Report on
Interpretation of knowledge on endocrine disrupting substances (EDs) – what is the risk?

DANISH CENTRE ON ENDOCRINE DISRUPTERS

Ulla Hass and Sofie Christiansen
Division for Diet, Disease Prevention and Toxicology, National Food Institute,
Technical University of Denmark (DTU)

Anna-Maria Andersson
Department of Growth and Reproduction, Rigshospitalet (GR)

Henrik Holbech and Poul Bjerregaard
Department of Biology, University of Southern Denmark (SDU)

Table of Contents

1. Terms of reference and scope

This report has been prepared by the Danish Centre on Endocrine Disrupters (CeHoS) as a project contracted by the Danish Environmental Protection Agency. The Danish Centre on Endocrine Disrupters is an interdisciplinary scientific network without walls. The main purpose of the Centre is to build and gather new knowledge on endocrine disrupters (EDs) with the focus on providing information requested for the preventive work of the regulatory authorities. The Centre is financed by the Ministry of Environment and Food of Denmark and the scientific work programme is followed by an international scientific advisory board.

The overall scope of this project was to provide a science based input to the ongoing work in EU with regards to specifically risk assessment of endocrine disruptors.

The project has been carried out by a project team: Ulla Hass (project leader, DTU) and Sofie Christiansen (DTU), Anna-Maria Andersson and Katrine Bay (GR), Poul Bjerregaard and Henrik Holbech (SDU).

This report describes the status as of autumn 2017. The final report was handed in to The Danish Environmental Protection Agency September 2017. Just before publication (in February 2019), some mainly editorial chances were made.

2. Background and aim

In a previous CeHoS report, it was concluded that there are major limitations as to the ability of the current testing requirements to adequately screen for endocrine disrupting properties and that effect sizes of human relevance may be present at the NOAEL (Hass et al. 2013). Several other uncertainties exist when it comes to assessment of the risk due to exposure to EDs. In addition to general knowledge gaps related to EDs there is also ongoing scientific discussions about in particular: the existence of a threshold for ED-effects, the existence of non-monotonic dose responses (NMDRs) for EDs, the existence of low dose effects of EDs, all issues relating to whether the current approach for risk assessment of substances regarded to have a threshold is appropriate to apply also for risk assessment of endocrine disruptors. Despite all these uncertainties there is at the same time an ongoing assessment of substances under various regulations, e.g. REACH. Criteria for the identification of endocrine disrupting substances under the biocides and pesticide regulations entered into force in June 2018 (biocides) and November 2018 (pesticides) and the ECHA/EFSA Guidance for the identification of endocrine disruptors was accepted in 2018[1]. However, under REACH there is a common agreement to identify endocrine disruptors based on the WHO/IPCS-definition from 2002[2] and the recommendations from the EU Commission Endocrine Disruptors Expert Advisory Group (JRC, 2013), but it is still under discussion how the risk of exposure to

[1] After completion of this report.
[2] The identification as ED in REACH also requires identification as SVHC cf. Art 57(f) and ELoC as CMR substances.

endocrine disruptors should be assessed. The Danish EPA has therefore requested scientific advice and guidance on how to handle the issue of risk assessment of endocrine disruptors.

The project was a so-called scoping project, i.e. a project that gathers and discusses information. The aims were - based on the current knowledge and the CeHoS (Centre on Endocrine Disrupters) expertise on epidemiology, toxicology and ecotoxicology - to:

- develop specific recommendations with regards to toxicology studies and derivation of tolerable exposure levels for endocrine disrupters (EDs)
- discuss considerations in regard to uncertainties in risk assessment of EDs, such as the presence or lack of threshold for effect
- discuss the draft report for the project at an international workshop
- develop clear recommendations for tolerable exposure levels for humans to be used for risk assessment of EDs
- make a final report that is useful for further discussions within EU at e.g. stakeholder meeting and within the scientific community

According to these aims, the scoping project considers EDs. The focus is mainly on effects on reproduction and sexual development in relation to risk for humans. As criteria for defining a substance as an ED has only recently[3] entered into force in the European Commission under the biocides and pesticide regulations, the project team has assumed that EDs means substances evaluated as EDs based on the WHO/IPCS definition of an endocrine disruptor (WHO/IPCS, 2002) and the recommendations from the EU Commission Endocrine Disruptors Expert Advisory Group (JRC, 2013), which will cover substances belonging to group 1 according to the proposed Danish Criteria for EDs (Hass et al. 2011). Suspected EDs are not considered in this scoping project. Moreover, combined exposure scenarios are not considered in this scoping project.

The process for the project was structured in 3 parts with the following main work tasks:

- Part 1:
  a) Initial work and preparation of material for internal workshops
  b) Two internal workshops with experts from the Danish Centre on Endocrine Disrupters and Danish authorities in October 2014 and February 2015
  c) A final discussion paper including the presentations and discussions from the two internal workshops.
- Part 2:
  a) Further work and preparation of draft workshop material incl. the draft discussion report paper for circulation before workshop
  b) A one-day workshop in November 2015, where the project team together with invited experts from Sweden, the Netherlands and Great Britain and from the Danish EPA discussed uncertainties and possible approaches for risk assessment of endocrine disruptors
  c) A final discussion report covering all aspects of the project so far

[3] In 2018 after completion of this report

- Part 3:
  a) Organization, preparation and conduction of an international workshop with 40-50 invited risk assessors from European authorities and experts with research background in endocrine disruption in May 2017
  b) Workshop Report
  c) Update of the discussion report with focus on an additional case study and the outcome of the international workshop

3. Risk assessment of EDs based on toxicological data - uncertainties

The topic for this section is the derivation of 'tolerable' dose levels for EDs based on toxicology studies, i.e. the potential need for additional assessment factors (other than the default factors for inter- and intraspecies differences). It is assumed that the toxicological data available are of sufficient quality and that the substance is evaluated as being an ED based on these data. Evaluation of data quality is beyond the scope of this project.

Depending on the regulatory context the terms Derived No effect Level (DNEL), Derived Minimal Effect Level (DMEL), Acceptable Daily Intake (ADI) and Tolerable Daily Intake (TDI) are used for the dose considered as tolerable (sufficiently safe) for humans. This dose is derived by applying assessment factors related to the extrapolation from experimental animals to humans and is used for comparison with human exposure levels to evaluate the risk for humans. However, there are a number of uncertainties related to the assessment of EDs and derivation of ´tolerable´ dose levels for EDs is therefore a challenge.
The uncertainties related to assessment of EDs include, but may not be limited to:

1. Thresholds or non-threshold assumption for ED effects?
2. Exposure during sensitive windows
3. Limited sensitivity of (many) regulatory testing methods
4. Non-monotonic dose-response (NMDR)

These uncertainties related to the assessment of EDs have been discussed in e.g. the previous CeHoS-report "Input for the REACH-review in 2013 on endocrine disrupters" (Hass et al. 2013). The following is mainly extracted from section 7 "Summary, conclusions and recommendations" in this report. However, some updates especially with regards to regulatory testing methods (OECD Test Guidelines) have been made.

3.1 Thresholds or non-threshold assumption for ED effects?

The presence of thresholds for effects can never be confirmed or rejected by experimental data, because all methods for measuring effects have a limit of detection below which effects cannot be observed. Thus evaluations on whether effects of EDs should be assumed to exhibit a threshold or not have to be based on a combination of biological plausibility and experimental observations. A general argument for assuming no biological threshold for EDCs is that low doses of endogenous hormones are present and fluctuating and therefore, small additions (or subtractions) to their actions will have a significant impact. The validity of assuming no biological threshold for EDs is supported by the very important organizing role hormones have during development at a time point where the homeostatic control is not effective or not yet developed. In addition, experimental data for e.g. procymidone and DEHP indicate non-thresholded dose-response relationships for some endpoints for adverse effects on sexual differentiation such as anogenital distance and nipple retention at the dose levels studied so far. It is therefore concluded - based on a combination of biological plausibility and experimental observations - that an assumption of no threshold for effects appears more plausible for the effects of EDs during development than an assumption of existence of a threshold for effects (Hass et al. 2013).

Regardless of the ED mode of action, it is uncertain whether or not there is a threshold for effects of EDs. For EDs, where the MoA (Mode of Action) directly involves the receptor, the interaction with the receptor is likely to be without having a threshold. Irrespective of the existence of threshold or non-threshold for the response, the dose response curves for EDs not displaying non-monotonic dose response relationship seem generally to be best described as sigmoid curves, i.e. the effect decreases asymptotically with dose towards zero but does not become zero, as supported by several types of experimental data. Such curves have a “threshold-like” appearance, but a threshold cannot be inferred from the shape of the dose-response curves. However, for risk assessment purposes a benchmark approach may be used for estimating a tolerable human exposure level.

For the environment, the text above is also valid for individual animals of different wildlife species but it should be taken into account that it is generally the population level effects that are of concern for the environmental hazard- and risk assessment.

3.2 Exposure during sensitive windows

The majority of the effects potentially related to human exposure to EDs during *in utero* and early postnatal development become manifest later in life, e.g. behavioural effects in children or adults, alterations of puberty timing, low sperm quality, decreased fertility, increased risk for cancer in mammary tissue, prostate and testes, endometriosis and effects on menopause in women. This reflects that exposure during early development can lead to irreversible developmental programming affecting the health for the rest of the individual’s lifetime and possibly also future generations. Thus, there may be a time lag of many years or several decades from the time when regulatory decisions on risk reduction are taken, to the time when this risk reduction will be

achieved. This is of particular concern when the regulation aims for reduction of risks to chemicals causing severe and delayed effects of EDs.

As regards effects on reproduction comparison of LOAELs for ED effects after exposure during sensitive windows of development to exposure outside these windows (e.g. in adults) have been done for a limited number of EDs by Hass et al. (2004). An additional safety factor of 10 seemed in most cases to be appropriate to compensate for lack of information from studies during sensitive windows of development. However, there were some exceptions such as for the 5-alpha reductase inhibitor, finasteride, and the steroid synthesis inhibitor, DEHP, where an additional safety factor needed to be higher than 10 to provide sufficient protection until further results are obtained in the reproductive and developmental toxicity studies.

The above comparison was done before the OECD TG 407 was updated in 2008 with the aim to include endpoints in adult animals that can detect effects of substances with presumed (anti)oestrogenic, (anti)androgenic, and thyroid disrupting mode of action. The validation of the updated OECD TG 407 based on studies of 10 substances showed that this assay is relatively insensitive and would only detect chemicals that are moderate and strong EDs as regards (anti)-oestrogenicity and (anti)-androgenicity (e.g. ethinylestradiol and flutamide) (OECD, 2018a, GD 150[4]). However, the assay did detect EDs that were weak and strong modulators of thyroid hormone-related effects (e.g. propylthiouracil and methyl testosterone). It may also detect steroidogenesis inhibition although only one (potent) chemical was used in the validation study.

Based on the above, it should still be considered to use the NOAELs obtained in the *in vivo* tests in adults, until further results are obtained in the reproductive and developmental toxicity studies, however inclusion of an additional factor of 10 is proposed to compensate for the lower sensitivity in adults. Absence of ED related effects in adults should not be used to conclude lack of effects during sensitive periods of development due to the lower sensitivity in adults.

For the environment, exposure during sensitive life-stages as for example development and sexual differentiation can be of crucial importance for detection of ED effects. As described for human health, effects of exposure during early life-stages often become manifested in adulthood. One example is phenotypic sex changes in fish after developmental exposure to (anti-)oestrogens, (anti)androgens or steroidogenesis inhibitors (Holbech et al., 2006).

[4] This GD 150 has been updated in 2018 (after completion of the report) and the reference list is updated with a new link to GD 150: Revised Guidance Document 150 on Standardised Test Guidelines for Evaluating Chemicals for Endocrine Disruption, OECD Series on Testing and Assessment, No. 150, OECD Publishing, Paris, https://doi.org/10.1787/9789264304741-en

3.3 Limited sensitivity of (many) regulatory test methods

3.3.1 Identification of EDs

The current standard information requirements in REACH are not designed for the identification of EDs, although certain endpoints and assays may provide information on adverse effects potentially related to endocrine disruption or information on endocrine mode of action. It is, however, evident that important endpoints needed for the detection of ED effects have not been included in many of the available studies using regulatory test methods. Especially, important effects (e.g. sexual maturation, sperm quality, oestrus cyclicity, follicular count, mammary gland development) after exposure that cover windows of susceptibility during development are not investigated at tonnage levels below 1000 tons per year. This raises major uncertainty as to the ability to adequately screen for endocrine disrupting properties of substances at tonnage levels below 1000 tons per year.

Until recently, a two-generation reproduction toxicity study was generally required for chemicals with a supply tonnage level above 1000 tons per year and 10-1000 tons if triggered. In the new standard information requirements, this study has been replaced with the extended one-generation reproductive toxicity study (EOGRTS), but for the majority of compounds tested for reproductive toxicity, test results from the EOGRTS study are not available. Both types of studies include exposure during sensitive windows of development and assessment of a number of endpoints sensitive to endocrine disruption in the offspring. However, as regards the two-generation study some endocrine sensitive endpoints were added only in 2001 as a result of an update of the guideline and others are not included in the updated version of the two-generation reproduction study, such as nipple retention, anogenital distance at birth (F1), and measurement of thyroid hormones (OECD, 2001). Thus, for the two-generation reproduction toxicity study there are uncertainties with regard to the ability to adequately detect EDs.

The extended one-generation reproduction toxicity study (OECD TG 443) includes the above-mentioned ED sensitive endpoints (OECD, 2018b[5]). The exposure of the foetus (which is a sensitive life-stage for endocrine disruptive effects), the long duration of dosing and the diversity of endpoints means that the extended one-generation study may be considered to be the best predictive test for ED-mediated adverse effects via oestrogen/ androgen/ thyroid/ steroidogenesis – EATS - modalities (OECD, 2018a). Therefore, using the extended one-generation study instead of the two-generation study will significantly enhance the ability for detection of endocrine disrupting substances at tonnage levels above 1000 tons per year.

[5] This TG 443 has been updated (with one sentence) in 2018 (after completion of the report) and the reference list is updated with a new link to TG 443 Test No. 443: Extended One-Generation Reproductive Toxicity Study, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, Paris, https://doi.org/10.1787/9789264185371-en. https://www.oecd-ilibrary.org/environment/test-no-443-extended-one-generation-reproductive-toxicity-study_9789264185371-en

Very recently, the reproduction/developmental toxicity screening studies (OECD TG 421/422) have been updated with inclusion of some sensitive endpoints for endocrine disruption (AGD, nipple retention and thyroid effects) (OECD, 2016a, OECD, 2016b). These endpoints were included based on a feasibility study addressing scientific and technical questions related to their inclusion as well as possible adaptations of the test design needed for their inclusion (OECD, 2015). These test methods are part of the REACH standard information requirements for substances at tonnage levels above 10 tons per year. The future use of the updated TGs will significantly enhance the ability for detection of indication of endocrine disrupting effects at tonnage levels above 10 tons per year.

For the environment, the development of ED test guidelines has been focussing on a few species for which there have been good laboratory experience. These model species like zebrafish (*Danio rerio*), African clawed frog (*Xenopus laevis*) and Japanese quail (*Coturnix japonica*) might not be representative for their respective phyla. For other potentially sensitive phyla, there is a total lack ED-sensitive standardised tests. For example, no tests with reptiles have yet been developed and tests on molluscs do not include ED-specific endpoints (Coady et al., 2017). Certain effects caused by EDs such as delayed or multigenerational effects underline the importance of testing in sensitive life stages or in full life cycle tests (Parrott et al., 2017). It should also be noticed that none of the OECD test guidelines developed to detect ED effects in wildlife species have been included in the standard information requirements in REACH, PPPR or BPR regulations to date.

3.3.2 Regulatory test methods, sensitivity for detection of EDs

The sensitivity for detecting an effect in a study can be determined by power calculations. The number of animals per group is important for the sensitivity. In OECD TG 416 and OECD TG 443 (two- and extended one generation study) 20 litters per group are included whereas the OECD TG 421/422 (Reproductive Screening Study) includes 8 litters per group (OECD, 2001, OECD 2018b, OECD 2016a, OECD 2016b).

Power calculations for studies using 8 or 20 litters per group indicate that the magnitude of effect needed for detection of statistically significant effects on quantal endpoints such as malformations or infertility have to be 25-37% with 20 litters per group and 50-75% with 8 litters per group (see Hass et al. 2013 for details). This clearly shows that the sensitivity for detecting quantal effects is very low with 8 litters per group and also quite low with 20 litters per group. Thus, such effects may go undetected, i.e. be present at the NOAEL, irrespective of the number of litters per group.

For continuous data (e.g. AGD), power calculations show that in studies with 20 litters per group, an effects around 4-7% will in most cases be statistically significant, whereas the magnitude of effect has to be 7-11% if only 8 litters per group are studied. Thus, such effects may be present at the NOAEL especially if only 8 litters per group are studied.

For the environment, the statistical power of the endpoints in the OECD tests is more related to the practicality and size of the experimental designs than to specific knowledge of the ecological relevance of certain changes in a given effect. For example, comprehensive work was done on the power of changes of phenotypic sex ratio in OECD TG 234, the fish sexual development test (FSDT) (OECD, 2012), where the final test design resulted in sufficient power (~80%) to detect phenotypic sex changes of approximately 30%. However, the sizes of these changes were not related to population level effects of wildlife fish species. I.e. what effects on the population will a change from 50% to 30% females in a specific fish species have? It is therefore possible that population level effects on sex ratio are present at or below NOEC.

3.4 Non-monotonic dose-response (NMDR)

Several mechanisms illustrate how hormones and EDs may cause NMDRs due to the function of the endocrine system. These mechanisms include receptor selectivity, receptor down-regulation and desensitization, receptor competition, and endocrine negative feedback loops.

The existence of NMDRs for endocrine active drugs has been recognized and used in human clinical practice for many years (reviewed by Vandenberg et al. 2012, Juul A et al. pers. com). A different specific term, i.e. flare, may be used. Flare is often reported in the therapy of hormone-dependent cancers such as breast and prostate cancer. Tamoxifen flare was described and named as a transient worsening of the symptoms of advanced breast cancer seen shortly after the initiation of therapy in some patients. If the therapy could be continued, the patients showing tamoxifen flare demonstrated a very high likelihood of subsequent response to tamoxifen, including arrest of tumour growth and progression of symptoms for some time. In addition, NMDR has been shown for many different ED-mediated *in vitro* and *in vivo* effects including binding to steroid hormone receptors and adverse effects on reproductive organ weights (prostate and testis), nipple retention and sexual maturation. In many of the cases, the observed NMDR is likely to directly reflect the way the endocrine system works. In other cases, the NMDR may reflect that the substance has multiple ED modes of action operating simultaneously, but with different dose-response curves. As detailed mechanistic knowledge is limited for most EDs it is often difficult to evaluate the MoA behind NMDR.

For the environment, NMDR has also been observed. For example with the aromatisable androgen methyl testosterone, where fish sex ratio changes from all male at medium exposure concentrations to mixed sex ratio at higher concentrations (Örn et al., 2003).

4. Approaches for derivation of tolerable exposure levels for EDs

The current EU legislation on industrial chemicals, REACH, already lays down that for substances for which it is not possible to determine a threshold for the adverse effects, the adequate control route for authorisation is not possible. It is up to applicants for authorisation to demonstrate that a threshold for the effects exists and to determine that threshold. Even though this might be

particularly difficult for EDs, it cannot be excluded based on current knowledge, that it will be possible (ED REACH Review COM 2016, 814 COM (2016))[6].

Based on the above position with regards to threshold or not for EDs it appears relevant to consider both the DNEL and DMEL (derived minimal effect level) approach for EDs, because at present a threshold has not been "demonstrated" for EDs in general. These two situations are dealt with separately in the next two subsections.

4.1 A threshold has been "demonstrated", i.e. a $DNEL_{ED}$ can be derived

If a threshold has been "demonstrated" a DNEL can be derived. A benchmark dose (BMD) approach where both effect size and severity is included is recommended as an option by Hass et al. (2013) and could be used when estimating human risk instead of a NOAEL approach. By using the BMD approach, the power of the data is taken into consideration. Thus poor data quality will lead to a lower benchmark dose low (BMDL), whereas better data, with their reduced degree of uncertainty, leads to a higher BMDL. In contrast, poor data quality usually results in higher NOAELs. However, in many cases derivation of a robust BMDL is not possible (e.g. due to the use of only 3 dose levels in regulatory test methods) and the DNEL is therefore often based on a NOAEL. This may in some cases be less precautionary, and it is recommended to consider whether this leads to the need for an additional assessment factor.

At the international workshops (details in Appendix 1 and 2) the most important uncertainties as regards assessment of EDs were - apart from the existence of a threshold - evaluated to be

- Exposure during sensitive windows
- Limited sensitivity of (many) testing methods with regards to relevant endpoints addressed and power/ robustness of study
- The potential for Non-monotonic dose-response (NMDR)

A large majority of the participants at the workshops found that there is a need to address these uncertainties related to assessment of EDs when deriving a "tolerable/safe" reference level for human risk assessment.

Table 1 shows the additional assessment factors discussed at workshop 1 to be used when deriving such a reference level, called $DNEL_{ED}$. Some of the additional assessment factors may also be relevant for some types of effects not mediated via the endocrine system, but considerations of this is outside the scope of this work. However, for ED effects the identified uncertainties are evaluated to call for additional assessment factors. So far, there is limited quantitative data to determine the appropriate magnitude of several of these factors. The suggested approach therefore also calls for further studies and evaluations that may lead to modifications of the proposed factors.

[6] https://ec.europa.eu/transparency/regdoc/rep/1/2016/EN/COM-2016-814-F1-EN-MAIN-PART-1.PDF

Table 1 Additional assessment factor to address uncertainties related to assessment of EDs when deriving $DNEL_{ED}$

Uncertainty	Proposed factor	Comments
Severity of effect (mainly categorical data)	**2**: if high severity, e.g. malformations, dystocia, decreased fertility	
Power of study for detecting ED effect	**3**: if the power is low, i.e. if only effect sizes larger than 10% can be detected	
Exposure during sensitive window?	**10**: if exposure is not during sensitive windows	Based on limited comparisons of dose levels causing ED related effects after adult exposure and exposure during sensitive windows
ED MOA, NMDR	**10**: data-driven additional concern in cases where the mode of action of the compound indicates that lower doses than the starting point used for deriving the $DNEL_{ED}$ may induce NMDR	Currently, there is very limited knowledge on the importance and magnitude of this uncertainty. The magnitude of the factor might be estimated based on known examples of NMDR seen at lower doses than regulatory NOAELs
Using NOAEL instead of BMDL	Depending on the data	Data-driven evaluation of the need for an additional factor to address cases where a NOAEL approach is less cautious than a BMDL approach
Using LOAEL or $BMDL_{10}$ as starting point	**10**	Because $BMDL_{10}$ is a LOAEL-like value, a similar factor as used to extrapolate from LOAEL to NOAEL is proposed

4.2 A threshold has not been ”demonstrated” i.e. a $DMEL_{ED}$ can be derived

If a threshold has not been ”demonstrated” a way forward could be to derive a DMEL (derived minimal effect level) similarly as for e.g. genotoxic carcinogens. The ECHA Guidance on information requirements and chemical safety assessment, Chapter R.8: Characterisation of dose [concentration]-response for human health (ECHA, 2012) provides guidance on that. For non-threshold effects, the underlying assumption is that a no-effect-level cannot be established and a DMEL therefore expresses an exposure level corresponding to a low, possibly theoretical risk. Cancer risk levels of 10^{-5} and 10^{-6} could be seen as indicative tolerable risk levels when setting DMELs for workers and the general population, respectively (ECHA, 2012).

Extrapolation from the effects seen at high dose levels used in toxicological studies, which are associated with high risks [i.e. 10% or higher] to the low dose levels of human exposure associated with risk level of very low concern has to be done using modelling. The assessment of dose levels associated with these low risks, i.e. in any quantitative terms, is not possible, as they cannot be verified either experimentally or by epidemiological studies (ECHA, 2012). The default approach is to extend a straight line, i.e. linear extrapolation. In general, there is no need for deviation from this default linear approach, but this may be done on a case-by-case basis, and clearly needs expert judgement. For systemic non-threshold effects, only an assessment factor for differences in metabolic rate (allometric scaling) is to be applied when using the linear approach (ECHA, 2012).

An alternative approach to linear extrapolation to low doses for cancer risk is also included in the ECHA Guidance. It is called The 'Large Assessment Factor' approach (or "EFSA" approach). The default additional assessment factors for cancer risk could also be used as basis for a similar approach for EDs, see table 2. $BMDL_{10}$ is used as starting point.

Table 2 The 'Large Assessment Factor' approach, factors for endocrine disruption when $BMDL_{10}$ is used as starting point (modified from a similar table using similar assessment factors for cancer in ECHA 2012)

Assessment factor	Default value
Interspecies	10
Intraspecies	10
Nature of endocrine process	10
The point of comparison ($BMDL_{10}$ is not a NOAEL)	10

5. Case studies

To illustrate the outcomes when using the different approaches described in section 4 for setting reference values, two EDs were selected, i.e. procymidone and DEHP. The toxicological data used as points of departure for the calculations for these EDs are considered to be realistic data and they come from published papers. A third case study on substance X was developed to illustrate the outcomes when there is a lack of studies with exposure during development. The data for substance x is mainly inspired by the data available for a specific ED before reproductive toxicity studies were performed. It should be emphasized that changes of any of the data for the three case studies will not change the overall picture as the same data has been used as point of departure for the calculation under each of the various approaches.

5.1 Procymidone

The pesticide procymidone is used here as an illustrative example, even though it is no longer approved for use in EU. The ADI (acceptable daily intake) for procymidone is 2.8 µg/kg.

The following data from the rat study in Hass et al. (2007) were used for procymidone:
In a reproductive toxicity study in rats a dose of 25 mg/kg bw/day caused a non-significant 7% incidence of genital malformations in the male offspring (NOAEL). The malformation incidence was about 10 % at 28 mg/kg bw/day. The NOAEL for decreased anogenital distance (AGD) and increased nipple retention (NR) was 10 mg/kg bw/day.

Thus the following data were used:

- Decreased AGD and increased nipple retention: NOAEL, = 10 mg/kg, LOAEL = 25 mg/kg; BMD_{10} ~ 10 mg/kg. $BMDL_{10}$ ~ 5 mg/kg (estimated to be $BMD_{10}/2$)

- Malformations: NOAEL, 7% = 25 mg/kg, ca. 10% with malformations at 28 mg/kg; $BMDL_{10}$ ~ 14 mg/kg, (estimated to be $BMD_{10}/2$).

DNEL, $DNEL_{ED}$ and $DMEL_{ED}$ were derived using several approaches as shown in table 3. For the $DMEL_{ED}$ approach, only the malformation incidence was used as the basis as this appeared more similar to the use of the cancer incidence when deriving DMEL for cancer than using decreased AGD and increased nipple retention.

The results with regards to DNEL, $DNEL_{ED}$ and $DMEL_{ED}$ clearly illustrate that the use of the different approaches leads to marked differences in levels of procymidone considered safe for humans, i.e. from 0.04-100 µg/kg. However, it also illustrates that the lowest $DNEL_{ED}$ where relevant assessment factors are used, the $DMEL_{ED}$ based on the 'Large Assessment Factor' approach and the $DMEL_{ED}$ obtained by linear extrapolation to 10^{-5} incidence are of similar magnitude, i.e. they range from 0.4-1.4 µg/kg.

Table 3 DNEL, $DNEL_{ED}$ and $DMEL_{ED}$ for procymidone based on different approaches[a]

Approach	Effect	Starting point	Assessment factors	Result
1) DNEL, traditional	AGD/NR	NOAEL =10 mg/kg	10x10 = 100	100 µg/kg
2) $DNEL_{ED}$	AGD/NR	NOAEL = 10 mg/kg	10x10x10 (ED MOA) = 1000	10 µg/kg
3) $DNEL_{ED}$	AGD/NR	$BMDL_{10}$ ~ 5 mg/kg	10x10x10 (ED MOA)x10 ($BMDL_{10}$ is an effect level) = 10,000	0.5 µg/kg
4) $DNEL_{ED}$	Malformations	NOAEL = 25 mg/kg	10x10x10 (ED MOA) x 2(severity) x 3(low power) = 6,000	4.1 µg/kg
5) $DNEL_{ED}$	Malformations	$BMDL_{10}$ ~ 14 mg/kg	10x10x10 (ED MOA)x10 ($BMDL_{10}$ is an effect level) x2 (severity) = 20,000	0.7 µg/kg
6) $DMEL_{ED}$, 'Large Assessment Factor'	Malformations	$BMDL_{10}$ ~ 14 mg/kg	10x10x10x10 = 10,000	1.4µg/kg
7) $DMEL_{ED}$, risk ~10^{-5}	Malformations	$BMDL_{10}$ ~14 mg/kg -> 1.4 µg/kg by linear extrapolation to 10^{-5} incidence	4[b]	0.4 µg/kg
8) $DMEL_{ED}$, risk ~10^{-6}	Malformations	$BMDL_{10}$ ~14 mg/kg -> 0.14 µg/kg by linear extrapolation to 10^{-6} incidence	4[b]	0.04 µg/kg

a. Note that $BMDL_{10}$ is estimated to be $BMD_{10}/2$
b. Factor for allometric scaling. Factor 4 is used as the studies are in rats (ECHA, 2012).

5.2 DEHP

A threshold has not been "demonstrated" for the anti-androgenic effects of the phthalates. DEHP is used an illustrative phthalate, because there are many relevant data available for this substance. The dose-response curves for the effects of DEHP on anogenital distance (AGD) and the frequency of external genital malformations in Christiansen et al. (2010) does not indicate a threshold at the doses studied (see figure 1).

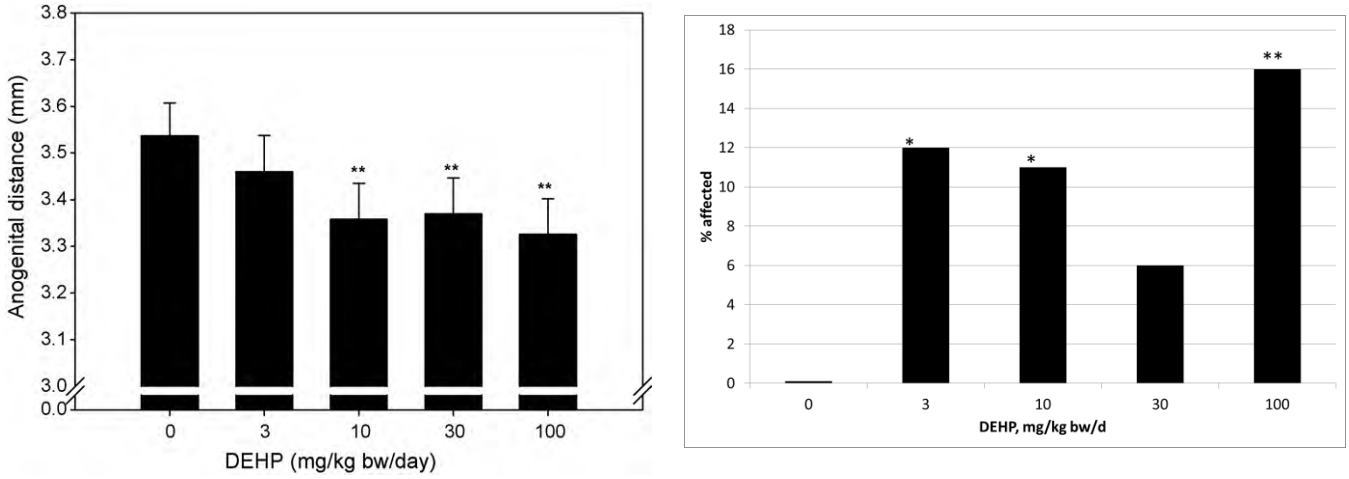

Fig. 1. Mean anogenital distance (AGD) on PND 1 (left) and mild external dysgenesis on PND 16 (right) in male rat offspring of dams administered corn oil (control), 3, 10, 30 or 100 mg/kg-d DEHP from GD 7 to PND 16 (from Christiansen et al. (2010). Least square means + SEM are shown for AGD and the data are corrected for body weight and litter effect. Frequency of affected male offspring is shown for mild external dysgenesis at PND 16; *Indicates $p < 0.05$; **Indicates $p < 0.01$.

Various outcomes with regards to estimating tolerable or low risk levels for humans exposed to DEHP have been estimated to evaluate whether using the DNEL or DMEL approach leads to differences of importance (see Table 4).

The following data from the rat studies in Christiansen et al. (2010) were used as starting point:

- Decreased AGD and increased nipple retention (NR): NOAEL for = 3 mg/kg, BMD_{10}~10 mg/kg., $BMDL_{10}$ ~ 5 mg/kg (estimated to be $BMD_{10}/2$)
- Malformations (mild dysgenesis of external genitalia): LOAEL for ~10% = 3 mg/kg; $BMDL_{10}$ ~ 1.5 mg/kg (estimated to be $BMD_{10}/2$)

DNEL, $DNEL_{ED}$ and $DMEL_{ED}$ were derived using several approaches as shown in table 4. For the $DMEL_{ED}$ approach, only the malformation incidence was used as the basis as this appeared more similar to the use of the cancer incidence when deriving DMEL for cancer than using decreased AGD and increased nipple retention. The results illustrate that the use of the different approaches leads to marked differences in the external dose value considered as a safe or low risk level for humans, i.e. the values range from 0.004 to 50 µg/kg. As such this illustrates that the DNEL of 50 µg/kg bw/d applied for DEHP in the EU risk assessment is not too cautious. The same would most

likely be the cases for DBP, DiBP and BBP as they have similar anti-androgenic effects as DEHP. However, the results shown in table 4 also illustrate that the lowest $DNEL_{ED}$ where relevant assessment factor are used, the $DMEL_{ED}$ based on the 'Large Assessment Factor' approach and the $DMEL_{ED}$ obtained by linear extrapolation to 10^{-5} incidence are of similar magnitude, i.e. they range from 0.04-0.5 µg/kg.

Table 4. DNELs and DMELs for DEHP based on different approaches[a]

Approach	Effect	Starting point	Assessment factors	Result
1) DNEL, applied by ECHA	Small male reproductive organs and minimal testis atrophy	NOAEL = 4.8 mg/kg	10 x 10 = 100	50 µg/kg
2) $DNEL_{ED}$	AGD/NR	NOAEL =3 mg/kg	10 x 10 x 10 (ED MOA) = 1,000	3 µg/kg
3) $DNEL_{ED}$	AGD/NR	$BMDL_{10}$ ~ 5 mg/kg	10 x 10 x10 (ED MOA)x 10 ($BMDL_{10}$ is an effect level) = 10,000	0.5 µg/kg
4) $DNEL_{ED}$	Malformations (mild)	LOAEL = 3 mg/kg	10 x 10 x10 (ED MOA) x 2(severity) x 3 (low power) x 10 (LOAEL to NOAEL) = 60,000	0.05 µg/kg
5) $DNEL_{ED}$	Malformations (mild)	$BMDL_{10}$ ~ 1.5 mg/kg	10 x 10 x10 (ED MOA)x 10($BMDL_{10}$ is an effect level) x 2(severity) = 20,000	0.075 µg/kg
6) DMEL, 'Large Assessment Factor'	Malformations	$BMDL_{10}$ ~ 1.5 mg/kg	10*10*10*10 = 10,000	0.15 µg/kg
7) DMEL, risk level ~10^{-5}	Malformations	$BMDL_{10}$ ~1.5 mg/kg -> 0.15 µg/kg by linear extrapolation (from 10^{-1} incidence to 10^{-5} incidence)	4[b]	0.04 µg/kg
8) DMEL, risk level ~10^{-6}	Malformations	$BMDL_{10}$ ~1.5 mg/kg -> 0.015 µg/kg by linear extrapolation (from 10^{-1} incidence to 10^{-6} incidence)	4[b]	0.004 µg/kg

a. Note that $BMDL_{10}$ is estimated to be $BMD_{10}/2$
b. Factor for allometric scaling. Factor 4 is used as the studies are in rats.

5.3 Substance X (example inspired mainly by old data for some phthalates later on identified as EDs)

Substance X has been evaluated as an ED based on anti-androgenic activity *in vitro*, adverse effects on reproductive organs in male animals exposed in adulthood (oral toxicity, 28 days, young adult animals) and a plausible link between the ED MoA and the adverse effects. A threshold has not been demonstrated.

The following data were used as starting point:

- Testicular atrophy and decreased weight of epididymis in several repeated dose toxicity studies in adult male rats (OECD TG 407, Repeated Dose 28-Day Oral Toxicity Study in Rodents). NOAEL = 500 mg/kg, BMD_{10}~800 mg/kg, $BMDL_{10}$ ~ 400 mg/kg (estimated to be $BMD_{10}/2$)
- There are no reproductive toxicity studies available

DNEL, $DNEL_{ED}$ and $DMEL_{ED}$ were derived using several approaches as shown in table 5. The results illustrate that the use of the different approaches leads to marked differences in the external dose value considered as a safe or low risk level for humans, i.e. the values range from 1.0 to 5,000 µg/kg. However, the results also illustrate that the lowest $DNEL_{ED}$ where relevant assessment factor are used (i.e. approach 2 and 3), the $DMEL_{ED}$ based on the 'Large Assessment Factor' approach and the $DMEL_{ED}$ obtained by linear extrapolation to 10^{-5} incidence are of similar magnitude, i.e. they range from 4-40 µg/kg.

Table 5. DNELs and DMELs for substance X based on different approaches[a]

Approach	Effect	Starting point	Assessment factors	Result
1) DNEL, as normally applied by ECHA	Testicular atrophy, decreased weight of epididymis	NOAEL = 500 mg/kg	10 x 10 = 100	5,000 µg/kg
2) $DNEL_{ED}$	Testicular atrophy, decreased weight of epididymis	NOAEL =500 mg/kg	10 x 10 x10 (ED MOA) x 2(severity) x 3 (low power) x 10 (not developmental exposure) = 60,000	8.3 µg/kg
3) $DNEL_{ED}$	Testicular atrophy, decreased weight of epididymis	$BMDL_{10}$ ~ 400 mg/kg	10 x 10 x10 (ED MOA)x 10 ($BMDL_{10}$ is an effect level) x 10 (not developmental exposure) = 100,000	4 µg/kg
4) DMEL, 'Large Assessment Factor'	Testicular atrophy, decreased weight of epididymis	$BMDL_{10}$ ~ 400 mg/kg	10*10*10*10 = 10,000	40 µg/kg
5) DMEL, risk level ~10^{-5}	Testicular atrophy, decreased weight of epididymis	$BMDL_{10}$ ~400 mg/kg -> 40 µg/kg by linear extrapolation (from 10^{-1} incidence to 10^{-5} incidence)	4[b]	10 µg/kg
6) DMEL, risk level ~10^{-6}	Testicular atrophy, decreased weight of epididymis	$BMDL_{10}$ ~1.5 mg/kg -> 4 µg/kg by linear extrapolation (from 10^{-1} incidence to 10^{-6} incidence)	4[b]	1.0 µg/kg

a. Note that $BMDL_{10}$ is estimated to be $BMD_{10}/2$, b. Factor for allometric scaling. Factor 4 is used as the studies are in rats.

6. Outcome of the 1st international workshop

The project team was gathered together with invited experts from abroad and from the Danish EPA for a workshop day to discuss uncertainties and possible approaches for risk assessment of endocrine disruptors on the basis of prepared workshop material that was circulated before the workshop. The set-up for the workshop is further detailed in appendix 1.

The workshop discussed the various uncertainties related to assessment of EDs i.e. the existence of thresholds for effects, the issue of exposure during sensitive window during development, the existence of non-monotonic dose responses as well as the limited sensitivity of various test methods. A large majority of the participants in the international workshop concluded that there is a need for addressing these uncertainties when assessing the risk of EDs. The problems that most of the data available come from old studies where test methods without ED-relevant endpoints have been used, that many substances have not been tested with *in utero* exposure of the animals and the general lack of data from comprehensive testing with sensitive endpoints for ED effects were highlighted. The existence of non-monotonic dose-responses was also seen as a challenge for setting appropriate reference values as it will be very difficult to get data from low dose studies and there may be a need for an extra assessment factor to address this uncertainty.

When discussing assessment factors it was mentioned that the use of several specific uncertainty factors makes the assessment more transparent than the use of one overall factor. In addition, this implies that when more information is available for e.g. a specific assessment factor, this assessment factor can be reduced. The case studies were very useful in demonstrating how the use of different approaches for setting reference values can lead to very different outcomes even though the same initial data are used for each approach. This demonstration prompted a lively debate that ended up by anonymous voting for the preferred approach. In the end it turned out that the workshop participants generally found that the approach for setting reference values should not be based on whether there exists a threshold for effects or not, because it is impossible to determine a threshold for EDs with reasonable certainty. Therefore, a unified approach for all EDs was recommended. An anonymous voting with regards to the non-threshold approaches for EDs showed that there were equal numbers of the supporters of the large assessment factor approach and the approach with a linear extrapolation.

7. Outcome of the 2nd international workshop

The participants in the two-day workshop were from 14 countries and included risk assessors from European authorities and experts with research background in endocrine disruption. The workshop comprised introductory presentations by international experts, plenum discussions and group discussions. The workshop report including details on the set-up is shown in Appendix 2.

In plenary sessions, all discussion groups presented summaries of their deliberations for debate. The discussions were lively and constructive. In addition, there appeared to be a general consensus or a clear majority agreement on most of the main issues discussed.

The following conclusions were drawn at the workshop:

• The four uncertainties discussed, i.e. 1) Lack of studies with exposure during sensitive windows, 2) Limited sensitivity of many regulatory testing methods with regard to relevant endpoints addressed and power/robustness of study, 3) The occurrence of ED-induced NMDR and 4) The existence of thresholds, are the most important uncertainties related to ED assessment

- The two first uncertainties are considered as more important than the two latter ones, but those are also of importance.
- The current risk assessment of EDs only address the uncertainties specific for ED assessment to a limited extend
- There is an urgent need for addressing uncertainties related to ED assessment due to the risk of irreversible and delayed effects
- It is recommended to use a non-threshold approach as default to address specific uncertainties related to assessment of ED when deriving reference doses for EDs. This could be a rebuttable assumption, thus if strong evidence for threshold is/becomes available then a threshold approach could be used
- The participants were equally divided between recommending two approaches for the derivation of references levels, i.e. 1) linear extrapolation (to e.g. 10^{-6} incidence) or 2) derivation of a reference dose (DMEL) using additional factors covering specific uncertainties related to assessment of ED including also as default an additional ED UF of 10-100.
- Using BMDL as point of departure is recommended (when possible) as this takes the power of the data into consideration

8. Summary, conclusions and recommendations

Several uncertainties exist when it comes to the assessment of EDs, however, despite this there is at the same time an ongoing assessment of substances under various regulations, e.g. REACH, and it is under discussion how the risk of exposure to EDs should be assessed.

This project has discussed some major uncertainties related to derivation of tolerable exposure levels for EDs which include:
- Threshold or no threshold
- Exposure during sensitive windows
- Limited sensitivity of (many) testing methods: relevant endpoints addressed and power/ robustness of study
- Potential for NMDR

Different approaches for setting reference values based on whether a threshold for effects exists have been discussed. Three case-studies were used to demonstrate the different outcomes when using different approaches for setting tolerable reference values even by using the same data for each approach. This clearly illustrated that the different approaches, i.e. DNEL, $DNEL_{ED}$ with relevant additional assessment factors and $DMEL_{ED}$ approach, lead to marked differences in the reference values, i.e. those dose levels considered tolerable for humans. However, the case studies also illustrated that the $DNEL_{ED}$ where relevant assessment factor for ED uncertainties are used, the $DMEL_{ED}$ based on the 'Large Assessment Factor' approach as well as the $DMEL_{ED}$ obtained by linear extrapolation to 10^{-5} incidence are of similar magnitudes. This will generally be the case for substances where the same point of departure is used for the derivation of the reference value, because the assessment factors will be similar to those for the 3 case studies. Thus each of these three approaches could be used as a unified approach for deriving tolerable reference levels for risk assessment of all EDs irrespective of a threshold assumption or not.

The uncertainty aspects and the case studies were discussed at two workshops (appendix 1 and 2) including an international workshop (in May 2017), where the around 40 participants were from 14 countries and included risk assessors from European authorities and experts with research background in endocrine disruption. There appeared to be a general consensus or a very clear majority agreement on the main issues discussed.

The workshop participants generally found that the approach used for setting reference values should not be based on whether there exists a threshold for effects or not, because it is difficult or impossible to determine a threshold with reasonable certainty. Therefore, a unified approach using a non-threshold approach for deriving a $DMEL_{ED}$ as default for all EDs was recommended. Concerning derivation of a $DMEL_{ED}$, there was an equal distribution between those preferring to use an approach with relevant assessment factor for ED uncertainties and those preferring to make a linear extrapolation to address the uncertainties related to assessment of EDs.

It is difficult to choose between these two approaches based on scientific arguments as they both have strength and limitations. For a $DMEL_{ED}$ derived using additional assessment factors for uncertainties related to assessment of ED the strength and limitations include:
- Transparency about the uncertainties considered
- May lead to better promotion of new studies
- May be complicated as each additional assessment factor has to be decided

For the linear extrapolation approach, the strength and limitations include:
- Transparency with regards to the risk level considered tolerable by risk managers (e.g. 10^{-5}or 10^{-6})
- May be complicated as the approach requires data and methods that allows derivation of $BMDL_{10}$
- Easy to calculate from $BMDL_{10}$ to $DMEL_{ED}$ because $BMDL_{10}$ is equal to 10^{-1} in animals and e.g. risk level 10^{-5} or 10^{-6} is derived by dividing by 10^4 or 10^5 and the factor for allometric scaling, e.g. 4 for rats.

Thus, the overall conclusions and recommendations for this project are:

- The most important uncertainties related to ED assessment are:

1) Lack of studies with exposure during sensitive windows,
2) Limited sensitivity of many regulatory testing methods with regard to relevant endpoints addressed and power/robustness of study,
3) The potential for ED-induced NMDR and
4) The existence of threshold

- There is an urgent need for addressing the uncertainties related to ED assessment due to the risk of irreversible and delayed effects
- It is recommended to use a non-threshold approach as default to address specific uncertainties related to assessment of ED when deriving reference doses for EDs. This could be a rebuttable assumption, thus if strong evidence for threshold is/becomes available then a threshold approach could be used
- One of two approaches for the derivation of references levels (DMEL) are recommended, i.e. 1) linear extrapolation (to e.g. 10^{-5} or 10^{-6} incidence) or 2) derivation of a reference dose using additional factors covering specific uncertainties related to assessment of ED including also as default an additional ED assessment factor of 10-100.
- It is difficult to choose between these two approaches based on scientific arguments as they both have strength and limitations that also include non-scientific issues (e.g. feasibility and risk level considered tolerable by risk managers).
- Using BMDL as point of departure is recommended (when possible) as this takes the power of the data into consideration

9. References

Christiansen S, Boberg J, Axelstad M, Dalgaard M, Vinggaard AM, Metzdorff SB and Hass U. Low-dose perinatal exposure to di(2-ethylhexyl) phthalate induces anti-androgenic effects in male rats. Reprod Toxicol, 2010; 30(2) 313-21. 10.1016/j.reprotox.2010.04.005

Coady KK, Biever RC, Denslow ND, Gross M, Guiney PD, Holbech, H, Karouna-Renier NK, Katsiadaki I, Krueger H, Levine SL, Maack G, Williams M, Wolf JC, Ankley GT, (2017) Current Limitations and Recommendations to Improve Testing for the Environmental Assessment of Endocrine Active Substances. Integrated Environmental Assessment and Management, 13(2), 302-316

ECHA, 2012. Guidance on information requirements and chemical safety assessment. Chapter R.8: Characterisation of dose [concentration]-response for human health. Version 2.1. November 2012. https://echa.europa.eu/documents/10162/13632/information_requirements_r8_en.pdf

Hass U, Dalgaard M, Jarfelt K, Kledal TSA (2004). OECD conceptual framework for testing and assessment of endocrine disruptors as a basis for regulation of substances with endocrine disrupting properties: Nordic Chemicals Group Report. 2004. 70 p. TemaNord 2004:555 http://norden.diva-portal.org/smash/get/diva2:702046/FULLTEXT01.pdf

Hass, U., Scholze, M., Christiansen, S., Dalgaard, M., Vinggaard, A.M., Axelstad, M., Broeng Metzdorff, S., Kortenkamp, A. (2007) Combined exposure to anti-androgens exacerbates disruption of sexual differentiation in the rat. Environ Health Perspect 115, 122-28.

Hass U, Christiansen S, Boberg J, Vinggaard AM, Andersson A-M, Skakkebæk NE, Katrine Bay, Holbech H, Bjerregaard P (2011). Criteria for Endocrine Disrupters. Report from Danish Centre on Endocrine Disrupters for Danish EPA. Annex A in: https://mst.dk/media/mst/9106718/danskeforslag.pdf

Hass U, Christiansen S, Axelstad M, Sørensen KD and Boberg J. Input for the REACH-review in 2013 on endocrine disrupters, Danish Centre on Endocrine Disrupters. Final report 21 March 2013, p. 1-50 http://mst.dk/media/mst/9106721/rapport_input_for_the_reach-review.pdf

Holbech H, Kinnberg K, Petersen GI, Jackson P, Hylland K, Norrgren L, Bjerregaard P. Detection of endocrine disrupters: evaluation of a fish sexual development test (FSDT) Comp. Biochem. Physiol. C Pharmacol. Toxicol. Endocrinol., 144 (2006), pp. 57-66

JRC (2013) Thresholds for Endocrine Disrupters and Related Uncertainties, Report of the Endocrine Disrupters Expert Advisory Group. Publications Office of the European Union, 2013 http://publications.jrc.ec.europa.eu/repository/handle/JRC83204

OECD test guidelines (human health, section 4): http://www.oecd-ilibrary.org/environment/oecd-guidelines-for-the-testing-of-chemicals-section-4-health-effects_20745788

OECD 2001, Test No. 416: Two-Generation Reproduction Toxicity, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, Paris, https://doi.org/10.1787/9789264070868-en.

OECD (2012) Validation Report (Phase 2) For The Fish Sexual Development Test For The Detection Of Endocrine Active Substances. 2012. Series on Testing and Assessment No. 142 http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2011)23/REV1&doclanguage=en

OECD (2015). Feasibility Study for Minor Enhancements of TG 421/422 with ED Relevant Endpoints. Environment, Health and Safety Publications, Series on Testing and Assessment (No. 217), Organisation for Economic Cooperation and Development, Paris. http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2015)24&doclanguage=en

OECD 2016a, Test No. 421: Reproduction/Developmental Toxicity Screening Test, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, Paris, https://doi.org/10.1787/9789264264380-en.

OECD 2016b, Test No. 422: Combined Repeated Dose Toxicity Study with the Reproduction/Developmental Toxicity Screening Test, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, Paris, https://doi.org/10.1787/9789264264403-en.

OECD 2018a. OECD (2018), Revised Guidance Document 150 on Standardised Test Guidelines for Evaluating Chemicals for Endocrine Disruption, OECD Series on Testing and Assessment, No. 150, OECD Publishing, Paris, https://doi.org/10.1787/9789264304741-en. https://www.oecd-ilibrary.org/environment/guidance-document-on-standardised-test-guidelines-for-evaluating-chemicals-for-endocrine-disruption-2nd-edition_9789264304741-en

OECD 2018b, Test No. 443: Extended One-Generation Reproductive Toxicity Study, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, Paris, https://doi.org/10.1787/9789264185371-en. https://www.oecd-ilibrary.org/environment/test-no-443-extended-one-generation-reproductive-toxicity-study_9789264185371-en

Örn S, Holbech H, Madsen TH, Norrgren L, Petersen GI. 2003. Gonad development and vitellogenin production in zebrafish (Danio rerio) exposed to ethinylestradiol and methyltestosterone. Aquat Toxicol 65:397–411.

Parrott, J, P Bjerregaard, K Brugger, LE Gray, T Iguchi, S Kadlec, L Weltje, J Wheeler.(2017). Uncertainties in biological responses that influence hazard and risk approaches to the regulation of endocrine active substances. Integrated Environmental Assessment and Management. 13(2): 293-301.

Vandenberg LN, Colborn T, Hayes TB, Heindel JJ, Jacobs, Jr DR., Lee D-H, Shioda T, Soto AM, vom Saal FS, Welshons WV, Zoeller RT, and Myers JP. Hormones and Endocrine-Disrupting Chemicals: Low-Dose Effects and Nonmonotonic Dose Responses. Endocrine Reviews, June 2012, 33(3):378-455

WHO/IPCS 2002. Global assessment of the state-of-the-science of endocrine disruptors. WHO/IPCS/EDC/02.2. Geneva: World Health Organisation. http://www.who.int/ipcs/publications/new_issues/endocrine_disruptors/en/

Appendix 1. Workshop on Interpretation of knowledge on endocrine disrupting substances (EDs) – what is the risk? DTU, November 10th, 2015

1) List of invited participants

Present:

From DTU Food: Ulla Hass, Sofie Christiansen, Marta Axelstad, Anne Marie Vinggaard, Karen Mandrup

From Southern University of Denmark: Henrik Holbech, Poul Bjerregaard, Karin Lund Kinnberg

University Hospital, Denmark: Niels Erik Skakkebæk, Katrine Bay, Anna-Maria Andersson

Danish EPA: Pia Juul Nielsen, Rikke Holmberg, Marie Holmer

Richard Sharpe, The University of Edinburgh, Scotland, Great Britain

Betty Hakkert, Centre for Safety of Substances and Products, The Netherlands

Anna Beronius, Department of Environmental Science and Analytical Chemistry (ACES), Stockholm University, Sweden

Absent:

Julie Boberg, DTU Food

Sharon Munn, Joint Research Centre, Institute for Health and Consumer Protection, Systems Toxicology, Ispra, Italy

Claire Beausoleil and Christophe Rousselle, French Agency for Food, Environmental and Occupational Health & Safety (ANSES), Paris, France

2) Agenda

9.00-9.20	Welcome and tour de table	DTU
9.20 – 9.40	Introduction and presentation of draft report	Ulla Hass
9.40 -10.00	NMDR and EDs	Sofie Christiansen
10.00-10.30	Input from Richard, Anna and Betty	
10.30-11.00	Coffee Break	
11.00 – 12.00	Round table discussion: Need for and how to address uncertainties?	All
12.00 – 12.40	Lunch Break	
12.40 – 13.40	Round table discussion: How to derive reference/tolerable levels for EDs, i.e. a threshold-based approach without or with additional factors or a non-threshold-based approach?	All
13.40 -14.30	For the two case studies, i.e. procymidone and DEHP, which reference level would you recommend?	All
14.30-15.00	Coffee Break	
15.00 – 16.00	Final discussion	All
16.00-16.30	Conclusion of the Workshop and how to proceed	DTU

3) Introduction (the draft report was distributed before the workshop)

Focus: Make recommendations on how to derive "tolerable" dose-levels for humans based on toxicology studies.
Questions addressed:

1. The uncertainties for EDs include
    i) Threshold or no-threshold
    ii) Sensitive windows
    iii) Limited sensitivity of (many) testing methods
    iv) NMDR
    v) Others?

    Is there a need for and if so, how can the uncertainties be addressed?
2. The draft report describes several possibilities for deriving reference/tolerable levels for EDs, i.e. a threshold-based approach without or with additional factors and a non-threshold-based approach. Which of these methods is scientifically most relevant for deriving a reference level for an ED? Threshold or non-threshold approach? Do you have alternative proposals?
3. Specifically, for the two case studies, i.e. procymidone and DEHP, which reference level would you recommend?

4) Meeting notes

These meeting notes aim to reflect the many topics discussed and the various inputs given. The overall input from the workshop was very useful and has been considered when developing the final report.

With regards to the need to address the *uncertainties for EDs* an anonymous vote was made showing that the majority of the participants found that there is a need for this (Yes: 13; No: 2; Mixed yes & no: 1)

As regards *thresholds for EDs* it was mentioned that these are debated for EDs as well as for carcinogens and that there are fewer mechanistic data for EDs. The background for this debate appears somewhat similar for EDs and for carcinogens, i.e. the endogenous hormones/ genotoxicants gives a background 'effect' and therefore small effects can have a significant impact. It was mentioned that it seems that we can agree that there is sufficient grounds for uncertainty if there is a threshold or not. The possibility for defending that a threshold factor is specifically relevant for EDs and not for other chemicals was also discussed. It was mentioned that such a factor may also be relevant for assessment of other toxicities as well. It was however, recognized further discussion of this was beyond the scope of this project. It was also expressed that in case we recommend a method that is currently used for assessment of non-threshold effects, this does not imply that our method assumes non-threshold. An anonymous vote regarding non-threshold approaches for EDs, i.e. 1) Large Assessment Factor method or 2) linear extrapolation method showed an equal distribution of preference (7; 7).

It was recognized that EDs may in some cases induce *NMDR*. As guideline studies cover only 3 doses, NMDR may not be detected and therefore an assessment factors can be relevant to address

this uncertainty. If NMDR has been found, e.g. a Tamoxifen-like effect, it was proposed to use the low end of the NMDR as point of departure.

It was generally agreed that the issue of *sensitive windows* needs to be addressed and it was discussed how to do it. If only data from adult exposure is available, an assessment factor was considered needed. The size of this factor may be estimated based on existing data, but the most reliable data would come from studies with exposure during sensitive windows.

The *quality of data* was identified an issue to be considered, because in several cases assessment of chemicals with old and insufficient data has to be done. This means, that if an ED effect is detected, it may not be the most sensitive ED effect. In such cases, an additional assessment factor appears relevant.

It was mentioned that a method with very few and maybe higher *assessment factors* would lead to an easier method. On the other hand, it was mentioned that the use of more factors reflecting each uncertainty makes the method more transparent. Also, it means that better data for one of the uncertainties makes it easier to modify or leave out the assessment factor for that uncertainty.

It was agreed that the use of *BMDL* as point of departure instead of NOAEL was preferred, because BMDL takes the power of the study and the magnitude of effect into consideration. However, it was discussed whether it will always be possible to obtain BMDL based on data from e.g. guideline studies with only 3 dose levels. In addition, it was mentioned that depending on the various (curve fitting) models used, the variation may become quite high and this would lead to a very low BMDL. It was stated that a BMD_{10} can in most cases be calculated based on such data. It was proposed that in cases where the $BMDL_{10}$ is too uncertain, an 'estimated $BMDL_{10}$' (e.g. BMD10/2) could be used. This approach is used for the case studies in the report, but was not further discussed. Also, it was asked whether it would be possible to start from a NOAEL and somehow calculate that into a BMDL-like level. This was considered as an interesting idea, but also recognized as a difficult task.

Recommendations/conclusions from workshop

The uncertainties for EDs include

- Threshold or no threshold
- Sensitive windows
- Limited sensitivity of (many) testing methods: relevant endpoints addressed and power/ robustness of study
- NMDR

A large majority of the participants found that there is a need for addressing these uncertainties (first vote) when setting tolerable exposure levels for humans.

Among the non-threshold approaches, there was an equal distribution between preferring the large assessment factor approach or linear extrapolation to address the uncertainties (second vote)

The assessment factors etc. will be revised to reflect the discussion at the workshop

The participants generally found that the approach should not be based on threshold or not, because it is impossible to determine a threshold. Therefore, a unified approach for all EDs was recommended.

Appendix 2: Workshop Report for International workshop on Risk Assessment of Endocrine Disruptors: derivation of reference doses (tolerable exposure levels) for humans, 22-23 May 2017

Auditorium 93, University Hospital (Rigshospitalet), Blegdamsvej 9, Copenhagen 2100 Ø, Denmark

Overview:

(The above mentioned appendices (2a-2e) is not part of the published report)

---

1. Background

The workshop was an important part of an ongoing project based on a commission from the Danish Environmental Protection Agency and was directed towards risk assessors from European authorities and experts with research background in endocrine disruption.

The main purpose of the workshop was to discuss the derivation of reference doses (tolerable exposure levels) for endocrine disruptors (EDs) including the need for addressing specific uncertainties related to the assessment of ED in the process of hazard and risk assessment. The focus was to make recommendations with regards to the derivation of reference doses for humans based on toxicology studies.

In November 2015, the project group organized a small workshop on this topic and a draft report was made (the unpublished report was distributed as part of the workshop material). Here, the most important uncertainties as regards derivation of reference doses for EDs were evaluated to be:

- Exposure during sensitive windows
- Limited sensitivity of (many) regulatory testing methods with regards to relevant endpoints addressed and power/robustness of study
- The potential for ED-induced Non-monotonic dose-response (NMDR)
- The existence of a threshold

Depending on the regulatory context the terms Derived No effect Level (DNEL), Derived Minimal Effect Level (DMEL), Acceptable Daily Intake (ADI) and Tolerable Daily Intake (TDI) are used for the reference dose considered as sufficiently safe for humans. This dose is derived by applying assessment factors related to the extrapolation from experimental animals to humans and the differences within the human population and it is used for comparison with human exposure levels to evaluate the risk for humans. However, for EDs it appears relevant to discuss whether this is protective enough in light of the identified uncertainties related to ED assessment, and especially if or how this can be addressed.

The participants in the workshop were from 14 countries and included risk assessors from European authorities and experts with research background in endocrine disruption (see List of participants in Appendix 2a).

The agenda for the workshop is shown in Appendix 2b and comprised introductory presentations (see Appendix 2c), plenum discussions and group discussions. During the workshop the participants were distributed in to four discussion groups (DGs) to work in separate parallel sessions dealing with the same topic, and the DGs were the same for the 1st and 2nd discussion rounds. Background and topic outline for the discussions were distributed to the participants at the workshop (appendix 2d and 2e). For each DG, a moderator was selected by the organisers prior to the workshop, and each DG selected a group member as rapporteur. The selected moderators were Andreas Kortenkamp, Betty Hakkert, Laura Vandenberg and Anna-Maria Andersson.

Material supporting the discussions was sent out before the workshop and included the draft report from 2016 from the project, i.e. ‘Interpretation of knowledge on endocrine disrupting substances (EDs) – what is the risk?’(2016).

2. Introduction and welcome to the workshop

This was done by project manager Ulla Hass, Technical University of Denmark, facilitator/ moderator Line Friis Frederiksen, Line•Friis ApS, and Pia Juul Nielsen, Danish Environmental Protection Agency.

Ulla Hass welcomed all participants to the workshop and introduced Line Friis Frederiksen as the workshop moderator/facilitator. Line introduced herself and stressed that the workshop is important not only from a scientific and regulatory perspective, but also from the consumer’s perspective referring to her own concerns as a mom of two kids. She then explained her role as the facilitator/moderator and briefly summed up the workshop program. Next, Pia Juul Nielsen introduced herself and presented the background for the project and the workshop. She explained how endocrine disruptors have been on the agenda in Denmark for many years and how Denmark has contributed with policy and scientific inputs to the EU regulatory work on endocrine disruptors. She pointed out that substances are already identified as endocrine disruptors under REACH and that there is a need to consider whether the current risk assessments approaches are adequate to

apply to endocrine disruptors. The Danish EPA has therefore asked the Centre on Endocrine Disruptors (CEHOS) to give scientific advice on how to best make human risk assessment of endocrine disruptors. This resulted in the organisation of the present workshop where various experts, member states and scientific expert groups and committees have been invited to contribute to the discussions and give inputs.

3. Brief summary of the workshop presentations

In this section, a brief summary of formal workshop presentations is given. The presentations are available in Appendix 2c (one PDF file).

Introduction and aims of the workshop, Ulla Hass, Technical University of Denmark, DK
Ulla Hass presented the aims and focus of the workshop, which were:

- To discuss derivation of reference doses (tolerable exposure levels) for endocrine disruptors (EDs) with focus on uncertainties related to ED assessment
- To come up with recommendations with regards to the derivation of reference doses for humans based on *in vivo* toxicology studies

Moreover, she stated that at this workshop the focus was on substances evaluated as an ED based on the WHO/IPCS definition of an ED (WHO/IPCS, 2002) and as belonging to group 1 according to the proposed Danish Criteria for EDs (Hass et al. 2011). Suspected EDs and mixtures should not be considered in the workshop discussions.

Endocrine disrupting effects in humans: sensitive windows of development and late manifestations, Anna-Maria Andersson, Rigshospitalet, DK
Anna-Maria Andersson explained why EDs can/may pose problems to human health and be involved in the increasing incidence rates of male reproductive orders such as testicular cancer, reduced semen quality etc. She then briefly introduced the pathogenesis of testicular germ cell cancer and testicular dysgenesis syndrome (TDS) as well as the developmental windows in male reproduction. Hereafter the evidence for links between chemical exposure and male reproductive disorders were explored by referring to a number of published epidemiological studies. The importance of exposure estimates in relation to sensitive windows of development as well as the long time lag between exposure and manifestation of health outcome were emphasized. Also, she mentioned that effects of mixed exposures/complex exposure scenarios and sex dimorphic and non-monotonic effects may be difficult to ”catch” in human studies.

Non-monotonic dose responses: the role of study design & influence of study quality, Laura N. Vandenberg, UMass Amherst School of Public Health & Health Sciences, USA
Laura N. Vandenberg presented her extensive work on non-monotonic dose responses (NMDR) and explained that non-monotonic curves are common in medicine, pharmacology, nutrition and endocrinology, as well as for ED effects. Some of the mechanisms behind NMDR curves were introduced and included cytotoxicity, receptor competition and negative feedback loops. Different discussion points regarding NMDRs were explored such as the reproducibility and frequency of

NMDR curves, how study design influences the ability to detect NMDR curves, and whether NMDR curves at low doses are relevant to humans.

Review of the non-monotonic dose-responses of substances for human risk assessment GP/EFSA/SCER/2014/01, Clémence Varret, LIST, FR
Clémence Varret presented the NMDR review undertaken by ANSES, Karolinska and others contracted by EFSA and aiming at making an overview, according to the systematic review methodology, on the state of science on NMDR hypothesis. In task 1, previous reports were analysed and used for gathering information regarding NMDR. In task 2 extensive literature search and selection was performed, and in task 3 the selected literature underwent assessment of relevance and reliability, and data were extracted from the relevant and reliable studies. The main reason for exclusion of studies was the use of too few dose groups where the criteria used was at least 5 doses + 1 control. During the final task 4, the extracted data underwent a dose-response analysis, that compared non-monotonic models with a monotonic model, and six checkpoints were developed and tested for the evaluation of the evidence of NMDR. The final part of the presentation consisted of a summary of the results from the project, discussions regarding the methods applied, and finally a number of recommendations.

EDs and risk assessment -how?, Betty Hakkert, Bureau REACH, Centre for Safety of Substances and Products, NL
Betty Hakkert first talked about the different information requirements of some relevant regulatory frameworks such as REACH and whether the framework was hazard or risk based. The information requirements are not specifically designed to assess substances with respect to the WHO-definition/draft EU ED criteria and even if a chemical has been assessed as an ED multiple uncertainties come into play such as the limited coverage of ED-relevant exposure windows and endpoints, the power/dose response curves, and the effects or diseases that should be considered in health impact (HI) assessments and socio economic analyses (SEA). Next, the debate about thresholds or non-threshold for EDs was introduced and put in the context of REACH. Two non-threshold approaches, linearized and large assessment factor, for derivation of DMELs were explained.

4. Discussion Groups I: Uncertainties for EDs with regards to setting references doses for human health

Sofie Christiansen introduced the task and summarised the key points from the distributed background material, where the major uncertainties related to assessment of ED are evaluated to include:

- Lack of studies with exposure during sensitive windows
- Limited sensitivity of (many) regulatory testing methods with regards to relevant endpoints addressed and power/ robustness of study
- The potential for ED-induced non-monotonic dose-response (NMDR)
- The existence of a threshold

The suggested discussion topics and questions (shown in Appendix 2d) were:

1. Are some of the above mentioned uncertainties for EDs more important than others? Which ones and in which cases?
2. Are there any other uncertainties of relevance for setting references doses for human health?
3. How does current risk assessment of ED address the uncertainties?
4. What is the need for addressing uncertainties for EDs?

Outcome I

Concerning the major uncertainties specific for assessment of ED, there was general agreement in all groups that the two first ('Lack of studies with exposure during sensitive windows' and 'Limited sensitivity of many regulatory testing methods with regard to relevant endpoints addressed and power/robustness of study') are more important than the two latter ones (NMDRs and threshold). It was further agreed that many different adverse effects can be induced through one endocrine mode of action. Further, only few of these adverse effects are currently included as endpoints in standard guideline toxicity studies. Therefore, information about all the possible adverse effects induced by an observed mode of action is only very rarely available for a substance under evaluation, even if an ED mode of action and an adverse effect is observed. This introduces a substantial uncertainty around the observed adverse effect level. However, the groups also found that the potential for ED-induced NMDR and the uncertainty with regards to the existence of a threshold is of importance.

The groups agreed that these four uncertainties are the most important ones when it comes to assessment of EDs. However, some of the groups also discussed some additional uncertainties such as:

- sex differences and gene/ethnic variations (as seen for drugs)
- the late onset of some ED effects during e.g. senescence
- metabolic diseases, where toxicology studies use optimal food, whereas 'humans have a lifestyle'
- iodine deficiency in some studies

The groups generally found that the current risk assessment of EDs only address the specific uncertainties related to assessment of ED to a limited extend.

The groups generally found that there is an urgent need for addressing specific uncertainties related to assessment of ED due to the risk of irreversible and delayed effects.

5. Discussion groups II: The need to specifically address uncertainties for EDs when setting references doses for human health? How?

Ulla Hass introduced the task and summarised the key points from the distributed background material incl. the draft report from 2016:

- The approach used for setting reference doses should not be based on whether there exists a threshold for effects or not, because it is impossible to determine a threshold with reasonable certainty. Therefore a unified approach for all EDs is recommended.
- The two case studies illustrate that different approaches, i.e. DNEL, a $DNEL_{ED}$ with additional assessment factors covering uncertainties related to the ED assessment and a $DMEL_{ED}$ approach, lead to marked differences in the reference doses of procymidone and DEHP considered tolerable for humans.
- The two case studies also illustrate that a $DNEL_{ED}$ where relevant assessment factor covering uncertainties related to the ED assessment are used, a $DMEL_{ED}$ based on the 'Large Assessment Factor' approach as well as a $DMEL_{ED}$ obtained by linear extrapolation to 10^{-5} incidence are of similar magnitudes. Thus each of these three approaches can be considered as a unified approach for deriving tolerable reference levels for risk assessment of all EDs irrespective of a threshold assumption or not.

In addition, Ulla also presented a newly developed hypothetical case study 3 (shown in Appendix 2e), which illustrates other limitations than those in the two case studies in the draft report (e.g. lack of studies with exposure during sensitive windows).

The suggested discussion topics and questions (shown in Appendix 2d) were:

1. How to address specific uncertainties related to assessment of ED when deriving reference doses for EDs? You can include the case studies for this.
2. A threshold-based approach without or with additional factors for ED uncertainties or a non-threshold-based approach? Which of these methods is scientifically most relevant for deriving a reference level for an ED? Do you have alternative proposals?
3. Should there be a unified approach for all EDs irrespective of a threshold assumption or not? Or a case-by-case approach? Or a unified approach that can be modified case-by-case?
4. Which approach appears most feasible for regulatory work?

Outcome II

The groups generally recommended using a non-threshold approach as default to address uncertainties related to assessment of ED when deriving reference doses for EDs. This could be a rebuttable assumption, thus if strong evidence for threshold is/becomes available then a threshold approach could be used.

Some of the groups found that the point of departure may be based not only on adverse effect, but also on an earlier key event (KE) (e.g. Thyroid hormone insufficiency) in an adverse outcome

pathway (AOP), especially if the KE occurs at lower doses than the adverse effect. However, if the KE occurs, but the relevant adverse effects have been studied and not observed, a case-by-case evaluation would then be needed.

Two of the four groups recommended using linear extrapolation (to e.g. 10^{-6} incidence) using e.g. BMDL (Bench Mark Dose Low) as point of departure. One of these groups also recommended that an additional uncertainty factor should be included in cases where sensitive windows of exposure were not studied.

The two other groups recommended deriving a DMEL(Derived Minimum Effect Level)/reference dose using additional factors covering uncertainties related to assessment of ED. This was seen as a unified approach, as more transparent and leading to better promotion of new studies. Using BMDL as point of departure was the preferred approach as this takes the power of the data into consideration.

It was recognized that the uncertainties related to ED assessment can be divided into two groups, i.e. some where the uncertainty factor (UF) could be reduced if more data becomes available and others where there is a lower plausibility that UFs can be reduced even by the occurrence of more data. The uncertainties that may be reduced included e.g. lack of studies with exposure during (all) sensitive windows, where one of the groups evaluated that an UF of 10 or higher would be appropriate. The uncertainties considered to have lower plausibility to be reduced include insensitivity of models and limitation of test guidelines, the limited number of ED endpoints assessed the potential for late effects and the possibility for NMDR. The two groups recommended including as default an additional ED UF of 10-100.

Some of the groups (or participants) also gave some recommendations and ideas for the future such as:

- Elaborate on a guidance document to define the derivation of a threshold
- Explore whether a TTC (threshold of toxicological concern) approach for EDs could be applicable. For example, data coming from the pesticide or pharmaceutical field might be used to group compounds with e.g. estrogenic, anti-androgenic or thyroid mode of action. Following that, the lowest level where ED effects appear could be defined and used as a TTC for this specific ED MoA.
- Explore the IPCS (International Programme on Chemical Safety) probabilistic approach associating the magnitude of effects and the incidence and distribution of uncertainty factors to see if it is applicable for EDs and how.

6. Wrap up and conclusions

In plenary sessions, all discussion groups presented summaries of their deliberations for debate. The discussions were lively and constructive. In addition, there appeared to be a general consensus or a clear majority agreement on most of the main issues discussed.

Thus the following conclusions are drawn:

- The four uncertainties discussed, i.e. 1) Lack of studies with exposure during sensitive windows, 2) Limited sensitivity of many regulatory testing methods with regard to relevant endpoints addressed and power/robustness of study, 3) The occurrence of ED-induced NMDR and 4) The existence of thresholds, are the most important uncertainties related to ED assessment.
- The two first uncertainties are considered as more important than the two latter ones, but those are also of importance.
- The current risk assessment of EDs only addresses the uncertainties specific for ED assessment to a limited extend.
- There is an urgent need for addressing uncertainties related to ED assessment due to the risk of irreversible and delayed effects.
- It is recommended to use a non-threshold approach as default to address specific uncertainties related to assessment of ED when deriving reference doses for EDs. This could be a rebuttable assumption, thus if strong evidence for threshold is/becomes available then a threshold approach could be used.
- The participants were equally divided between recommending two approaches for the derivation of references levels, i.e. 1) linear extrapolation (to e.g. 10^{-6} incidence) or 2) derivation of a reference dose (DMEL) using additional factors covering specific uncertainties related to assessment of ED including also as default an additional ED UF of 10-100.
- Using BMDL as point of departure is recommended (when possible) as this takes the power of the data into consideration.