44. WHO, Harmonisation Project Document No 11. 2014. Edler, L., et al., Selection of appropriate tumour data sets for Benchmark Dose Modelling (BMD) and derivation of a Margin of Exposure (MoE) for substances that are genotoxic and carcinogenic: considerations of biological relevance of tumour type, data quality and uncertainty assessment. Food Chem Toxicol, 2014. 70: p. 264-89 Meek, M.E., et al., New developments in the evolution and application of the WHO/IPCS framework on mode of action/species concordance analysis. J Appl Toxicol, 2014. 34(1): p. 1-18
4.	Generally, in vitro screens are relevant for effects in both humans and vertebrate wildlife because many of them 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 adverse effects in vivo both in non-mammalian vertebrate species and mammalian species (including wildlife species and humans). A novel paper concludes that when comparing data from fish and rat assays, a high concordance was seen with respect to identifying chemicals that impacted specific endocrine pathways of concern (Ankley and Gray, 2013 see below). Although most chemicals were detected as positive in both the rat and fish assays, eliminating data from one class of vertebrate would weaken the battery. 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 (Ankley and Gray, 2013). Due to the similarities between different vertebrate species it seems to be relevant to consider ED related concerns identified in one species (e.g. in fish or amphibians) as of potential relevance for other species (such as rat or humans) - in particular in cases when no or only limited data is available for more than one species. The current ED Conceptual Framework and Guidance Document do not at the higher levels (from level 3 and onwards) reflect this adequately. Instead the current approach is that "human health concerns" are based on rat /rodent TGs/data only, whereas concern for the environment is based on available data on fish, amphibian/birds and to some extent invertebrates only. In our view, this needs further discussion e.g. of the need for supplementing the current Conceptual Framework or GD 150 with some (more) considerations, e.g. for substances where there are concerns in fish for estrogenicity or for thyroid effects in amphibians and at the same time absence / or clear limitations in availability of relevant data in rodents. In such cases there may be concerns for potential ED related effects in mammalians (e.g. rat and humans) - and vice versa. Another issue is that currently ED related effects in rodents are normally only regarded in the context of their relevance for human health. But, as already mentioned some years ago at an EDTA meeting, such rodent data should in general also be regarded in the context of their relevance for mammalian wildlife species (i.e. in relation to environmental concerns). Reference: Ankley GT, Gray LE. Cross-species conservation of endocrine pathways: A critical analysis of tier 1 fish and rat screening assays with 12 model chemicals. Environ Toxicol Chem. 2013 Feb 7. doi: 10.1002/etc.2151.
5.	Development of methods to detect multifactorial disorders like development of hormone related cancers after in utero exposure, which are challenging and might require another framework to address. More basic longer term research is needed. REFERENCES: Ankley GT, Gray LE. Cross-species conservation of endocrine pathways: A critical analysis of tier 1 fish and rat screening assays with 12 model chemicals. Environ Toxicol Chem. 2013 Feb 7. doi: 10.1002/etc.2151. Hass et al., OECD Conceptual Framework for Testing and Assessment of Endocrine Disrupters as a basis for regulation of substances with endocrine disrupting properties, TemaNord 2004:555, http://www.diva-portal.org/smash/get/diva2:702046/FULLTEXT01.pdf Hass et al., Information/testing strategy for identification of substances with endocrine disrupting properties, Danish Centre on Endocrine Disrupters, 2013, http:// http://www.cend.dk/files/EDtestingstrategy.pdf Greally & Jacobs. In vitro and in vivo Testing Methods of epigenetic Endpoints for evaluating endocrine Disruptors, ALTEX (2014), 7/13.
6.	In my experience the behavior is a very sensitive endpoint that needs to be evaluated always in this type of study.
7.	Good luck!
8.	We are looking forward to a fruitful meeting in November
9.	The priority must be given to those axes for which there is a clear understanding of the long term adverse effects and the AOPs/MOAs responsible for the effects - ie the focus should continue to be on the HPG, HPA, HPT axes. It is currently difficult to say whether the other axes

	are as relevant due to the paucity of mechanistic data. Relevant test methods cannot be developed until AOPs/MOAs are established for a given toxicity and as the scope of each test method consequently developed has been ascertained. Evaluation of EDs cannot rest solely on the use of in vitro data - characterizing the adverse effects in relevant in vivo studies, determining the severity of the effects and whether the observations are reversible as well as establishing that the MOA for the adverse effects is ED related are critical elements not only for identifying EDs but also determining those which are of greatest regulatory concern.
10.	Birds play a critical role in our ecosystem. We currently do not have a validated test that evaluates effects of edcs in birds. Also there is a data gap regarding the MOAs and associated specific endpoints that give rise to effects in birds that would not be seen in mammals and other taxa. I believe this should be an area of focus for endocrine test development that will make a significant contribution in the short and the long term.
11.	Consider that we are not experts for the tests of the OECD conceptual framework, as we do not perform the tests ourselves, our answers should provide more a general view from our site as all-round regulatory toxicologists. Asking the advice of scientists performing regularly the assays could bring high value and strong expertise to the further development of methods for the evaluation of ED.
12.	we should go for a step-wise approach developing/validating further in vitro screening methods, such as reporter assays for other modes of action than so far developed ER & AR assays, but relevant to the whole endocrine axis and hormone signaling pathways (as listed in document). Next to that, based on relevant modes of action detected by in vitro assays, verification should follow by animal test but considering 3R. Not necessary to develop new in vivo tests, but most of current in vivo tests can be upgraded with other apical endpoints (e.g. measurement of hormone levels or biomarkers) representing these other pathways. Upgrading test methods for other pathways, at different levels (according to OECD framework) will allow to anticipate to the AOP concept applicable for EDC evaluation.
13.	UBA scientific report "Substances of very high concern under REACH - an evaluation of uncertainties in the environmental risk assessment of endocrine active substances, Oct. 2012"
14.	References: OECD, New Scoping Document on in vitro and ex vico assays for the 39 identification of medulators of thyroid hormone signalling. 2014, Series on Testing and Assessment. Greally, J.M. and M.N. Jacobs, In vitro and in vivo testing methods of epigenomic endpoints for evaluating endocrine disruptors. ALTEX, 2013. 30(4): p. 445-71. Jacobs, M.N., et al., In vitro metabolism and bioavailability tests for endocrine active substances: what is needed next for regulatory purposes? ALTEX, 2013. 30(3): p. 331-51. OECD, The use of metabolising systems for in vitro testing of endocrine disruptors. A detailed review paper for the OECD No. 97. Paparella, M., et al., Uncertainty of testing methods--what do we (want to) know? ALTEX, 2013. 30(2): p. 131-44. WHO, Harmonisation Project Document No 11. 2014. Edler, L., et al., Selection of appropriate tumour data sets for Benchmark Dose Modelling (BMD) and derivation of a Margin of Exposure (MoE) for substances that are genotoxic and carcinogenic: considerations of biological relevance of tumour type, data quality and uncertainty assessment. Food Chem Toxicol, 2014. 70: p. 264-89 Meek, M.E., et al., New developments in the evolution and application of the WHO/IPCS framework on mode of action/species concordance analysis. J Appl Toxicol, 2014. 34(1): p. 1-18
15.	When pesticides come up for re-authorisation the test data supplied should be as for new pesticides. It is important to enforce the re-testing of such substances with updated endpoints focused on ED as if only tested in old tests, ED properties will be missed. We need to update and find improved test methods, BUT ALSO there is a need to ensure substances at re-authorisation are re-tested with the most up to date test methods.

4 Conclusions

The respondents submitted many valuable and detailed comments. We thank all survey participants for their contributions.

The received information was further evaluated and is currently used to develop a thought-starter document for the workshop on prioritising test method development for the identification of endocrine disrupting substances to be held on 30 May – 1 June 2017 in Brussels.

5 Literature references provided by experts in the survey

In the end of each section the experts provided literature references useful for the prioritising. All references provided are listed here in alphabetical order since the same references were repeatedly mentioned in several sections.

Anand-Ivell R and Ivell R. (2014) Insulin-like factor 3 as a monitor of endocrine disruption. Reproduction 147:R87-95.

Ankley GT, Gray LE. (2013) Cross-species conservation of endocrine pathways: A critical analysis of tier 1 fish and rat screening assays with 12 model chemicals. Environ Toxicol Chem.. doi: 10.1002/etc.2151.

Ashby J, Lefevre PA, Tinwell H, Odum J, Owens W. (2004) Testosterone-stimulated weanlings as an alternative to castrated male rats in the Hershberger anti-androgen assay. Regul Toxicol Pharmacol.39(2):229-38.

Blaauboer BJ, Boekelheide K, Clewell HJ, Daneshian M, Dingemans MML, Goldberg AM, et al. (2012) The use of biomarkers of toxicity for integrating in vitro hazard estimates into risk assessment for humans. ALTEX. 29(4):411-25

DeFur PL, Crane M, Tattersfield LJ (1999). Conclusions and recommendations. In: Endocrine disruption in invertebrates: endocrinology, testing and assessment (DeFur PL, Crane M, Ingersoll C, Tattersfield L, eds.), pp. 271-279. Society of Environmental Toxicology and Chemistry, Pensacola, U.S.A.

EFSA Scientific Committee (2013) Scientific Opinion on the hazard assessment of endocrine disruptors: scientific criteria for identification of endocrine disruptors and appropriateness of existing test methods for assessing effects mediated by these substances on human health and the environment. EFSA Journal 2013;11(3):3132.

Fetter E, Baldauf L, Da Fonte DF, Ortmann J, Scholz S. (2015) Comparative analysis of goitrogenic effects of phenylthiourea and methimazole in zebrafish embryos. Reprod Toxicol. 57:10-20

Greally JM and Jacobs MN (2013) In Vitro and In Vivo Testing Methods of Epigenomic Endpoints for Evaluating Endocrine Disruptors, ALTEX. 30(4):445-71.

Guerra MT, de Toledo FC, Kempinas Wde G (2011) In utero and lactational exposure to fenvalerate disrupts reproductive function in female rats. Reprod Toxicol. 32:298-303

Hass U, Christiansen S, Bjerregaard P, Holbech H (2013) Information/testing strategy for identification of substances with endocrine disrupting properties, Danish Centre on Endocrine Disrupters, http:// http://www.cend.dk/files/EDtestingstrategy.pdf

Head JA, Dolinoy DC, Basu N (2012) Epigenetics for ecotoxicologists. Environmental Toxicology and Chemistry, 31: 221–227.

Hofer D, Münzker J, Schwetz V, Ulbing M, Hutz K, Stiegler P, Zigeuner R, Pieber TR, Müller H, Obermayer-Pietsch B. (2014) Testicular synthesis and vitamin D action. J Clin Endocrinol Metab. 99(10):3766-73.

Hutchinson TH (2002). Reproductive and developmental effects of endocrine disrupters in invertebrates: in vitro and in vivo approaches. Toxicol. Lett. 131, 75-81.

Hutchinson TH (2007). Small is useful in endocrine disrupter assessment – four key recommendations for aquatic invertebrate research. Ecotoxicology 16, 231-238.

IPCS (2002). Global assessment of the state-of-the-science of endocrine disruptors. (Damstra T, Barlow S, Bergman A, Kavlock R, Van der Kraak G, eds.). International Programme on Chemical Safety. WHO/PCS/EDC/02.2, World Health Organisation.

Jacobs MN, Laws SC, Willett K, Schmieder P, Odum J, Bovee TF (2013) In vitro metabolism and bioavailability tests for endocrine active substances: What is needed next for regulatory purposes? ALTEX. 30(3):331-51.

Jonas A, Buranova V, Scholz S, Fetter E, Novakova K, Kohoutek J, Hilscherova K (2014) Retinoid-like activity and teratogenic effects of cyanobacterial exudates. Aquat Toxicol. 155:283-90.

Klüver N, König M, Ortmann J, Massei R, Paschke A, Kühne R, Scholz S (2015) Fish embryo toxicity test: identification of compounds with weak toxicity and analysis of behavioral effects to improve prediction of acute toxicity for neurotoxic compounds. Environ Sci Technol. 49(11):7002-11.

Kortenkamp et al., State of the Art Assessment of Endocrine Disrupters, 2011 http://ec.europa.eu/environment/chemicals/endocrine/pdf/sota_edc_final_report.pdf

Koohi MK, Walther N, Ivell R (2007) A novel molecular assay to discriminate transcriptional effects caused by xenoestrogens. Mol Cell Endocrinol. 276:45-54.

Lafont R (2000). The endocrinology of invertebrates. Ecotoxicology 9, 41-57.

LeBlanc GA (2007). Crustacean endocrine toxicology: a review. Ecotoxicology 16, 61-81.

Leblond CP and Clermont Y (1952) Spermiogenesis of rat, mouse, hamster and guinea pig as revealed by the periodic acid-fuchsin sulfurous acid technique. Am J Anat 1952, 90(2):167-215.

Lee HJ, Macbeth AH, Pagani JH, Young WS (2009) Oxytocin: the great facilitator of life. Prog Neurobiol. 88(2):127-151.

Lengqvist J, Mata De Urquiza A, Bergman AC, Willson TM, Sjövall J, Perlmann T, Griffiths WJ (2004) Polyunsaturated fatty acids including docosahexaenoic and arachidonic acid bind to the retinoid X receptor alpha ligand-binding domain. Mol Cell Proteomics. 3(7):692-703.

Lorenzetti S, Marcoccia D, Narciso L, Mantovani A. (2010) Cell viability and PSA secretion assays in LNCaP cells: a tiered in vitro approach to screen chemicals with a prostate-mediated effect on male reproduction within the ReProTect project. Reprod Toxicol. 30(1):25-35.;

Lowe XR, de Stoppelaar JM, Bishop J, Cassel M, Hoebee B, Moore D and Wyrobek AJ (1998). Environmental and Molecular Mutagenesis 31: 125-132.

Mantovani A and Fucic A (2014) Puberty dysregulation and increased risk of disease in adult life: possible modes of action. Reprod Toxicol. 44:15-22.

Maranghi F and Mantovani A (2012) Targeted toxicological testing to investigate the role of endocrine disrupters in puberty disorders. Reprod Toxicol. 33(3):290-6.

Matthiessen P (2003). Historical perspective on endocrine disruption in wildlife. Pure Appl. Chem. 75, 2197-2206.

Mogi K, Ooyama R, Nagasawa M, Kikusui T (2014) Effects of neonatal oxytocin manipulation on development of social behaviors in mice. Physiol Behav. 133:68-75.

OECD (2006). Detailed review paper on aquatic arthropods in life cycle toxicity tests with an emphasis on developmental, reproductive and endocrine disruptive effects. OECD series on testing and assessment, No. 55. ENV/JM/MONO(2006)22. Organisation for Economic Co-Operation and Development, Paris, France.

OECD (2008): Detailed review paper on the use of metabolising systems for in vitro testing of endocrine disruptors. OECD Series on testing and assessment, No. 97. ENV/JM/MONO(2008)24. Organisation for Economic Co-Operation and Development, Paris, France.

OECD (2010). Detailed review paper on molluscs life-cycle toxicity testing. OECD Series on testing and assessment, No. 121. ENV/JM/MONO(2010)9. Organisation for Economic Co-Operation and Development, Paris, France.

OECD (2012) 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 disruptors" OECD Series on testing and assessment, No. 178. ENV/JM/MONO(2012)23. Organisation for Economic Co-Operation and Development, Paris, France.

OECD (2014) New scoping document on in vitro and ex vivo assays for the identification of modulators of thyroid hormone signaling, Series on Testing and Assessment No. 207, ENV/JM/MONO(2014)23. Organisation for Economic Co-Operation and Development, Paris, France.

Oehlmann J, Schulte-Oehlmann U, Bachmann J, Oetken M, Lutz I, Kloas W, Ternes TA (2006). Bisphenol A induces superfeminization in the ramshorn snail Marisa cornuarietis (Gastropoda, Prosobranchia) at environmentally relevant concentrations. Environ. Health. Persp. 114, 127-133.

Pereira-Fernandes A, Demaegdt H, Vandermeiren K, Hectors TL, Jorens PG, Blust R, Vanparys C (2013) Evaluation of a screening system for obesogenic compounds: screening of endocrine disrupting compounds and evaluation of the PPAR dependency of the effect. PLoS One. 8(10):e77481.

Robb GW, Amann RP, Killian GJ (1978) Daily sperm production and epididymal sperm reserves of pubertal and adult rats. J Reprod Fertil. 54(1):103-107.

Ross CA, Taylor CL, Yaktine AL, Del Valle HB (2011) Dietary Reference Intakes for Calcium and Vitamin D, ISBN 978-0-309-16394-1

Saternus R, Pilz S, Gräber S, Kleber M, März W, Vogt T, Reichrath J (2015) A closer look at evolution: Variants (SNPs) of genes involved in skin pigmentation, including EXOC2, TYR, TYRP1, and DCT, are associated with 25(OH)D serum concentration. Endocrinology. 156(1):39-47.

Schiffer C, Müller A, Egeberg DL, Alvarez L, Brenker C, Rehfeld A, Frederiksen H, Wäschle B, Kaupp UB, Balbach M, Wachten D, Skakkebaek NE, Almstrup K, Strünker T (2014) Direct action of endocrine disrupting chemicals on human sperm. EMBO Rep. 5(7):758-65.

Schiffelers MJ, Blaauboer BJ, Bakker WE, Hendriksen CF, Krul C (2015) Regulatory acceptance and use of the Extended One Generation Reproductive Toxicity Study within Europe. Regul Toxicol Pharmacol. 71(1):114-24.

Smeriglio A, Trombetta D, Marcoccia D, Mantovani A, Lorenzetti S (2014) Intracellular distribution and biological effects of plant bioactives in a sex steroid-sensitive model of human prostate adenocarcinoma. Anticancer Agents in Medicinal Chemistry, 14:1386-96.

Soin T, Smagghe G (2007). Endocrine disruption in aquatic insects: a review. Ecotoxicology 16, 83–93.

Thienpont B, Tingaud-Sequeira A, Prats E, Barata C, Babin PJ, Raldúa D (2011) Zebrafish eleutheroembryos provide a suitable vertebrate model for screening chemicals that impair thyroid hormone synthesis. Environ Sci Technol. 45(17):7525-32.

Tinwell H, Friry-Santini C, Rouquié D, Belluco S, Elies L, Pallen C, Bars R (2007) Evaluation of the antiandrogenic effects of flutamide, DDE, and linuron in the weanling rat assay using organ weight, histopathological, and proteomic approaches. Toxicol Sci. 100:54-65.

Trummer O, Schwetz V, Walter-Finell D, Lerchbaum E, Renner W, Gugatschka M, Dobnig H, Pieber TR, Obermayer-Pietsch B (2012) Allelic determinants of vitamin d

insufficiency, bone mineral density, and bone fractures. J Clin Endocrinol Metab. 97(7):E1234-40.

UBA (2012) Substances of very high concern under REACH – an evaluation of uncertainties in the environmental risk assessment of endocrine active substances. Project No. 3710 63 416.

Vandegehuchte MB, Janssen CR (2011) Epigenetics and its implications for ecotoxicology. Ecotoxicology. 20(3):607-24.

Vidal JD, Mirsky ML, Colman K, Whitney KM, Creasy DM (2013). Reproductive system and mammary gland. Chapter 18 in Toxicologic pathology –Nonclinical safety assessment, (edited by Sahota PS, Popp JA, Hardisty JF, Gopinath C), CRC Press, Taylor and Francis Group.

Weibel ER (1963) Principles and methods for the morphometric study of the lung and other organs. Lab Invest 1963, 12(1):131-155.

Yeh J, Kim B, Liang YJ, Peresie J (2006) Mullerian inhibiting substance as a novel biomarker of cisplatin induced ovarian damage. Biochem Biophys Res Commun 348: 337-344

Yeh J, Kim B, Peresie J, Liang YJ, Arroyo A (2007) Serum and ovarian Mullerian inhibiting substance, and their decline in reproductive aging. Fertil Steril 87: 1227-1230 98.

6 References used in designing the survey

WHO/UNEP (2012) State of the Science of Endocrine Disrupting Chemicals – 2012. Edited by Bergman A, Heindel JJ, Jobling S, Kidd KA, Zoeller RT.

OECD (2012) 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 disruptors" OECD Series on testing and assessment, No. 178. ENV/JM/MONO(2012)23. Organisation for Economic Co-Operation and Development, Paris, France.

OECD (2014) New scoping document on in vitro and ex vivo assays for the identification of modulators of thyroid hormone signaling, Series on Testing and Assessment No. 207, ENV/JM/MONO(2014)23. Organisation for Economic Co-Operation and Development, Paris, France.

Abbreviations

ACTH	Corticotropin
ADHD	Attention deficit hyperactivity disorder
AFSS	Androgenised female stickleback screen
AhR	Aryl hydrocarbon receptor
AMA	Amphibian metamorphosis assay
AOP	Adverse Outcome Pathway
ATGT	Avian two-generation test
CAR	Constitutive androstane receptor
CF	Conceptual Framework
EASZY	Detection of endocrine active substances, acting through estrogen receptors, using transgenic cyp19a1b-GFP zebrafish embryos
EFSA	European Food Safety Authority
EIA	Enzyme immunoassay
EMSA	Electrophoretic mobility shift assay
EP	Endpoint
Eq	Equation
ER	Estrogen receptor
EROD	Ethoxyresorufin-O-deethylase
ESTAF	EURL ECVAM Stakeholder Forum
EURL ECVAM	European Union Reference Laboratory for alternatives to animal testing
FLCTT	Fish life cycle toxicity test
FRPLT	Fish reproduction partial lifecycle test
FSDT	Fish sexual development test
FSTRA	Fish short term reproduction assay
GR	Glucocorticoid receptor
HPA	Hypothalamus-pituitary-gonad axis
HPG	Hypothalamus-pituitary-adrenocortical axis
HPT	Hypothalamus-pituitary-thyroid axis

IATA	Integrated approaches to testing and assessment
IGF-1	Insulin-like growth factor 1
LAGDA	Larval amphibian growth and development assay
MEOGRT	Medaka extended one-generation reproduction test
OECD	Organisation for Economic Co-operation and Development
PARERE	EURL ECVAM's Network for Preliminary Assessment of Regulatory Relevance
PPAR	peroxisome proliferator-activated receptor
PR	progesterone receptor
RAR	retinoic acid receptor
RIA	radioimmunoassay
RXR	retinoid X receptor
TG	test guideline
TH	tyrosine hydroxylase
TM	Test method
TR	thyroid hormone receptor
WHO/IPCS	World Health Organisation/ International Programme on Chemical Safety
XETA	Xenopus embryonic thyroid signalling assay

List of boxes

List of figures

List of tables

Annexes

Annex 1: Overview of the expert relevance rating of test methods included in the OECD CF and in OECD DRP No 178 (2012).

Source or Pathway	Endpoint (EP) or Test Method (TM)	**Relevance Ranking** based on Ratio relevant/not relevant **[Eq5, Eq7]**	**% practical problems [Eq6]**[a]
CF	Amphibian metamorphosis assay (AMA, TG 231)	- [b]	
CF	Fish sexual development test (FSDT, TG 234)	- [b]	
CF	Medaka extended one-generation reproduction test (MEOGRT)	- [b]	
CF	Extended one-generation reproductive toxicity study (TG 443)	29.0	
HPA	(EP) adrenal steroid synthesis (in vivo)	24.0	20.8
HPG gestagenic	(EP) reduced fertility in fish (TG229; MEOGRT)	22.0	18.2
PPAR	(TM) Adipocyte differentiation in cultured pre-adipocyte cells	21.0	19.0
CF	Larval amphibian growth and development assay (LAGDA)	20.0	
CF	Xenopus embryonic thyroid signalling assay (XETA)	20.0	
CF	Fish short term reproduction assay (FSTRA, TG 229)	17.0	
HPT	(TM) neurite extension assay	14.0	28.6
HPT	(TM) neural progenitor cell proliferation assay	14.0	28.6
HPA	(EP) Stress response (in vivo)	12.5	44.0
retinoid	(TM) RXR transactivation assay	12.5	12.0
HPG gestagenic	(TM) Progesterone receptor (PR) transactivation test	11.5	17.4
HPG estrogenic	(EP) Gonad histopathology in chronically exposed amphibians (TG 231 AMA; included in LAGDA)	10.5	14.3
HPT	(TM) XETA (Xenopus Embryonic Thyroid Signaling Assay)	9.5	10.5
CF	Developmental neurotoxicity (TG 426)	8.7	
HPT	(TM) thyroid peroxidase assay	7.7	8.7
HPT	(TM) iodine uptake assays	7.7	8.7
CF	Pubertal development and thyroid function assay in intact juvenile/peripubertal male rats (OCSPP 890.1500)	7.7	
HPT	(TM) Tadpole tail explant resorption assay	7.5	26.7
CF	Detection of endocrine active substances, acting through estrogen receptors, using transgenic cyp19a1b-GFP zebrafish embryos (EASZY)	7.5	

CF	Pubertal development and thyroid function assay in intact juvenile/peripubertal female rats (OCSPP 890.1450)	7.3	
retinoid	(TM) RAR transactivation assay	7.0	14.3
PPAR	(TM) PPARalpha,beta/delta,gamma transactivation assay	7.0	9.5
retinoid	(EP) weight gain, increased adipose tissue mass, increased lipid accumulation, reduced retinoid levels in vivo (TG 415, 416, 443, fish and amphibians)	6.8	14.8
HPT	(TM) dendritic arborization assay	6.0	41.7
HPT	(EP) TH production in thyroid gland explants	6.0	33.3
HPA	(EP) corticotropin (ACTH) release (in vivo)	5.8	21.7
HPT	(TM) T4 binding protein displacement assay	5.8	13.0
HPG gestagenic	(TM/EP) assessment in exposed oocytes and sperm ex vivo or in oocytes/sperm derived from adult fish from TG 229 and MEOGRT	5.7	29.4
HPT	(TM) T-screen assay	5.7	17.6
CF	Androgenised female stickleback screen (AFSS, GD 140)	5.3	
HPA	(TM) adrenal steroid synthesis (in vitro), e.g. modified TG 456	5.3	19.0
Vitamin D	(TM) VDR transactivation assay	5.0	0.0
HPG gestagenic	(TM) membrane PR binding assay	4.8	21.1
HPG estrogenic	(EP) GnRH neuron development in brain of chronically exposed fish (MEOGRT)	4.7	14.3
CF	21-day fish assay (TG 230)	4.7	
CF	Two-generation reproduction toxicity study (TG 416)	4.6	
CF	Prenatal developmental toxicity study (TG 414)	4.4	
Vitamin D	(TM) Vitamin D hydroxylase assay (in vivo)	4.3	15.4
PPAR	(EP) Weight gain in chronically exposed animals (TG 415, 416, 443, LAGDA)	4.2	4.8
HPG androgenic	(EP) behavioural assessments in any in vivo study	4.2	40.0
retinoid	(EP) EROD induction in in vivo assays	4.0	5.0
Vitamin D	(EP) RIA or EIA for serum vitamin D levels (could potentially be applied to any in vivo exposure assay)	3.8	13.3
Vitamin D	(EP) reduced bone length in juvenile rodent (TG 416, 443)	3.2	31.3
HPA	(TM) GR transactivation test (in vitro)	3.2	21.1
retinoid	(TM) AhR transactivation assay	3.2	15.8
HPT	(TM) TR reporter assays	3.2	15.8
CF	Fish reproduction partial lifecycle test (no validation	3.0	

	ongoing)		
retinoid	(EP) CYP1A mRNA or protein quantification in in vivo assays	2.9	10.0
PPAR	(TM) Peroxisome proliferation assay	2.8	23.5
CF	Uterotrophic bioassay in rodents (TG 440)	2.8	
CF	Avian reproduction test (TG 206)	2.8	
CF	Avian two-generation test (ATGT)	2.8	
CF	Combined 28 day reproductive screening test (TG 421+422)	2.7	
HPT	(TM) AhR reporter assays	2.7	18.8
CF	Hershberger bioassay in rats (TG 441)	2.6	
HPT	(TM) CAR reporter assays	2.3	14.3
CF	Combined chronic toxicity / carcinogenicity study (TG 453)	2.0	
Vitamin D	(EP) Brain size measurements in rodent offspring	1.9	46.2
CF	One-generation reproduction toxicity study (TG 415)	1.8	
somatotropic	(EP) Fetal birth weight and length in rodent multigeneration tests (TG 416, 443)	1.7	4.0
CF	Fish life cycle toxicity test (FLCTT) (no validation ongoing)	1.6	
CF	Repeated dose 28 day study (TG 407)	1.6	
Vitamin D	(TM) AhR transactivation assay	1.6	0.0
CF	Repeated dose 90 day study (TG 408)	1.4	
Vitamin D	(EP) EROD activity assay (biomarker, could potentially be applied to any in vivo exposure assay)	1.3	37.5
somatotropic	(EP) analyses of hepatic GR mRNA levels in fish/mammals in vivo assays	1.1	23.8
somatotropic	(EP) Analyses of hepatic IGF-1 mRNA levels in fish/mammal in vivo assays	1.1	14.3
somatotropic	(TM) TR and GR transactivation assays	1.0	25.0
somatotropic	(EP) Growth evaluation in fish assays (MEOGRT)	1.0	20.0
HPT	(TM) EMSA, DNA pull-down assay	1.0	50.0

[a] = Experts were asked to address practical problems only for the tests included in the OECD DRP no 178 (2012). Therefore, for OECD CF included tests no value is presented.

[b] = there were no answers stating low or no relevance, therefore no ratio number is given, but tests should be considered of high relevance.

Europe Direct is a service to help you find answers to your questions about the European Union.

Freephone number (*)**:**

00 800 6 7 8 9 10 11

(*) The information given is free, as are most calls (though some operators, phone boxes or hotels may charge you).

More information on the European Union is available on the internet (http://europa.eu).

HOW TO OBTAIN EU PUBLICATIONS

Free publications:

- one copy:
  via EU Bookshop (http://bookshop.europa.eu);
- more than one copy or posters/maps:
  from the European Union's representations (http://ec.europa.eu/represent_en.htm);
  from the delegations in non-EU countries (http://eeas.europa.eu/delegations/index_en.htm);
  by contacting the Europe Direct service (http://europa.eu/europedirect/index_en.htm) or
  calling 00 800 6 7 8 9 10 11 (freephone number from anywhere in the EU) (*).

  (*) The information given is free, as are most calls (though some operators, phone boxes or hotels may charge you).

Priced publications:

- via EU Bookshop (http://bookshop.europa.eu).

KJ-NA-28592-EN-N

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

EU Science Hub
ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub - Joint Research Centre

Joint Research Centre

EU Science Hub

Publications Office

doi:10.2760/844100

ISBN 978-92-79-68778-5