Centre on Endocrine Disrupters

Input for the REACH-review in 2013 on endocrine disrupters (tærskelværdi-projekt, j.nr. MST-621-00050)

Final report 21 March 2013

DANISH CENTRE ON ENDOCRINE DISRUPTERS

Ulla Hass, Sofie Christiansen, Marta Axelstad
Karin Dreisig Sørensen and Julie Boberg

Division of Toxicology and Risk Assessment, National Food Institute,
Technical University of Denmark

DTU Food
National Food Institute

DTU

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 with the focus on providing information requested for the preventive work of the regulatory authorities. The Centre is financed by the Ministry of the Environment and the scientific work programme is followed by an international scientific advisory board.

The overall scope of this project is to provide a science based input to the coming REACH review with regard to endocrine disruptors. In accordance with the terms of reference the scientific evaluation includes a review of the paper on low dose effects and non-monotonic dose responses of hormones and endocrine disrupting chemicals by Vandenberg et al. (2012) with regard to the level of evidence for non-monotonic dose responses (NMDR).

2. Background and aim

Endocrine disrupting substances are case-by-case covered by the authorisation scheme in REACH. If a substance is identified as an endocrine disruptor in accordance with Article 57 (f) an authorisation can only be given if adequate control of the risk can be demonstrated. However, before June 2013 the European Commission is obliged to review REACH with regard to endocrine disrupting substances. It is to be evaluated whether the application area for article 60.3 should be expanded to include endocrine disrupting substances in general, which means that authorisation can only be given if the socio-economic benefits “overrule” the risk and if there are no useful substitutes.

Only substances of particular concern are authorised via the socio-economic route, i.e. CMRs (Carcinogens, Mutagens, Reproductive toxicants) and substances of equivalent concern for which a threshold cannot be determined, PBTs (Persistent, Bioaccumulating, Toxic substances) or vPvBs (very Persistent and very Bioaccumulating substances) and substances of equivalent concern due to PBT or vPvB properties.

It is therefore relevant to examine if a threshold for endocrine disrupters (EDs) in general can be established with reasonable certainty based on current scientific knowledge on the mechanisms and modes of action for EDs - also taking into account current scientific knowledge on the possible so-called “low dose effects” and non-monotonic dose responses for EDs. If an assumption of threshold seems plausible, it is furthermore relevant to examine whether the currently available and used regulatory test methods are sufficiently sensitive for deriving a robust NOAEL (No Observed

Adverse Effect Level) or BMD (Bench Mark Dose) in relation to endocrine relevant adverse effects endpoints.

The so-called “low dose effects” and non-monotonic dose-response for EDs have been under discussion for several years. There are different evaluations among researchers in the field on the quality and the extent of evidence in relation to these topics. The discussion as well as the number of research papers has increased during recent years and in the beginning of 2012 a large review on low dose effects and non-monotonic dose responses with focus on human health effects was published in the journal ”Endocrine Reviews” (Vandenberg et al. 2012). This review was an important background paper for discussions at an EFSA colloquium on low-dose effects, June 12-15 2012, and at a joint EU/US workshop on low-dose effects and non-monotonic dose-response for endocrine active chemicals, September 11-13 2012, in Berlin. Some of the authors of this report participated in both workshops and also contributed to the planning of the joint EU/US workshop.

It is further relevant to evaluate whether EDs in general give rise to particular concern as the PBTs (persistent, bioaccumulating and toxic substances) or vPvBs (very persistent and very bioaccumulating substances) because this may or may not further support the management of EDs in accordance with REACH art. 60.3

The aim of this report is, from a scientific point of view, to discuss the topics expected to be relevant for the REACH review on EDs, i.e.:

- Thresholds or non-threshold assumption for ED effects
- Considerations concerning non-monotonic dose-response (NMDR)
- Uncertainties of the currently regulatory test methods with regard to determination of possible thresholds for EDs
- Whether there is particular concern for EDs

The REACH review on endocrine disruptors considers both human health and the environment. The focus in this report is, however, restricted to human health as both the review from Vandenberg et al. (2012) and the workshops mentioned above focused on human health. The report does not specifically discuss the so-called ‘low dose effects’ of EDs. However, this topic is indirectly covered in the sections on thresholds, NMDR and uncertainties.

3. Threshold for EDs

One of the key concepts in toxicology and risk assessment is the dose-threshold, which implies that chemicals can only cause (non-cancer) effects above a certain dose level (Slob 1999).

Historically, dose-response assessments have been conducted differently for cancer and non-cancer effects. For carcinogenic effects, it has earlier on, generally been assumed that there is no dose-threshold for effect, and dose-response assessments have focused on quantifying risk at low doses. However, during the last decade considerations of mode of action have highlighted that there may be a need to differentiate the approaches for genotoxic and non-genotoxic carcinogens. For non-

cancer effects, a dose-threshold has been assumed and no observed adverse effect levels (NOAEL) or benchmark doses (BMD) have been used as a point of departure for deriving levels below which effects are not expected to occur or are extremely unlikely in an exposed population (Abt et al. 2010). Evidence of mutagenicity is used to differentiate between genotoxic and non-genotoxic carcinogens and for genotoxic carcinogens the assumption of non-threshold effects precludes the establishment of a Derived No Effect Level (DNEL) (ECHA 2007).

It is currently discussed whether an assumption of non-threshold may also be valid for chemicals with endocrine mode of action. Also, the existence of non-monotonic dose response for EDs is currently discussed and this issue is dealt with in section 4. In this section, focus is on threshold or non-threshold at the low end of the dose-response curve irrespective of whether this is a monotonic or non-monotonic dose-response.

3.1 What is a threshold?

The threshold for effect may be defined in different ways, which may be relevant to the arguments for or against a threshold for EDCs. Slob (1999) provided three different definitions:

1. Mathematical definition: the dose below which the response is zero and above which it is non-zero.
2. Biological definition: the dose below which the organism does not suffer from any (adverse) effect.
3. Experimental definition: the dose below which no effects are observed.

The presence or absence of a threshold using the mathematical definition can never be experimentally proven or ruled out (Kortenkamp et al. 2012; Sheehan et al. 1999; Slob 1999). All methods for measuring effects have a limit of detection below which effects cannot be observed, which will obscure thresholds, if they exist (Kortenkamp et al. 2012). Also, to generate an exact dose-response curve would require an infinite number of doses and infinitely precise measures (Slob 1999). Additional complicating factors are related to normal biological variation and the limited power that is available with the size of dose groups normally used in toxicity testing (see section 5).

As it is not possible to experimentally prove the existence or absence of a threshold, 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.

3.2 Biological thresholds for EDs?

The reviewed literature provides arguments both for and against assuming a threshold for EDs. The general argument for assuming no biological threshold for EDCs is that because low doses of endogenous hormones are present and fluctuating, small additions (or subtractions) to their actions will have a significant impact (Zoeller et al. 2012). This “additivity-to-background” argument has also been made to defend a no-threshold-approach for genotoxic carcinogens (Slob 1999).

A central tenet of endocrinology is that hormones exert their physiological actions through receptors (Zoeller et al. 2012). This has several implications. First, hormone action is saturable in terms of both ligand-binding and effect. Moreover, the maximum effect of the hormone typically occurs at ligand concentrations well below those that result in receptor saturation. These observations impose several consequences for the expected shape of dose-response curves induced by hormones and by chemicals that interfere with hormone actions. First, the curves are never linear, although they may contain linear portions. Instead, they tend to be sigmoidal in shape but may depart from this basic form, as in the case of non-monotonic dose-responses (see section 4). It is the nature of sigmoidal dose responses that an equivalent change in hormone level (or action) at both the very low end and the high end of the curve will have a small effect, whereas at the mid part of the curve the effect is proportionally greater. Furthermore, because low doses of endogenous hormones are present and fluctuating, small additions (or subtractions) to their actions will have an impact (Zoeller et al. 2012, Kortenkamp et al. 2012, Vandenberg et al. 2012). If no homeostatic control occurs, this implies that endocrine disrupting chemicals can exhibit activity in a threshold-independent fashion. On the contrary, if homeostatic control occurs like protein binding of hormones or chemicals, buffering of hormone levels via feed-back mechanisms etc., a threshold of EDCs could be expected. It is important to note, that the presence of thresholds can never be confirmed or rejected by experimental data as indicated in the previous section.

The arguments made in support of a biological threshold for EDs are mainly that, as stated for example by Blair et al. (2001): "…a threshold could be expected if there is no endogenous hormone, if the endogenous hormone induces no adverse effect, or if there is effective homeostatic control". This may be of relevance for the function of hormones in adults, where there may be an effective homeostatic control. However, during development hormones have a very important organizing role in relation to the sexual dimorphic development of the reproductive system and the brain, the general development of the brain (e.g. thyroid hormones) and also for foetal programing of the endocrine system (Kortenkamp et al. 2012). Thus, during development, endogenous hormones are present and "wrong" levels of endogenous hormones may induce adverse effects. In humans, hormonal regulations and feedback interactions develop during foetal life and for the hypothalamus-pituitary axes this system is functional after 20 weeks of gestation (Siler-Khordr 1998). The steroidogenesis of androgens and oestrogens, however, occurs earlier and organizes the sexual dimorphic development of the reproductive system during 7-10 weeks of gestation (Moore 1983). This implies that during sensitive windows of prenatal development there is no effective homeostatic control, because the buffering of hormone levels via feed-back mechanisms is not developed yet. In conclusion, the above mentioned arguments for a biological threshold are not relevant during sexual development.

Conolly and Lutz (2004) state that the first interaction of a toxic agent with its primary biological target molecule is likely to have no threshold but imply that the complexity of a biological system makes non-threshold dose-response curves unlikely for many "higher" endpoints, such as behaviour, reproduction, organ weights and growth. In relation to effects of EDs, this would mean that although there is not necessarily a threshold for the primary biological action, the integration of chemical influences on several pathways of importance to development of a certain "higher" type of effect may lead to threshold-like response patterns. This may be the case when e.g. opposing effects

occur at different dose levels due to different specific mechanisms of action occurring, and the influence of one direction of effect overrides the opposing effect caused by another mechanism of action of the same compound. This point is related to the presence of non-monotonous dose-response curves discussed in section 4.

3.3 Do toxicological data indicate threshold or no-threshold?

Probably, because of its suggestive wording, the term NOAEL may be taken to imply an absence of effects, as expressed for example by Ashby et al. (2004): “If the statistical methods used are appropriate, the absence of significance should indicate the absence of a chemically induced effect” (as described in Scholze and Kortenkamp 2007). The term NOAEL is, however, not the same as a threshold, because a NOAEL as signalled by the O for “observed” is the dose level where no effects are observed and thus depend on the sensitivity of the methods for assessing the effects (see section 5).

To examine the threshold assumption for endocrine active chemicals with non-genotoxic endpoints, Sheehan (2006) examined dose-response data from the literature and the hypothesis was that no threshold exists when a substance acts through the same mechanism as endogenous oestradiol, i.e. has oestrogenic activity. The analysis was accomplished by fitting the dose-response data to a modified Michaelis-Menten equation, which has no threshold term. Thirty-one data sets from studies on 9 different substances were evaluated. The chemicals used included natural (oestradiol) and synthetic (e.g., diethylstilbestrol and conjugated oestrogens) hormones as well as several synthetic endocrine disruptors (e.g. dioxin, polychlorobiphenyls). Twenty-six of the data sets fitted the modified Michaelis-Menten equation with high multiple correlation coefficients ($r>0.90$). The endpoints included both physiological (e.g. plasma prolactin levels and cell proliferation), and adverse responses (e.g. presence of vaginal threads and adenomas). Sheehan (2006) state that it is not surprising to observe a good fit to the modified Michaelis-Menten equation without a threshold term for many of the examined dose-response data, since endocrine disruptors are capable of acting through receptor binding initiating a rate-limiting step that does not exhibit a threshold.

In the US NTP low dose peer review report (Melnick et al. 2002) it was evaluated that for finasteride, which acts as a 5α–reductase inhibitor, the dose-response curve for reduction in male anogenital distance (linear) was different from that for increased hypospadias (threshold-appearing). Also, exposure of pregnant rats to vinclozolin at six doses ranging from 3.125 to 100 mg/kg/day resulted in reduced anogenital distance and increased incidences of areolas and nipple retention in male offspring (Melnick et al.2002). For these effects, the dose–response curves appeared linear to the lowest dose tested. Reproductive tract malformations and reduced ejaculated sperm numbers were observed only at the two highest doses. These observations indicate that the shape of the dose–response curves may be low-dose linear for the effects on anogenital distance and nipple retention. In relation to hypospadias, the threshold-appearing response might indicate a threshold, or alternatively it may reflect the limited sensitivity for detecting rare quantal effects (see section 5). Thus, based on these data it is evaluated as uncertain whether there is a threshold or not for hypospadias.

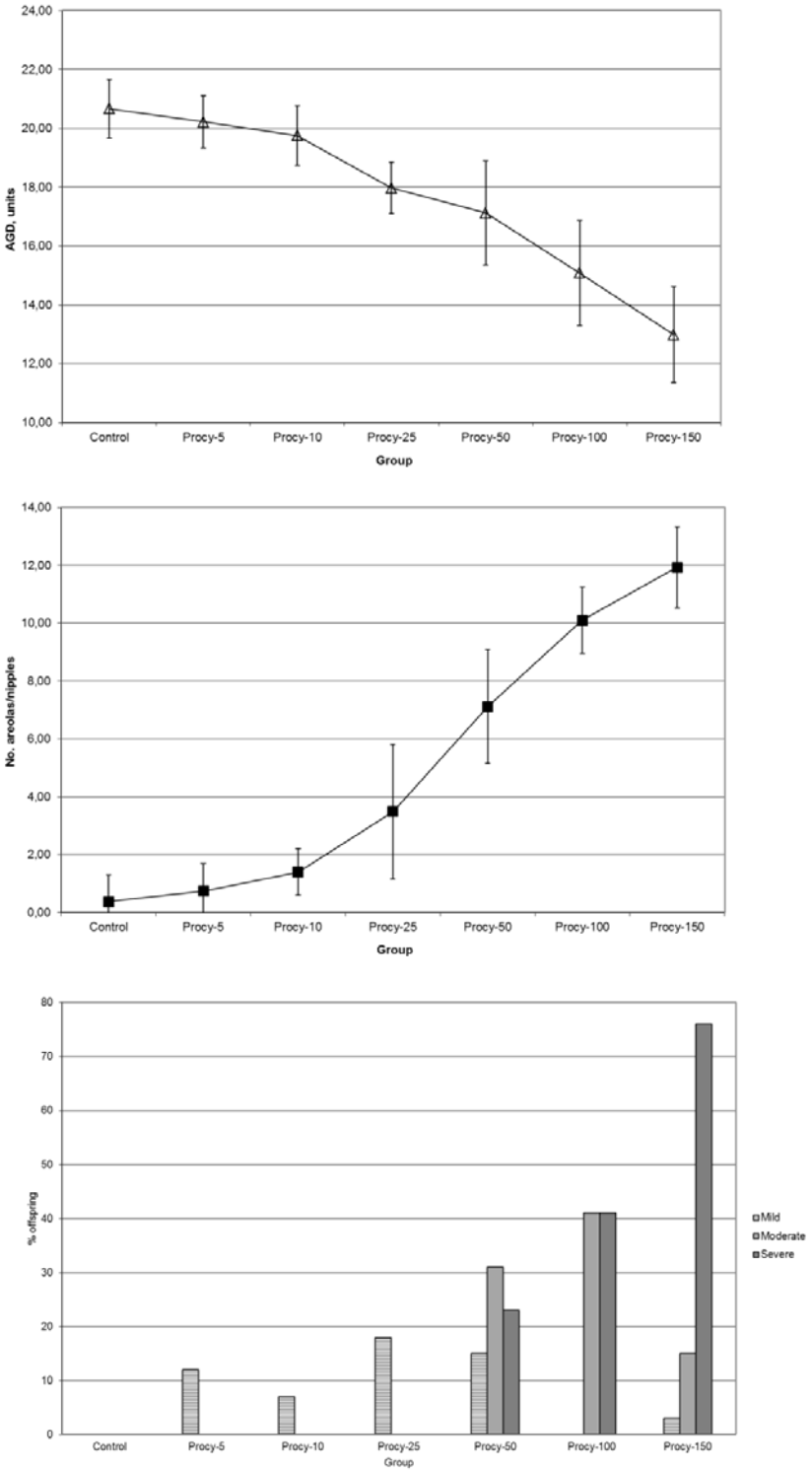

Figure 1. Dose-related decrease in anogenital distance day 1 (top), increase in nipple retention day 13 (middle) and dysgenesis of external sex organs day 16 in male offspring exposed perinatally to procymidone. Results shown for anogenital distance and nipple retention are mean + SD. For genital dysgenesis, the % offspring affected is shown. Based on Hass et al. 2007 and Metzdorff et al. 2007. the dose response curves appeared non-thresholded for AGD and nipple retention, but threshold-like for genital dysgenesis

Similar results as described above have been found in studies of the effects of the AR-antagonists flutamide, vinclozolin and procymidone on male sexual development (Hass et al. 2007, Metzdorff et al. 2007). At the doses studied, the dose-response curves appeared non-thresholded for AGD and nipple retention, but threshold-like for genital dysgenesis (hypospadias), see figure 1 where the results for procymidone are shown. These dose-response data were all fitted to nonlinear sigmoidal models. The arguments for not using a threshold model parameter were to keep the modelling simple and robust but also more importantly, that none of the data analysed justified inclusion of a threshold parameter (Martin Scholze, pers. com). Based on this, it is not possible to conclude whether the dose-response may have a threshold or not, i.e. it is simply unknown.

The problem of methodological limitations has made it difficult to reach conclusions about additivity to the background concerning morphological effects because it has not been possible to design experiments that have sufficient sensitivity to determine whether very small doses of a compound have any effect (Boobis et al. 2009). Gene expression has been analyzed in foetal rat testis exposed transplacentally to three different compounds with estrogenic activity - ethinyl oestradiol, genistein, and bisphenol A. Doses for each compound spanned five or six orders of magnitude, starting from a dosage known to have pharmacological activity, down to very low dose levels. All three compounds had effects on gene expression at the higher dose levels, and there were still some effects on gene expression at doses lower than those that had morphological effects; however, at the lowest dose levels of these compounds there were no significant changes from the control in relation to gene expression (Naciff et al., 2005b as described in Boobis et al. 2009). According to Boobis et al. (2009), this result strongly suggests a threshold for activity of oestrogens on gene expression during development. However, we find that similarly as for morphological effects such as anogenital distance it is not possible to determine whether the effect on gene expression effect actually has a threshold or the results just showed a threshold-like dose-response, because the variability in the measurement overwhelmed the ability to detect very small changes in a reasonable number of animals. Therefore, based on these data, we conclude that effects on gene expression appear to be more sensitive endpoints than morphological effects, but there are uncertainties with regard to threshold for both types of effects.

3.4 Conclusions

The presence of thresholds 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 because low doses of endogenous hormones are present and fluctuating, 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 of hormones during development at a time point where the homeostatic control is not effective or not developed yet. Also, experimental data indicate non-thresholded dose-response 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 appears more valid for the effects of EDs during development than an assumption of a threshold.

Regardless of ED mode of action, it is uncertain whether or not there is a threshold for EDs. For EDs, where the MoA (Mode of Action) directly involve the receptor, the interaction with the receptor is likely to have no threshold. For EDs affecting the hormone levels, the response pattern may appear threshold-like, because multiple pathways converge before seeing the final response and some of these pathways may have a threshold.

Irrespective of threshold or non-threshold, the dose response curves of EDs 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, however, have a "threshold-like" appearance, but a threshold cannot be inferred from the shape of the dose-response curves. However, a benchmark approach may be used for estimating a human exposure level with very low risk.

4. Non-monotonic dose-response (NMDR) for EDs

In the fields of toxicology and human health-risk assessment there is currently much debate about the shape of the dose-response curve. By a monotonic dose-response, the observed effects may be linear or non-linear, but the slope does not change sign. In contrast, a dose-response curve is non-monotonic when the slope of the curve changes sign somewhere within the range of doses examined (Vandenberg et al. 2012). NMDRs are often U-shaped (with maximal responses of the measured endpoint observed at low and high doses) or inverted U-shaped (with maximal responses observed at intermediate doses). Numerous toxicological studies show a NMDR curve with either a decrease in the response below control at low dose followed by an increase at high dose (U- or J-shaped) or *vice versa* (inverted U- or β-shape) (Conolly & Lutz, 2012).

4.1 ED mechanisms for NMDR

There are several mechanisms that illustrate how hormones and EDs may cause NMDRs. These mechanisms include cytotoxicity, cell and tissue-specific receptors and cofactors, receptor selectivity, receptor down-regulation and desensitization, receptor competition, and endocrine negative feedback loops (Vandenberg et al. 2012). In the following, these mechanisms are briefly described based on Vandenberg et al. (2012) with main focus on those mechanisms where the NMDR can be related to functions of the endocrine system. For further details and specific references, see Vandenberg et al. 2012.

Cytotoxicity

The simplest mechanism for NMDR derives from the observation that hormones can be acutely toxic at high doses yet alter biological endpoints at lower doses. As experimental results clearly

indicate that the effects of for example oestradiol at high doses are due to toxicity via non-ER-mediated mechanisms we do not consider such NMDRs as evidence for endocrine related NMDR.

Cell- and tissue-specific receptors and cofactors

Some NMDRs may be due to the combination of two or more monotonic responses that overlap, affecting a common endpoint in opposite ways via different pathways. For example, oestrogens have been shown to induce cell proliferation and inhibit apoptosis in several cell populations, but inhibit proliferation and induce apoptosis in others, with the combined effect being an inverted U-shaped curve for cell number. In many cases, it is difficult to evaluate whether observed NMDR for an ED endpoint is due to two or more monotonic endocrine related responses as mechanistic data is scarce. In the absence of mechanistic data, it is proposed to assume that such NMDRs are considered as evidence for endocrine related NMDR until proven otherwise.

Receptor selectivity

NMDRs can occur because of differences in receptor affinity, and thus the selectivity of the response, at low *vs.* high doses. Thus, the effects seen at high doses may be due to action via the binding of multiple receptors in contrast to the effects of low doses, which may be caused by action via only a single receptor or receptor family. If NMDR is seen due to such action this is evaluated as clearly related to the function of the endocrine system.

Receptor down-regulation and desensitization

When hormones bind to nuclear receptors, the outcome is a change in the transcription of target genes. After this, the reaction must cease; *i.e.* the bound receptor must be inactivated in some way. Nuclear hormone receptors can be degraded and as hormone levels rise, the number of receptors being inactivated and degraded also rises, and the number of new receptors being produced may not maintain the pace of the degradation.

There can also be receptor desensitization, where a decrease in response to a hormone is due to biochemical inactivation of a receptor. Desensitization typically occurs when repeated or continuous exposure to the ligand occurs. Receptor desensitization has been observed for a range of hormones including glucagon, FSH, human chorionic gonadotropin, and prostaglandins. Receptor down-regulation and desensitization may occur in the same cells for the same receptor, and therefore, both can play a role in the production of NMDRs. In such cases, NMDR is related to the function of the endocrine system.

Receptor competition

Mathematical modelling studies suggest that endogenous hormones and EDs establishes a natural environment to foster NMDRs. Using mathematical models, Kohn and Melnick (as described in Vandenberg et al. 2012) proposed that when ED exposures occur in the presence of endogenous hormone and unoccupied hormone receptors, some unoccupied receptors become bound with the ED, leading to an increase in biological response. At low concentrations, both the endogenous hormone and the ED bind to receptors and activate this response, but at high doses, the ED may outcompete the natural ligand. The model predicts that inverted U-shaped curves may occur and

would be abolished only if the concentration of natural hormone were raised such that all receptors were bound.

Endocrine negative feedback loops

In several cases, the control of hormone synthesis is regulated by a series of positive- and negative feedback loops. Studies indicate that these negative feedback loops could produce NMDRs when the duration of hormone administration is changed. For example, short exposures of oestrogen induce proliferation in the uterus and pituitary, but longer hormone regimens inhibit cell proliferation. Thus, the exposure to a single hormone concentration – or an ED - may stimulate an endpoint until negative feedback loops are induced and the stimulation ends. As endocrine feed-back loops are not developed before the late part of foetal life (e.g. around week 20), NMDRs due to this function of the endocrine system is not to be expected during foetal life.

4.2 Human evidence

The existence of NMDRs for endocrine active drugs has been recognized and used in human clinical practice for many years (Vandenberg et al. 2012, Juul 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. The recognition of this dual dose-response range for tamoxifen led to the definition of the term selective oestrogen response modulator or SERM, activity. These observations defined three separate dose-response ranges for tamoxifen in human clinical use. The lowest dose-response range, the range of flare, stimulated breast cancer growth and symptoms in some patients with hormone-dependent cancer. The next higher dose-response range is the therapeutic range where tamoxifen inhibits oestrogen-dependent tumour growth and the highest dose range causes acute toxicity by the SERM (Vandenberg et al. 2012).

4.3 NMDR *in vitro*

U-shaped or inverted U-shaped dose-response curves are often observed in *in vitro* studies, which is usually due to various mechanisms of actions involved for the same chemical. The typical situation is a low concentration effect due to the primary mechanism tested and cytotoxicity (i.e. cell death) at higher concentrations. However, other cases exist in which two or more mechanisms of action that do not include cytotoxicity are into play.

One example is seen for the antagonistic effect of hydroxyflutamide on the androgen receptor. Hydroxyflutamide is the hydroxyl-metabolite of flutamide, which is a drug used for treatment of prostate cancer. A non-monotonic dose-response curve for androgen-receptor-mediated gene transcription by hydroxyflutamide was seen in HepG2 human hepatoma cells. This effect is a general phenomenon happening in several androgen receptor reporter gene assays. Low

hydroxyflutamide concentrations partially antagonized the effect of dihydroxytestosterone, while agonistic activity was observed with a further increase in hydroxyflutamide concentration. The biphasic dose-response curve was explained by the hypothesis that only the receptor dimer that carry two DHT or two hydroxyflutamide ligands, but not mixed-ligand dimers, are transcriptionally active (Maness et al., 1998). Such a dose-response relationship for antagonism of the androgen receptor is found for vinclozolin and progesterone as well. The mechanism of progesterone is believed to be comparable to that of hydroxyflutamide and this is illustrated in Figure 2.

Another example is the antagonistic action of two adenosine receptor subtypes that regulate adenylate cyclase in opposite directions, given appropriate differences in ligand affinity and in efficacy of signal transduction, resulting in a clearly biphasic dose-response curve (Ebersolt et al., 1983). Many other examples from *in vitro* and *in vivo* studies can be presented and explained as evidence for the existence of NMDRs.

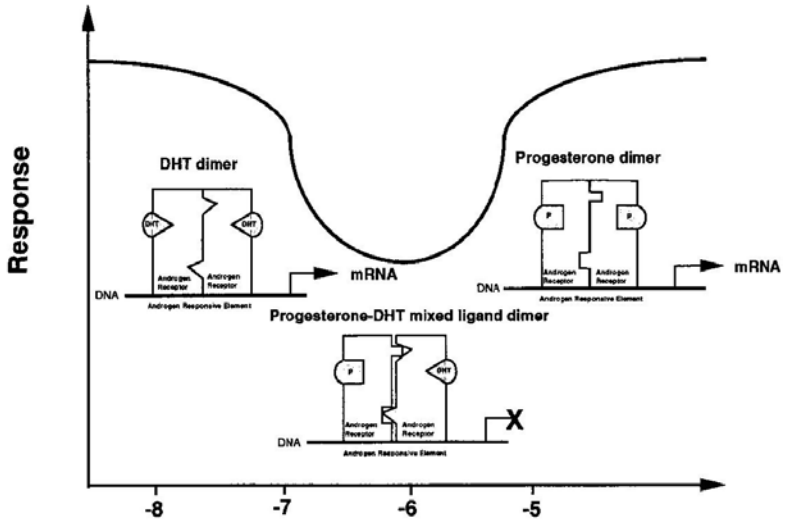

Figure 2: Mixed-ligand hypothesis: Formation of ligand dimers and the resultant response for low, medium, and high concentrations of progesterone in combination with an inducing concentration of dihydrotestosterone. With low concentrations of progesterone DHT–DHT dimers are more likely to form and induce a response. As the concentration of progesterone increases, mixed progesterone–DHT ligand dimers form, which block androgen receptor activity. At high progesterone concentrations progesterone–progesterone dimers are more likely to form and induce a response. From Maness et al.(1998).

4.3.1 Critical review of in vitro studies included in Vandenberg et al. 2012

In the review by Vandenberg et al., (2012) a comprehensive table summarizing results of *in vitro* studies giving rise to NMDR is presented. The definition of a NMDR used in this review is based upon the mathematical definition of non-monotonicity: that the slope of the dose-response curve changes sign from positive to negative or *vice versa* at some point along the range of doses examined. There are several adequate studies in the table that add confirmation for the hypothesis that EDs are capable of eliciting NMDR *in vitro* including Jeng et al., (2009), Boettcher et al., (2011), and Almstrup et al., (2002) (ref. 744, 719 and 730 in the Vandenberg review).

However, there are several important points worth emphasizing regarding the criteria used for including many of the *in vitro* studies to a list of studies showing NMDR. The broad definition used to define NMDRs does not seem to distinguish between the mechanisms that underlie the curve i.e.

the definition also allows for inclusion of studies with inverted U-shaped dose-response curves, which are the result of cytotoxicity at high concentrations. Thus, there are some studies, which have shown cytotoxicity at the highest concentrations and even discuss its importance to the shape of the biphasic curve (e.g. Asp et al., (2010) (ref. 754 in the Vandenberg review), and Alm et al., (2008)). In relation to *in vitro* investigations, inverted U-shaped curves caused by general toxicity towards the cells should not be regarded as "true" NMDRs, as this is merely a reflection of the concentration level of the compound. If the definition is applied as suggested by Vandenberg et al., most compounds will give rise to NMDRs. Thus the definition used in the review by Vandenberg et al. is in our view too broad to be applied on *in vitro* studies.

Another important point to mention is the testing of hypotheses and the associated statistics. Many of the studies used ANOVA followed by Dunnett's post hoc test to test for significant differences between the control and test concentrations. Yet, to demonstrate a NMDR according to the definition, a different kind of statistics e.g. testing for significant positive/negative slopes on either side of the curve peak to determine whether there is a real shift in the slope of the curve, will have to be applied. In other words, to significantly demonstrate a NMDR, a completely different approach in hypothesis testing and statistics will have to be used. Somjen et al., (1998) (ref. 721 in the Vandenberg review) describe the creatine kinase specific activity in vascular smooth muscle cells as a result of increased concentrations of ethinyl oestradiol. The curve first shows a slight decrease followed by an increase in activity and then a minor drop that might be due to cytotoxicity. The first point may be a coincidence and due to simple variation, since no statistical significance was found. Similarly, no evidence has been provided that there was a real decrease in the curve at the last point, since the statistics said nothing about the difference between the 10 nM and 100 nM concentrations.

In contrast, Leung et al., (2008) (ref. 728 in the Vandenberg review), compared all concentrations in the dose-response curve showing the IGF-1 expression as a result of growth hormone exposure using a Student–Newman–Keuls test. However, there was no significant difference between the three highest concentrations (i.e. 10, 100, and 1000 ng/mL). The change in slope was not significant, but only due to random variation. So even though the statistic that was used here is better suited for the detection of biphasic curves, it could not be proven that this was in fact a NMDR.

Lastly, NMDR may be the result of several mechanisms coming into play. Chemical mixtures can consist of substances that possess different modes of action, and can therefore interfere with the *in vitro* assay in many different ways. Thus, studies investigating mixtures are not very suitable for evaluating the existence of NMDR *in vitro*. Again, a too broad definition was in our opinion used when including mixture studies that showed biphasic dose-response curves *in vitro*. Mixture studies listed in the Vandenberg-review as showing NMDRs *in vitro* include Campagna et al., (2007) and Ohlsson et al., (2010) (ref. 757 and 750 in the Vandenberg review).

Figure 3 is a diagram showing the number of examples from the Vandenberg review describing NMDR *in vitro* (n=80 totally) allocated into four groups based on an evaluation according to our definition, i.e. NMDR where cytotoxicity cannot explain the change in curve slope (blue), NMDR

where cytotoxicity is or might be the cause of the change in the slope of the curve (green), NMDR that for some of the above reasons may or may not be evidence for NMDR (red), and no evidence for NMDR of EDs (purple). Almost half of the examples (45%) could not according to our definition be regarded as showing a true non-monotonic dose-response, as the NMDR was evaluated as due to cytotoxicity. Furthermore, some examples were evaluated as "false NMDR", because of e.g. testing of mixtures or limitations in the study design. The remaining examples were evaluated to either show evidence for NMDR (16%) or a dose-response that may or may not be due to NMDR of EDs (17%). More details can be found in Appendix 1.

In conclusion, the Vandenberg et al. review gives several good examples showing the existence of NMDRs *in vitro*. Thus, there are well-conducted studies showing biphasic curve patterns, which are supported by possible explanatory models. However, our critical evaluation based on the use of a less broad definition of NMDR leads to fewer cases than those included in the Vandenberg et al. review.

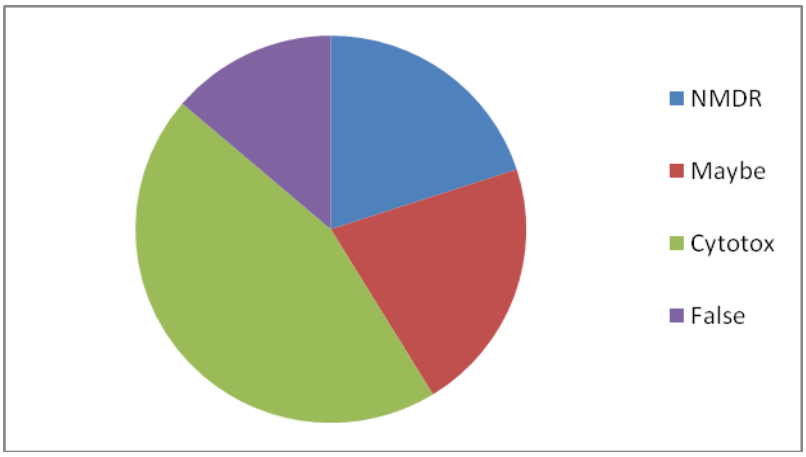

Figure 3: Pie chart showing the studies from the Vandenberg-review described as showing NMDR *in vitro* allocated into four groups. See text for further explanation.

4.4 NMDR *in vivo*

4.4.1 Critical review of in vivo studies included in Vandenberg et al. 2012

The papers selected for a critical evaluation included 34 from table 7 in the Vandenberg review. The papers selected were the majority of the studies related to human toxicity, with main focus on studies where the effects were regarded as sufficiently severe in relation to the definition of adverse effects.

The evaluations were based on a weight of evidence assessment and considered many aspects incl. the number of animals per group, the number of dose levels, the statistical significance of the effects and the plausibility for NMDR based on mode of action consideration in the paper. The evaluation did not include weight of evidence across papers for a specific chemical. Based on this, the evidence in the papers was allocated into one of 3 groups according to the level of evidence, i.e.:

Group1: Clear evidence for NMDR

- the effects found can be regarded as adverse effects
- a sufficient number of animals per group. This was not a fixed number as this depend on the power for detection of the specific effects
- statistically significant effect(s) at the peak (in case of inverted U-shaped dose-response) and also a high plausibility of a significant difference between the effect at the peak and the effect at higher dose(s)
- plausible endocrine MoA(s) behind the observed NMDR

Group 2: Some evidence for NMDR

Mainly similar evidence as for group 1, but where there were limitations for some parts of the evidence needed for group 1.

Group 3: Poor or no evidence for NMDR

- insufficient number of animals leading to a the high probability for false-positive and false-negative findings
- lack of statistical significance
- the apparent NMDR was evaluated as due to general toxicity and thus not related to an endocrine MoA
- use of an animal model deprived of the natural hormone and where the NMDR was evaluated as due to normalization of the function followed by toxicity due to too high hormone level.

Figure 4 is a diagram showing the number of examples evaluated (n=34 totally) grouped into the three groups according to the level of evidence for NMDR based on our definition. The majority of the examples were evaluated to give some evidence for NMDR (n=22) and 5 studies showed clear evidence. Poor or no evidence for NMDR was concluded for 7 of the studies. More details can be found in Appendix 2.

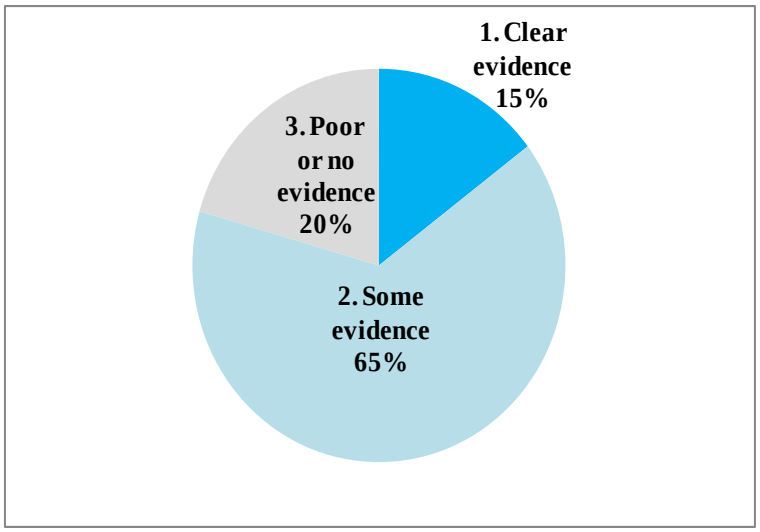

Figure 4: Pie chart showing 34 examples from the Vandenberg-review described as showing NMDR *in vivo* grouped into three groups. See text for further explanation.

Our evaluation shows that there was clear experimental evidence for endocrine induced NMDR in a limited number of studies (5 studies), but also some evidence for NMDR in the majority of the studies, i.e. 22 studies. For the latter studies more experimental data are needed to evaluate whether the observed NMDRs were actually real findings related to endocrine mode of action. However, we also found poor or no evidence for some studies. This may reflect that a very broad definition was used to define NMDR in the Vandenberg paper. For example, it seems that studies with inverted U-shaped dose-response curves, which are most likely the result of general toxicity at high concentrations were are also included. Also, a few studies without any statistical significance were included. Thus the broad definition apparently used by Vandenberg et al. was too broad in our view to be correctly applied on *in vivo* studies. Nevertheless, our evaluation indicates that for the majority of the studies evaluated by Vandenberg et al. there was clear or some evidence for NMDR.

In the following sections some of the studies in the Vandenberg-review as well some published results from our own studies of endocrine disrupters that are not included in the Vandenberg paper are described and evaluated in relation to NMDR.

4.4.1.1. NMDR and reproductive organ weights

For reproductive organ weights, several cases of NMDR have been described. These curves could either be associated with differences in androgen action as described previously or could reflect how effects in the target organs are interrelated and cause changes in organ weights in one direction at low doses, and another effect at higher doses. This happens when another action of the compound appears and affects the organ in the opposite direction. For example, testis weight can be affected by chemically induced changes such as fluid accumulation or impaired proliferation/differentiation, and these changes will likely have opposing effects on testis weight. If these changes appear at different doses, it may be speculated that this could result in NMDR curves for testis weight.

- One example of NMDR has been found for the effects of procymidone on testis weight in two different studies. The first study showed no change in testis weight at 5 mg/kg bw/day, a statistically significant increase at 10 mg/kg bw/day, no change at 25, 50 and 100 mg/kg bw/day, and a statistically significant decrease at 150 mg/kg bw/day (Metzdorff et al., 2007). In another study from the same laboratory, a statistically significant increase in testis weights was observed in animals exposed to the lowest dose of 12.5 mg/kg bw/day of procymidone but not at 50 mg/kg bw/day (Jacobsen et al., 2012). As body weight was used as a covariate in these studies, these changes were not caused by differences in body weight, but could rather reflect NMDR due to endocrine disrupting effects.

Prostate weight has also been shown to be affected by estrogenic compounds in a non-monotonous manner.

- Exposure of neonatal male rats to oestradiol benzoate resulted in increased prostate weights at low doses (0.15 ug/kg bw) and decreased prostate weight at high doses (1500 and 15000 ug/kg bw) when examined at PND 35 (Putz et al. 2001). In adulthood, a comparable pattern of effects was seen, though only the weight reductions at high doses were statistically

significant. The Putz et al. (2001) study was included in the Vandenberg paper (ref. 780), and assessed by us as belonging to group 2.

- In mice exposed to diethylstilbestrol (DES) during gestation, low doses (0.02, 0.2 or 2 ug/kg bw) resulted in increased prostate weights and high dose exposure (200 ug/kg bw) resulted in reduced prostate weights in adulthood (Vom Saal et al. 1997). Likewise, low levels of 17β- oestradiol increased prostate weights in adult mice exposed in utero, whereas prostate weights were unchanged at higher doses. Prostatic androgen receptor expression was increased at low levels of 17β-estradiol compared to controls. The Vom Saal et al. (1997) study was included in the Vandenberg paper (ref. 689), and assessed by us as belonging to group 1.
- In androgen-responsive reporter mice exposed to hexachlorbenzene during gestation, lactation and prepuberty, an increase in prostate weight and androgenic activity was seen at low doses, but not at high doses (Ralph et al., 2003). With continued exposure to 8 weeks, a decreased androgenic activity was seen at high doses. Low doses also increased epididymis and testis weights at 4 weeks and induced early puberty, while high doses showed no change of epididymis or testis weights. (Ralph et al., 2003). The Ralph et al. (2003) study was only included in the Vandenberg paper (ref. 755) in table 6 (Examples of NMDRCs in cell culture experiments) and therefore the in vivo results from this study were not assessed by us in relation to our grouping of the in vivo examples.

Early onset of puberty due to increased gonadotropin levels or altered sensitivity to androgens (e.g. increased androgen receptor expression) could be the cause of increased reproductive organ weights at low doses. At high doses, androgen receptor levels are down regulated and it is suggested that these opposing high- and low-doses effects are due to different modes of action appearing at different dose levels. In contrast, the study on hexachlorbenzene (Ralph et al., 2003) indicates that there may also be cases when a non-monotonous response is caused by the same primary effect/mode of action (androgen receptor interaction), and that the opposing responses are due to the non-monotonicity of dose-response curves for partial agonists.

4.4.1.2. NMDR and timing of puberty and nipple retention in male offspring

Examples of NMDR have also been described for DEHP for two different endpoints, i.e.: preputial separation and nipple retention (Ge et al. 2007, Christiansen et al. 2010).

The Ge et al. (2007) study was included in the Vandenberg paper (ref. 789), and assesses by us as belonging to group 1. Male Long-Evans rat pups were chronically subjected to low or high doses of DEHP, with the androgen-driven process of preputial separation used as an index of pubertal timing. The results are averages from 2 experiments. Rats were treated with 0, 10, 500, or 750 mg/kg body weight DEHP for 28 days starting at day 21 postpartum. The average age at which the animals completed preputial separation was recorded in each group. The age of preputial separation was 41.5 ± 0.1 days postpartum in controls (vehicle). The 10 mg/kg DEHP dose advanced pubertal onset significantly to 39.7 ± 0.1 days postpartum, whereas the 750 mg/kg DEHP dose delayed pubertal onset to 46.3 ± 0.1 days postpartum (see fig 5).

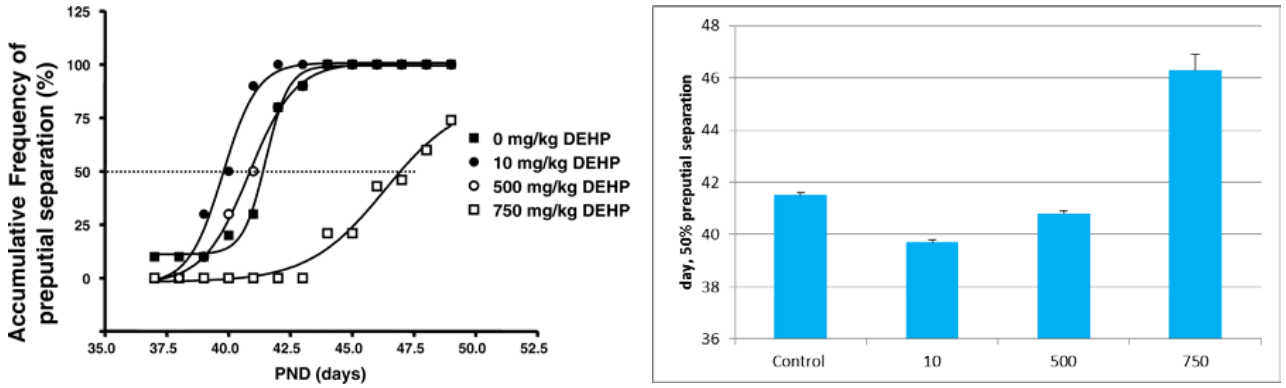

Figure 5. Left (from Ge et al. 2007): Biphasic effect of di(2-ethylhexyl)phthalate (DEHP) exposures on puberty onset assessed by preputial separation. Prepubertal rats were gavaged with DEHP (0, 10, 500, and 750 mg/kg/d). The time course of the accumulative frequency of rats with preputial separation was fitted by sigmoid nonlinear regression. Average age was calculated as the intercept at 50% accumulative frequency, shown as the dotted line. Right: The same results shown as mean values + SEM according to the day of preputial separation.

Moreover, a similar picture was seen for body weight, seminal vesicle weight and serum testosterone. The 10 mg/kg DEHP dose significantly increased serum testosterone (T) levels (3.13 ± 0.37 ng/mL) and seminal vesicle weights (0.33 ± 0.02 g) compared with control serum T (1.98 ± 0.20 ng/mL) and seminal vesicle weight (0.26 ± 0.02 g), while the 750 mg/kg dose decreased serum T (1.18 ± 0.18 ng/mL) as well as testes and body weights. The statistics are well performed however the results are averages from 2 experiments. Thus, this paper demonstrated NMDR as low-dose exposure to DEHP (10 mg/kg) induced increased serum T levels, precocious 2-day advancement in the timing of preputial separation, and increases in seminal vesicle weight in male rats, whereas higher doses of DEHP (750 mg/kg/d) had the opposite effect of lowering T levels and delaying puberty. However, it is important to keep in mind that the NMDR might be a secondary to the effect on bodyweight which follows the same pattern.

In the studies reported in Christiansen et al. 2010, the effects of perinatal DEHP exposure was studied in time-mated Wistar rats gavaged from gestation day 7 to postnatal day 16 with 0, 10, 30, 100, 300, 600 and 900 mg/kg bw/day (study 1) and 0, 3, 10, 30 and 100 mg/kg/day (study 2), respectively. The results showed that DEHP at a relatively low dose of 10 mg/kg caused adverse anti-androgenic effects on male rat sexual development. At this dose level, male anogenital distance was decreased, the incidence of nipple retention was increased, weights of levator ani/bulbocavernosus muscle (LABC) were reduced and mild external genitalia dysgenesis was observed.

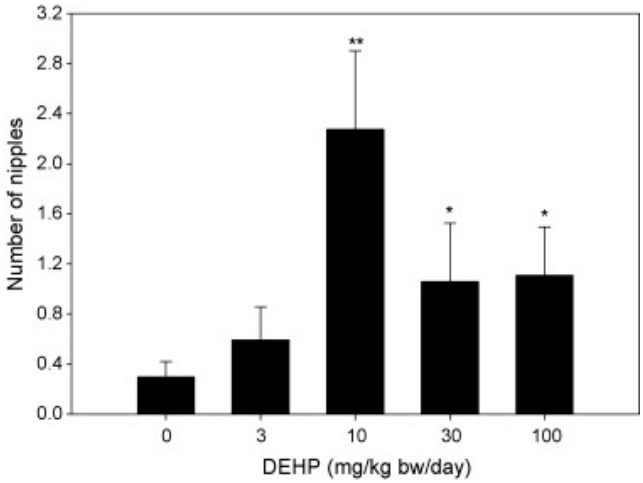

Figure. 6. Mean number of nipples in male rat offspring of dams exposed to corn oil (control), 3, 10, 30 or 100 mg/kg-d DEHP from GD 7 to PND 16. Results are based on analysis of litter means and are presented as mean + SEM. Data represents the combined analysis of study 1 and study 2. *Indicates $p \leq 0.05$, **indicates $p < 0.01$.The exposure to DEHP statistically significantly increased nipple retention in male rat offspring at 10, 30, and 100 mg/kg-d (figure shown here) but also at 300, 600 and 900 mg/kg-d (not shown).

In study 1, perinatal DEHP exposure induced nipple retention in male offspring at all dose levels, i.e. from 10 mg/kg. However, the dose–response relationship seemed non-monotonic, as 10 mg/kg induced a more marked effect than 30 and 100 mg/kg. In study 2, there also seemed to be a higher number of nipples at 10 mg/kg compared to controls, although the difference was not statistically significant. A combined analysis of the data from both studies showed that at doses above 3 mg/kg increased nipple retention was observed (figure 6). The combined dose–response curve also appeared as non-monotonic, as the dose of 10 mg/kg still seemed to induce a more marked effect than 30 mg/kg ($p = 0.053$) and 100 mg/kg, though not statistically significant (Christiansen et al.2010).

These DEHP results indicate NMDR in relation to nipple retention and a similar picture was seen for the incidence of male offspring with mild external genital malformations, and reductions in weight of LABC, with more pronounced effects at 10 mg/kg than at higher doses. These endpoints are all recorded during the last part of the lactation period and the NMDR might therefore be due to special mechanisms or toxicokinetics during this period or it might be due to biological variation and incidental difference in response in animals from these groups. However, the existence of a bi-phasic dose–response pattern for DEHP cannot be excluded. Ge et al. (reported above) found a non-monotonous (biphasic) effect on sexual maturation when exposing male rats to DEHP in the prepubertal period (PND21-48) (Ge et al. 2007). Moreover, Andrade et al. (2006) also found a NMDR curve, i.e. a J-shaped curve where the aromatase activity was inhibited at low doses and increased at high doses in DEHP exposed (GD6-PND21) male rats on PND1 (Andrade et al. 2006 ref. 788 in Vandenberg et al). In the latter study, the decreased aromatase activity at 0.1 and 0.4 mg/kg and increased at 15, 45, 135 and 405 mg/kg in male offspring may not be considered as an adverse effect, but more evidence for an ED mode of action. Taken together these three studies indicate that DEHP induces NMDR for some ED endpoints.

4.5 Conclusions

There are several mechanisms that 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.

NMDR for EDs exists and have been shown and used in human endocrinology as a basic principle behind the pharmaceutical treatment of severe diseases. Also, 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.

5. Uncertainties related to regulatory requirements and test methods

One of the aims of the present report was also to give scientific input on the uncertainties of the currently used regulatory test methods, with regard to determination of possible thresholds for EDs.

In regulatory practise NOAELs are generally used as part of risk assessment or as point of departure for deriving acceptable human exposure levels. NOAELs are, however, not fixed values, but are sensitive to the specific features of the chosen experimental design, the choices of statistical methods and significance criteria. Thus, when there is no statistically significant difference in response between treated groups and controls, it can only be concluded that the magnitude of effect was below the detection limit of the particular experimental arrangement used (Scholze and Kortenkamp 2007).

If the effects of EDs are to be identified within various kinds of regulations, including REACH, it is essential that the testing requirements include studies where the exposure covers windows of increased susceptibility and the relevant endpoints are assessed (Kortenkamp et al. 2012). In addition, it is important that the power for detecting a relevant threshold-like dose is sufficient for the endpoints assessed.

5.1 Current REACH information requirements, test methods

The information requirements for substances for registration under REACH are differentiated according to supply tonnage. Generally, testing requirements at a lower tonnage level apply to the higher tonnage level, unless exemptions are clearly stated. The current information requirements in REACH is not designed for the identification of endocrine disrupters, but some relevant test methods for detection of endocrine disrupters are mentioned in relation to testing for repeated dose

toxicity, carcinogens and reproductive toxicants. The minimum information requirements for repeated dose toxicity and reproductive toxicity are summarized in Table 1. It should be noted that interpretation of the testing requirements by the registrants in practice depends on a weight-of-evidence evaluation of existing data and may therefore be different to the minimum requirements as presented here.

Table 1. Repeated dose toxicity and reproductive toxicity testing minimum information requirements under REACH by tonnage level

	≥ 1 t/year	≥ 10 t/year	≥ 100 t/year	≥ 1000 t/year
Repeated dose toxicity	None	28-day repeated dose oral toxicity study in rodents (OECD TG 407)	90-day repeated dose oral toxicity study in rodents (OECD TG 408)	90-day repeated dose oral toxicity study in rodents (OECD TG 408)
Reproductive toxicity	None	Screening for reproductive/developmental toxicity (OECD TGs 421 or 422)	Prenatal development toxicity study (OECD TG 414) in one species, and if appropriate in a second species	Prenatal development toxicity study (OECD TG 414) in one species, normally in a second species Two-generation reproduction toxicity study (OECD TG 416)

5.1.1 Repeated dose toxicity

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 validated in relation to identification of endocrine disrupters. The validation of OECD TG 407 in relation to endocrine endpoints showed that this assay is relatively insensitive and would only detect chemicals that are moderate and strong EDs for (anti)-estrogenicity and (anti)-androgenicity (e.g. ethinylestradiol and flutamide) (OECD GD 150). The assay did, however, 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 (OECD TG 407).

The OECD TG 407 measures some parameters which are relevant to endocrine-mediated toxicity such as the weight and histopathology of the pituitary, adrenals, ovaries and ventral prostate. Some of the endpoints, particularly those related to the thyroid, are optional, and the lack of relevant endpoints is particularly striking for those most relevant to the testicular dysgenesis syndrome (Kortenkamp et al. 2012). In conclusion, there are major limitations for these studies in terms of screening for endocrine disrupting properties and these are mainly related to the fact that only adult animals are exposed and the limited sensitivity of the gross endocrine endpoints.

5.1.2 Carcinogenicity

There are no standard information requirements in relation to identification of carcinogenic properties for substances produced or imported in quantities of less than 1000 tons per year. A carcinogenicity study for substances produced or imported in quantities ≥ 1000 tons per year may be required if the substance has a widespread dispersive use or there is evidence of frequent or long-

term human exposure and the substance is classified as mutagen category 3 or there is evidence from the repeated dose toxicity study(ies) that the substance is able to induce hyperplasia and/or pre-neoplastic lesions.

Some endpoints that are relevant in relation to hormonally mediated cancers are included in the repeated dose toxicity tests and may trigger a carcinogenicity study if information on use and human exposure warrant it. The limitations of standard repeated dose studies (OECD TG 407 and OECD TG 408) in terms of the timing of exposure and the sensitivity of the endpoint have already been mentioned above and this raise doubt over the likelihood that potential effects on hormonally mediated carcinogenesis will be detected on the basis of those tests (Kortenkamp et al. 2012).

5.1.3 Reproductive toxicity

The minimum information requirements in relation to reproductive toxicity are summarised in Table 1.

Neither the combined repeated dose toxicity/reproductive developmental toxicity screening tests (OECD TGs 421/422) nor the prenatal development toxicity study have yet been validated for the detection of endocrine disrupters. In the prenatal development toxicity study (OECD TG 414), animals are exposed from implantation to two days before expected birth and in the combined repeated dose toxicity/reproductive toxicity screening studies animals are exposed from two weeks prior to mating to four days postnatally. Although these tests include exposure during pregnancy, the endpoints related to fertility and gestation maintenance are measured in the parent generation. Thus, a major limitation of these studies is that they do not include exposure during critical windows of development for those endpoints.

In the prenatal development toxicity study (OECD TG 414), the foetuses are inspected for gross anomalies. However, important differences between humans and rodents concerning the timing of birth compared to developmental stage should be borne in mind. Rodents are compared to humans born at a relatively immature stage and some parts of the sexual differentiation of the brain and reproductive organs that take place during the third trimester of human pregnancy occur after birth in the rat. This means that data from the prenatal development toxicity study have very limited use for evaluating effects of EDs during the third trimester of human pregnancy.

Gross evaluation of anogenital distance is generally used for sexing the offspring in reproductive toxicity studies, because anogenital distance is normally twice as long in males compared to females. Thus major effects on male sexual differentiation induced by potent anti-androgens may be detected as all offspring may display female-like anogenital distance (e.g. Hass et al. 2007).

In conclusion, although these studies include endocrine relevant endpoints for fertility effects and developmental effects, they have major limitations with regards to the endpoints related to fertility as the exposure is not during critical windows of development and they also have very limited sensitivity for detecting effects of EDs on sexual differentiation. Furthermore, the lower number of animals used (8-10 parental males and females) decrease their statistical power compared to e.g. the two-generation study and the extended one-generation study. These considerations raise uncertainty

as to the ability of the current testing requirements to adequately screen for endocrine disrupting properties at tonnage levels below 1000 tons per year (Kortenkamp et al. 2012).

For chemicals with a supply tonnage level over 1000 tons per year, a two-generation study is generally required. This study includes exposure during sensitive time windows of development and assessment of a number of endpoints sensitive to endocrine disruption in the offspring. Results of two-generation reproduction toxicity studies (OECD TG 416) should nonetheless be interpreted with caution: some endocrine sensitive endpoints were added only in 2001 as a result of an update of the guideline. Further, some endpoints sensitive to endocrine disruption are not included in the updated version of the two-generation reproduction study, such as nipple retention, anogenital distance at birth, and measurement of thyroid hormones. Thus, for the two-generation reproduction toxicity study there are uncertainties with regard to the ability to adequately detect endocrine disrupters.

The new extended one-generation reproduction toxicity study (OECD TG 443) includes the above mentioned ED sensitive endpoints as well as assessment of neurodevelopment and immunotoxicity. Thus, the new EOGRT study (OECD TG 443) is preferable for detecting endocrine disruption because it provides an evaluation of a number of endocrine endpoints in the juvenile and adult F1, which are not included in the 2-generation study (OECD TG 416) adopted in 2001(OECD GD 150). This test is also expected to have greater sensitivity than OECD TG 416 as it requires an increased number of pups to be examined. In summary, the exposure of the foetus (which is a sensitive life-stage for endocrine disruption effects), the long duration of dosing and the diversity of endpoints means that the extended one-generation study may be considered to be the most predictive test for ED-mediated adverse effects via EATS modalities (OECD GD 150). Therefore, the use of the extended one-generation reproduction toxicity study (OECD TG 443) instead of the two-generation study would significantly enhance the ability for detection of endocrine disrupters at tonnage levels above 1000 tons per year.

Delayed effects of developmental exposure to EDs that can manifest themselves only with ageing such as premature reproductive senescence are currently not included in any guideline study. Such ED effects are clearly severe, however, there is at present not sufficient scientific knowledge for evaluating whether effects observed earlier in life may protect also towards such late effects or whether they may occur at lower doses than early effect.

5.1.4 Sensitivity for finding a relevant threshold-like dose for EDs, using power analysis

Power analysis can be used to calculate the minimum sample size required, in order to likely detect an effect of a given size. Power analysis can also be used to calculate the minimum effect size that is likely to be detected in a study using a given sample size.

A number of endpoints relevant for ED provide quantal data, i.e. they are results providing yes/no answers, like for example data on malformations of reproductive organs, or fertility- and pregnancy index. Assessment of quantal endpoints is generally expected to lead to a lower power than assessment of continuous data (e.g. anogenital distance (AGD) or reproductive organ weights). To explore this, we have calculated the effect size needed for finding significant effects, i.e. $p < 0.05$,

for yes/no endpoints and continuous endpoints. The methods and results from these calculations are described in Appendix 3, whereas the next section will only provide the conclusions.

To illustrate the importance of sample size, the power calculations were done for studies using either 8 or 20 litters per group, which are the group sizes required in the OECD TG 421/422 (Reproduction/Developmental toxicity Screening study) and in the OECD TG 416 and OECD TG 443(two- generation and extended one generation study), respectively. Overall, the results illustrated that the effect size for a quantal effect had to be 25-37% in studies with 20 litters per group, and even higher (50-75%) with only 8 litters per group. This implies that the sensitivity for detecting quantal effects is very low and that effect sizes of human relevance may be present at the NOAEL.

For continuous endpoints the statistical power for detecting significant effects depends on the group size, and on the coefficient of variation (CV) in the control group. For AGD data, the CV is normally around 5- 7% and the calculations performed in appendix 2 show that in studies with 20 animals per group, an effects size of ca. 4-7% will in most cases be statistically significant, whereas the effect size has to be 7-11% if only 8 litters per group are studied. So for continuous data, effect sizes needed for detecting significant effects with 8 animals per group are approximately 1.6-1.8 times higher than when 20 animals per group are used.

Continuous effect data are generally expected to be more sensitive than quantal data and this was also found by the present calculations, as the effect sizes needed for continuous data ranged from 4-18%, whereas the effect sizes for quantal data were found to range from 25-75%. In spite of the increased sensitivity of continuous data compared to quantal data, effect sizes of human relevance may also be present at the NOAEL for continuous data.

5.2 Conclusions

The current information requirements in REACH are not designed for the identification of endocrine disrupters, although certain endpoints and assays may give some indication of endocrine disrupting effects. It is, however, evident that important endpoints needed for the detection of ED effects are not included. Especially, important effects after exposure that cover windows of susceptibility during development are not assessed. This raises major uncertainty as to the ability of the current testing requirements to adequately detect EDs at tonnage levels below 1000 tons per year. A two-generation reproduction toxicity study is generally required for chemicals with a supply tonnage level above 1000 tons per year and this study includes exposure during sensitive time windows of development and assessment of a number of endpoints sensitive to endocrine disruption in the offspring. However, 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, and measurement of thyroid hormones. Thus, for the two-generation reproduction toxicity study there are also uncertainties with regard to the ability to adequately detect endocrine disrupters.

The new extended one-generation reproduction toxicity study (OECD TG 443) includes the above mentioned ED sensitive endpoints. The exposure of the foetus (which is a sensitive life-stage for endocrine disruption effects), the long duration of dosing and the diversity of endpoints means that the extended one-generation study may be considered to be the most predictive test for ED-mediated adverse effects via EATS modalities (OECD GD 150). Therefore, using the extended one-generation study instead of the two-generation study would significantly enhance the ability for detection of endocrine disrupters at tonnage levels above 1000 tons per year.

Power calculations for studies using 8 or 20 litters per group, which are the group sizes required in the OECD TG 421/422 (Reproductive Screening study) and in the OECD TG 416 and OECD TG 443(two- and extended one generation study) illustrated that the effect size needed for detection of quantal effects had to be 25-37% with 20 litters per group and 50-75% with 8 litters per group. This clearly shows that the sensitivity for detecting quantal effects is very low and that effects sizes of human relevance may be present at the NOAEL. The effect sizes needed for continuous data ranged from 4-18%, so in spite of the increased sensitivity of continuous data compared to quantal data, effect sizes of human relevance may also be present at the NOAEL for continuous data.

6. Are EDs of particular concern?

The European Commission published in the beginning of 2012 a report on "State of the art assessment of endocrine disruptors", where it is indicated that EDs are of similar concern as CMRs (carcinogens, mutagens, reproductive toxicants) and PBTs. The arguments for this include that EDs induce irreversible and very severe effects and that exposure during sensitive windows of development can lead to occurrence of such effects also later in life.

Many endocrine disruptors are already or can be identified as carcinogens or reproductive toxicants due to the inherent endocrine disrupting properties. A common characteristic for CMRs is that effects may often occur with a time lag of several years after the exposure.

The majority of the effects potentially related to human exposure to EDs during 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 individuals life time. Thus, there may be a time lag of many years or several decades from regulatory decisions on risk reduction are taken, to the time when this risk reduction will be achieved and this is of particular concern when the regulation aims for reduction of risks to chemicals causing severe and delayed effects.

A common reason for considering both PBTs and vPvBs as substances of very high concern is expressed by the P, i.e. that the substances are persistent. A characteristic for both persistent and bioaccumulative substances is that exposure to these substances will occur long time after the initial

source of exposure has ceased. This means that there can be a time lag of many years or decades from regulatory decisions on risk reduction are implemented to the time when exposure to these chemicals diminishes and it is therefore difficult to control the risk. With regard to persistent and bioaccumulating chemicals that are toxic due to endocrine disrupting properties, fat-soluble, persistent EDs are accumulated in the body fat and humans will be exposed for a long time after the initial source of exposure to the substance has ceased. The consequences of long-term continued exposure to bioaccumulated EDs for the complex functioning of the endogenous hormonal system are largely unknown.

In conclusion, EDs are evaluated as being of particular concern, because exposure during sensitive time windows of development may cause irreversible developmental programming effects leading to severe health effects manifested late in life, and also because the consequences of long-term continued exposure on the complex hormonal system are largely unknown.

7. Summary, conclusions and recommendations

The aim of this report is, from a scientific point of view, to discuss the topics expected to be relevant for the REACH review on EDs, i.e.:

- Thresholds or non-threshold assumption for ED effects
- Considerations concerning non-monotonic dose-response (NMDR)
- Uncertainties of the currently regulatory test methods with regard to determination of possible thresholds for EDs
- Whether there is particular concern for EDs.

The presence of thresholds 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 because low doses of endogenous hormones are present and fluctuating, 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 of hormones during development at a time point where the homeostatic control is not effective or not developed yet. Also, experimental data indicate non-thresholded dose-response 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 appears more valid for the effects of EDs during development than an assumption of a threshold.

Regardless of ED mode of action, it is uncertain whether or not there is a threshold for EDs. For EDs, where the MoA (Mode of Action) directly involve the receptor, the interaction with the receptor is likely to have no threshold. For EDs affecting the hormone levels, the response pattern

may appear threshold-like, because multiple pathways converge before seeing the final response and some of these pathways may have a threshold.

Irrespective of threshold or non-threshold, the dose response curves of EDs 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, however, have a “threshold-like” appearance, but a threshold cannot be inferred from the shape of the dose-response curves. However, a benchmark approach may be used for estimating a human exposure level with very low risk.

There are several mechanisms that 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. NMDR for EDs exists and have been shown and used in human endocrinology as a basic principle behind the pharmaceutical treatment of severe diseases. Also, 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.

The current information requirements in REACH are not designed for the identification of endocrine disrupters, although certain endpoints and assays may give some indication of endocrine disrupting effects. It is, however, evident that important endpoints needed for the detection of ED effects are not included. Especially, important effects after exposure that cover windows of susceptibility during development are not assessed. This raises major uncertainty as to the ability of the current testing requirements to adequately screen for endocrine disrupting properties at tonnage levels below 1000 tons per year. A two-generation reproduction toxicity study is generally required for chemicals with a supply tonnage level above 1000 tons per year and this study includes exposure during sensitive windows of development and assessment of a number of endpoints sensitive to endocrine disruption in the offspring. However, 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, and measurement of thyroid hormones. Thus, for the two-generation reproduction toxicity study there are also uncertainties with regard to the ability to adequately detect endocrine disrupters.

The new extended one-generation reproduction toxicity study (OECD TG 443) includes the above mentioned ED sensitive endpoints. The exposure of the foetus (which is a sensitive life-stage for endocrine disruption effects), the long duration of dosing and the diversity of endpoints means that the extended one-generation study may be considered to be the most predictive test for ED-mediated adverse effects via EATS modalities (OECD GD 150). Therefore, using the extended

one-generation study instead of the two-generation study would significantly enhance the ability for detection of endocrine disrupters at tonnage levels above 1000 tons per year.

Power calculations for studies using 8 or 20 litters per group, which are the group sizes required in the OECD TG 421/422 (Reproductive Screening study) and in the OECD TG 416 and OECD TG 443(two- and extended one generation study) illustrated that the effect size needed for detection of quantal effects have to be 25-37% with 20 litters per group and 50-75% with 8 litters per group. This clearly shows that the sensitivity for detecting quantal effects is very low and that effects sizes of human relevance may be present at the NOAEL. The effect sizes needed for continuous data range from 4-18%, so in spite of the increased sensitivity of continuous data compared to quantal data, effect sizes of human relevance may also be present at the NOAEL for continuous data.

The majority of the effects potentially related to human exposure to EDs during 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 individuals life time and possibly also future generations. Thus, there may be a time lag of many years or several decades from regulatory decisions on risk reduction are taken, to the time when this risk reduction will be achieved and this is of particular concern when the regulation aims for reduction of risks to chemicals causing severe and delayed effects. In conclusion, EDs are evaluated as being of particular concern due to the ability for causing severe and irreversible effects that may possibly also be manifested through next generations, because exposure are especially problematic during sensitive time windows of development, and because the consequences of long-term continued exposure on the complex hormonal system are largely unknown.

Overall, it concluded that there are major uncertainties in relation to the detection of safe levels for human exposure to EDs. These uncertainties include that:

- During development an assumption of no threshold appears more valid than an assumption of a threshold.
- For EDs, where the MoA directly involve the receptor, the interaction with the receptor is likely to have no threshold. For EDs affecting the hormone levels, there may be a threshold, if the substances affect the hormone levels via a mechanism where there is a threshold.
- NMDR for EDs exists and this knowledge is used in human endocrinology as a basic principle behind the pharmaceutical treatment of severe diseases. Also, NMDR has been shown for many ED-mediated *in vitro* and *in vivo* effects including binding to steroid hormone receptors and adverse effects and this can directly reflect the way the endocrine system works.
- There are major limitations as to the ability of the current testing requirements to adequately screen for endocrine disrupting properties and effect sizes of human relevance may be present at the NOAEL.

- Delayed effects of developmental exposure to EDs that can manifest themselves only with ageing such as premature reproductive senescence are currently not included in any guideline study.
- EDs are evaluated as being of particular concern, because exposure during sensitive time windows of development may cause irreversible developmental programming effects leading to severe health effects manifested late in life, and also because the consequences of long-term continued exposure on the complex hormonal system are largely unknown.

Based on these conclusions, we recommend that:

- A sufficient regulatory testing scheme should be developed for detection of EDs. Assessment of adverse ED effects as well as investigations of MoA are relevant as both are important for evaluating whether a substance is an ED
- Enhancement of the Reproduction/developmental toxicity Screening studies (OECD TG 421/422) and the Prenatal developmental toxicity study (OECD TG 414) with regard to detection of ED effects should be considered
- The new extended one-generation reproduction toxicity study (OECD TG 443) should replace the two-generation reproduction toxicity study (OECD TG 416), as this would significantly enhance the ability to identify endocrine disrupting substances
- The scientific knowledge needed for evaluating whether ED effects observed early in life also protects towards delayed effects that becomes manifest only with ageing such as premature reproductive senescence should be improved
- Effects seen at low doses, but not at higher doses of EDs should be carefully evaluated and interpreted, as EDs may cause NMDR
- The number of dose levels in experimental studies should be increased to better characterize the dose-response and increase the possibility for detection of NMDR
- A benchmark dose (BMD) approach where both effect size and severity is included should be used when estimating human risk instead of a NOAEL approach. Using the BMD approach, poor data quality will lead to a lower BMD and better data, with their reduced degree of uncertainty, are “rewarded” with higher BMDs whereas poor data quality usually results in higher NOAELs. However, sufficient dose-response data for a BMD approach may in many cases not be available and NMDR may not be detected with current regulatory testing where only three dose levels are required.

8. References

Abt E, Rodricks JV, Levy JI, Zeise L, Burke TA. 2010 Science and decisions: advancing risk assessment. Risk Anal. 2010 Jul;30(7):1028-36. doi: 10.1111/j.1539-6924.2010.01426.x. Epub 2010 May 20.

Alm, H.; Kultima, K.; Scholz, B.; Nilsson, A.; Andren, P. E.; Fex- Svenningsen, A.; Dencker, L.; Stigson, M. Exposure to brominated flame retardant PBDE-99 affects cytoskeletal protein expression in the neonatal mouse cerebral cortex. Neurotoxicology 2008, 29 (4), 628–37.

Almstrup K, Fernandez MF, Petersen JH, Olea N, Skakkebaek NE, Leffers H 2002 Dual effects of phytoestrogens result in U-shaped dose-response curves. Environ Health Perspect 110:743–748

Andrade AJM, Grande SW, Talsness CE, Grote K, Chahoud I. A dose-response study following in utero and lactational exposure to di-(2-ethylhexyl)-phthalate (DEHP): Non-monotonic dose-response and low dose effects on rat brain aromatase activity. Toxicology 2006;227:185-192.

Ashby J, Tinwell H, Odum J, Lefevre P 2004 Natural vari- ability and the influence of concurrent control values on the detection and interpretation of low-dose or weak endocrine toxicities. Environ Health Perspect 112:847– 853

Asp V, Ullerås E, Lindström V, Bergström U, Oskarsson A, Brandt I 2010 Biphasic hormonal responses to the adrenocorticolytic DDT metabolite 3-methylsulfonyl-DDE in human cells. Toxicol Appl Pharmacol 242:281–289

Blair, R.M., Fang, H., Gaylor, D., Sheehan, D.M., 2001. Threshold analysis of selected dose–response data for endrocrine disruptors. Acta Path. Microbiol. Immunol. Scand. 109, 198–208.

Boettcher M, Kosmehl T, Braunbeck T 2011 Low-dose effects and biphasic effect profiles: Is trenbolone a genotoxicant? Mutat Res 723:152–157

Boobis AR, Datson GP, Preston RJ, Olin SS. 2009. Application of Key Events Analysis to Chemical Carcinogens and Noncarcinogens. Critical Reviews in Food Science and Nutrition, 49:690-707.

Campagna C, Ayotte P, Sirard MA, Arsenault G, Laforest JP, Bailey JL 2007 Effect of an environmentally relevant metabolized organochlorine mixture on porcine cumulus- oocyte complexes. Reprod Toxicol 23:145–152

Christiansen S, Boberg J, Axelstad M, Dalgaard M, Vinggaard AM, Metzdorff SB & Hass U. (2010) Low-dose perinatal exposure to di(2-ethylhexyl) phthalate induces anti-androgenic effects in male rats. Reprod Toxicol 30, 313–321.

Conolly RB, Lutz WK 2004 Nonmonotonic dose-response relationships: mechanistic basis, kinetic modeling, and im- plications for risk assessment. Toxicol Sci 77:151–157

Ebersolt C, Premont J, Prochiantz A, Perez M, Bockaert J. Inhibition of brain adenylate cyclase by A1 adenosine receptors: pharmacological characteristics and locations. Brain Res. 1983 May 9;267(1):123-9.

European Chemicals Agency, 2007. Guidance for the preparation of an Annex XV dossier on the identification of substances of very high concern. ECHA. Helsinki, Finland). http://echa.europa.eu/documents/10162/13638/svhc_en.pdf .

Ge RS, Chen GR, Dong Q, Akingbemi B, Sottas CM, Santos M, Sealfon SC, Bernard DJ, Hardy MP. Biphasic Effects of Postnatal Exposure to Diethylhexylphthalate on the Timing of Puberty in Male Rats. J Androl 2007;28:513-520.

Gray L, Ostby J, Monosson E, Kelce WR. Environmental antiandrogens: low doses of the fungicide vinclozolin alter sexual differentiation of the male rat. Toxicol Ind Health 1999;15:48-64.

Hass U, Scholze M, Christiansen S, Dalgaard M, Vinggaard AM, Axelstad M, Metzdorff SB, Kortenkamp A 2007 Combined exposure to anti-androgens exacerbates disruption of sexual differentiation in the rat. Environmental Health Perspectives 115 Suppl 1:122-128

Jacobsen PR, Axelstad M, Boberg J, Isling LK, Christiansen S, Mandrup KR, Berthelsen LO, Vinggaard AM, Hass U. Persistent developmental toxicity in rat offspring after low dose exposure to a mixture of endocrine disrupting pesticides. Reprod Toxicol. 2012 Sep;34(2):237-50. Epub 2012 Jun 4.

Jeng YJ, Kochukov MY, Watson CS 2009 Membrane estrogen receptor-alpha-mediated nongenomic actions of phytoestrogens in GH3/B6/F10 pituitary tumor cells. J Mol Signal 4:2

Kohn MC, Melnick RL 2002 Biochemical origins of the non-monotonic receptor-mediated dose-response. J Mol Endocrinol 29:113–123

Kortenkamp A, Martin O, Faust M, Evans R, McKinlay R, Orton F, Rosivatz E. (2012). State of the Art Assessment of Endocrine Disrupters. Final Report. [Online] Available at: http://ec.europa.eu/environment/endocrine/documents/4_SOTA%20EDC%20Final%20Report%20V3%206%20Feb%2012.pdf

Leung LY, Kwong AK, Man AK, Woo NY 2008 Direct actions of cortisol, thyroxine and growth hormone on IGF-I mRNA expression in sea bream hepatocytes. Comp Biochem Physiol A Mol Integr Physiol 151:705–710

Maness SC, McDonnell DP, Gaido KW(1998) Inhibition of androgen receptor-dependent transcriptional activity by DDT isomers and methoxychlor in HepG2 human hepatoma cells. Toxicol Appl Pharmacol 151:135–142.

Melnick R, Lucier G, Wolfe M, Hall R, Stancel G, Prins G, Gallo M, Reuhl K, Ho S, Brown T, Moore J, Leakey J, Haseman J, Kohn M (2002) Summary of the National Toxicology Program's Report of the Endocrine Disruptors Low-Dose Peer Review. Env Health Perspect 110: 427-431.

Metzdorff SB, Dalgaard M, Christiansen S, Axelstad M, Hass U, Kiersgaard MK, et al. 2007. Dysgenesis and histological changes of genitals and perturbations of gene expression in male rats after in utero exposure to antiandrogen mixtures. Toxicol Sci 98:87–98.

Moore, KL 1983. Before we are born: Basic Embryology and Birth Defects
Second edition Saunders (ISBN 10: 0721610242 / ISBN 13: 9780721610245)

Naciff, J. M.,Hess, K. A., Overmann, G. J., Torontali, S. M.,Carr, G. J., Tiesman, J. P., Foertsch, L. M., Richardson, B. D., Martinez, J. E. and Daston, G. P. (2005b). Gene expression changes induced in the testis by transplacental exposure to high and low doses of 17-alpha-ethynyl estradiol, genistein, or bisphenol A. Toxicol. Sci. 86:396–416.

National Toxicology Program 2001 National Toxicology Program's report of the endocrine disruptors low dose peer review. Research Triangle Park, NC: National Institute of Environmental Health Sciences

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

OECD 2012: Guidance Document on Standardised Test Guidelines for Evaluating Chemicals for Endocrine Disruption. Series on Testing and Assessment No. 150, ENV/JM/MONO(2012)22, 24-Aug-2012;
http://www.oecd.org/env/ehs/testing/seriesontestingandassessmentpublicationsbynumber.htm

Ohlsson A, Cedergreen N, Oskarsson A, Ullerås E 2010 Mixture effects of imidazole fungicides on cortisol and aldosterone secretion in human adrenocortical H295R cells. Toxicology 275:21–28

Putz O, Schwartz CB, Kim S, LeBlanc GA, Cooper RL, Prins GS 2001 Neonatal low- and high-dose exposure to estradiol benzoate in the male rat. I. Effects on the prostate gland.Biol Reprod 65:1496 –1505

Ralph JL, Orgebin-Crist MC, Lareyre JJ, Nelson CC 2003 Disruption of androgen regulation in the prostate by the environmental contaminant hexachlorobenzene. Environ Health Perspect111:461–466

Scholze M, Kortenkamp A. 2007 Statistical power considerations show the endocrine disruptor low-dose issue in a new light.
Environ Health Perspect. 2007 Dec;115 Suppl 1:84-90. doi: 10.1289/ehp.9364.

Sheehan DM et al. 1999. No threshold dose for estradiol induced sex reversal of turtle embryos: how little is too much? Environ Health Perspect 107: 155-159.

Sheehan DM. No-threshold dose-response curves for nongenotoxic chemicals: findings and applications for risk assessment.Environ Res. 2006 Jan;100(1):93-9. Epub 2005 Oct 26.

Siler-Khodr TM, Fetal Hormones page307-317 in Encyclopedia of Reproduction, 2nd Volume Set, 1998, Academic Press

Slob W. 1999. Thresholds in Toxicology and Risk Assessment. International Journal of Toxicology 18:259-268; Scholze M and Kortenkamp A. 2007. Statistical power considerations show the endocrine disrupter low dose issue in a new light. Environ Health Perspect 115 Suppl 1: 84-90.

Somjen D, Kohen F, Jaffe A, Amir-Zaltsman Y, Knoll E, Stern N 1998 Effects of gonadal steroids and their antag- onists on DNA synthesis in human vascular cells. Hyper- tension32:39 – 45

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

vom Saal FS, Timms BG, Montano MM, Palanza P, Thayer KA, Nagel SC, Dhar MD, Ganjam VK, Parmigiani S, Welshons WV 1997 Prostate enlargement in mice due to fetal exposure to low doses of estradiol or diethylstilbestrol and opposite effects at high doses. Proc Natl Acad Sci USA 94:2056 –2061

Zoeller RT, Brown TR, Doan LL, Gore AC, Skakkebaek NE, Soto AM, Woodruff TJ, Vom Saal FS. Endocrine-disrupting chemicals and public health protection: a statement of principles from The Endocrine Society Endocrinology. 2012 Sep;153(9):4097-110. doi: 10.1210/en.2012-1422. Epub 2012 Jun 25

Appendix 1

This appendix gives some more details on the 80 evaluated *in vitro* examples allocated into four groups. Almost half of the examples (45%) could not according to our definition be regarded as showing a true non-monotonic dose-response, as the NMDR was evaluated as due to cytotoxicity. Furthermore, some examples were evaluated as “false NMDR”, because of e.g. testing of mixtures or limitations in the study design. The remaining examples were evaluated to either show evidence for NMDR (16%) or a dose-response that may or may not be due to NMDR of EDs (17%).

Chemicals by chemical class	Nonmonotonic effect	Cell type	Refs.	Evaluation
Natural hormones				
17R-Estradiol	Cell number	MCF7 breast cancer cells	135, 716	Cytotox
	Dopamine uptake	Fetal hypothalamic cells (primary)	717	Maybe
	pERK levels, prolactin release	GH3/B6/F10 pituitary cells	41, 718, 719	NMDR
	R-Hexosaminidase release	HMC-1 mast cells	720	Cytotox?
	Cell number	Vascular smooth muscle cells	721	False
	Production of L-PGDS, a sleep-promoting substance	U251 glioma cells	722	Cytotox
5a-Dihydrotestosterone	Cell number	LNCaP-FGC prostate cancer cells	499	Cytotox
	Cell number, kinase activity	Vascular smooth muscle cells	721	False
5a-Androstenedione	Cell number	LNCaP-FGC prostate cancer cells	499	Cytotox
Corticosterone	Mitochrondrial oxidation, calcium flux	Cortical neurons (primary)	723	Cytotox
Insulin	Markers of apoptosis (in absence of glucose)	Pancreatic R-cells (primary)	724	Cytotox
Progesterone	Cell number	LNCaP-FGC prostate cancer cells	499	Cytotox
Prolactin	Testosterone release	Adult rat testicular cells (primary)	725	Cytotox
hCG	Testosterone release	Adult rat testicular cells (primary)	725	Cytotox
T3	Rate of protein phosphorylation	Cerebral cortex cells (primary, synaptosomes)	726	NMDR
	LPL mRNA expression	White adipocytes (rat primary)	727	False
GH	*IGF-I* expression	Hepatocytes (primary cultures from silver sea bream)	728	False
Pharmaceutical hormones				
DES	Cell number	MCF7 breast cancer cells	716	Cytotox
	Prolactin release	GH3/B6/F10 pituitary cells	41	NMDR
Ethinyl estradiol	CXCL12 secretion	MCF7 breast cancer cells, T47D breast cancer cells	729	Maybe
R1881 (synthetic androgen)	Cell number	LNCaP-FGC cells	499	Cytotox
Trenbolone	Induction of micronuclei	RTL-W1 fish liver cells	730	NMDR
Plastics				
BPA	Cell number	MCF7 breast cancer cells	135, 716	Cytotox
	Dopamine efflux	PC12 rat tumor cells	40	NMDR
	pERK levels, intracellular Ca^{2+} changes, prolactin release	GH3/B6/F10 pituitary cells	41, 718	NMDR
	Cell number	LNCaP prostate cancer cells	731	False
DEHP	Number of colonies	*Escherichia coli* and *B. subtilis* bacteria	732	Maybe
Di-*n*-octyl phthalate	Number of colonies	*E. coli* and *B. subtilis* bacteria	732	Maybe

Chemicals by chemical class	Nonmonotonic effect	Cell type	Refs.	Evaluation
Detergents, surfactants				
Octylphenol	Cell number	MCF7 breast cancer cells	716	Cytotox
	Dopamine uptake	Fetal hypothalamic cells (primary)	717	Maybe
	pERK levels	GH3/B6/F10 pituitary cells	718	NMDR
	HCG-stimulated testosterone levels	Leydig cells (primary)	733	NMDR
Propylphenol	pERK levels	GH3/B6/F10 pituitary cells	718	NMDR
Nonylphenol	pERK levels, prolactin release	GH3/B6/F10 pituitary cells	41, 718	NMDR
	R-Hexosaminidase release	HMC-1 mast cells	720	Cytotox
	Cell number	MCF7 breast cancer cells	135	Cytotox
PAH				
Phenanthrene	All-trans retinoic acid activity	P19 embryonic carcinoma cells	734, 735	Maybe
Benz(a)acridine	All-trans retinoic acid activity	P19 embryonic carcinoma cells	734	Maybe
Naphthalene	hCG-stimulated testosterone	Pieces of goldfish testes	736	Cytotox
B-naphthoflavone	hCG-stimulated testosterone	Pieces of goldfish testes	736	Cytotox
Retene	hCG-stimulated testosterone	Pieces of goldfish testes	736	Cytotox
Heavy metals				
Lead	Estrogen, testosterone, and cortisol levels	Postvitellogenic follicles (isolated from catfish)	737	Cytotox
Cadmium	Expression of angiogenesis genes	Human endometrial endothelial cells	738	Maybe
Phytoestrogens and natural antioxidants				
Genistein	Cell number	Caco-2BBe colon adenocarcinoma cells	739	Cytotox
	CXCL12 secretion, cell number	T47D breast cancer cells	729	Maybe
	Cell number, cell invasion, MMP-9 activity	PC3 prostate cancer cells	740	Cytotox
	pJNK levels, Ca^{2+} flux	GH3/B6/F10 pituitary cells	719	NMDR
Coumesterol	Prolactin release, pERK levels	GH3/B6/F10 pituitary cells	719	NMDR
Daidezin	Prolactin release, pERK levels	GH3/B6/F10 pituitary cells	719	NMDR
	Cell number	MCF7 breast cancer cells	135	Cytotox
	Cell number	LoVo colon cancer cells	741	Maybe
Resveratrol	Expression of angiogenesis genes	Human umbilical vein endothelial cells	742	NMDR
Trans-resveratrol	pERK levels, Ca^{2+} flux	GH3/B6/F10 pituitary cells	719	NMDR
Artelastochromene	Cell number	MCF7 breast cancer cells	743	Maybe
Carpelastofuran	Cell number	MCF7 breast cancer cells	743	Maybe
Biochanin A	Induction of estrogen-sensitive genes	MCF7 breast cancer cells	744	Maybe
Licoflavone C	Induction of estrogen-sensitive genes	Yeast bioassay	745	Maybe
Quercetin	Aromatase activity	H295R adrenocortical carcinoma cells	746	Cytotox
	Cell number	SCC-25 oral squamous carcinoma cells	747	Maybe

Chemicals by chemical class	Nonmonotonic effect	Cell type	Refs.	Evaluation
Dioxin				
TCDD	Cell number, gene expression	M13SV1 breast cells	748	Maybe
PCB				
PCB-74	Cell viability, GnRH peptide levels	GT1-7 hypothalamic cells	749	Cytotox
PCB-118	Cell viability, GnRH peptide levels	GT1-7 hypothalamic cells	749	Cytotox
Aroclor 1242 (PCB mixture)	β-Hexosaminidase release	HMC-1 mast cells	720	Cytotox
POP mixture	Apoptosis of cumulus cells	Oocyte-cumulus complexes (primary, from pigs)	750	False
Herbicides				
Glyphosphate-herbicide (Round-Up)	Cell death, aromatase activity, ERβ activity	HepG2 liver cells	751	Cytotox
Atrazine	Cell number	IEC-6 intestinal cells	752	False
Insecticides				
Endosulfan	Cell number	IEC-6 intestinal cells	752	False
	β -Hexosaminidase release	HMC-1 mast cells	720	Cytotox
	ATPase activity of P-glycoprotein	CHO cell extracts	753	Maybe
Diazinon	Cell number	IEC-6 intestinal cells	752	False
Dieldrin	β -Hexosaminidase release	HMC-1 mast cells	720	Cytotox
DDT	Cell number	MCF7 breast cancer cells	144	Not evaluated
DDE	β-Hexosaminidase release	HMC-1 mast cells	720	Cytotox
	Prolactin release	GH3/B6/F10 pituitary cells	41	NMDR
3-Methylsulfonyl-DDE	Cortisol and aldosterone release, steroidogenic genes	H295R adrenocortical carcinoma cells	754	Cytotox
Fungicides				
Hexachlorobenzene	Transcriptional activity in the presence of DHT	PC3 prostate cancer cells	755	Cytotox
Prochloraz	Aldosterone, progesterone, and corticosterone levels; expression of steroidogenic genes	H295R adrenocortical cells	756	Cytotox?
Ketoconazole	Aldosterone secretion	H295R adrenocortical cells	757	False
Fungicide mixtures	Aldosterone secretion	H295R adrenocortical cells	757	False
PBDE				
PBDE-49	Activation of ryanodine receptor 1	HEK293 cell (membranes)	758	Cytotox
PBDE-99	Expression of GAP43	Cerebral cortex cells (primary)	759	Cytotox

Appendix 2

This appendix gives some more details on the 34 evaluated examples (*in vivo)* allocated into the three groups. The majority of the studies, i.e. 22, were evaluated to give some evidence for NMDR (Group 2) and 5 studies showed clear evidence (Group 1). Poor or no evidence for NMDR was found for 7 of the studies (Group 3).

Chemical class	NMDR effect; Organ/sex/species	Refs.	Group	MoA and other relevant text (mainly as described in the paper)	No per group; no. doses	Further details and remarks
Natural hormones						
17-Estradiol	Morphological parameters; Mammary gland/female/mice	138, 541	1	The induction of estrogen-target genes in the mammary gland was monotonic in both strains. This type of dose–response curve suggests that estrogens can evoke different effects depending on the different doses at which these effects were tested. The combined effect of these variable responses is reflected in the overall cell number. Similarly, in the mammary gland, estrogens induce proliferation, manifested as ductal growth, while concurrently inducing apoptosis, manifested as lumen formation.	5 per group, 2 strains; 8 groups	Dose-related monotonic increase in uterine weight, non-monotonic (inverse U-shaped) dose-response for number of terminal end buds, ductal extension and ductal area in one of the mice strains. It was assessed whether the peak response of a given parameter occurred at one of the intermediate doses and whether the peak response could be statistically distinguished from the response at the highest dose. When both criteria were met, the parameter of interest was defined as having a non-monotonic response. However, response at the highest dose level was still increased compared to controls.
17-Estradiol	Prostate weight; male/mice	689	1	Potential mechanisms mediating a decrease in prostate weight in response to supraphysiological doses of estrogen include receptor down-regulation and the capacity for estradiol to bind to receptors for other steroids, such as androgen receptors, resulting in antagonistic effects mediated via other receptor systems.	6-8 litters, 1/litter; 5 doses	Caesarean section on GD 19, males reared by foster dams, castrated and given testosterone. At 50% increase in free serum estradiol in male mouse foetuses, the prostate was in adulthood enlarged by 30% relative to untreated males. As the free serum estradiol concentration in male foetuses was increased from 2- to 8-fold, adult prostate weight decreased relative to males exposed to the 50% increase in estradiol.
17-Estradiol	Uterine weight; female/mice	761	2		At least 5 per group; 6 doses	Dose-related increase in relative uterine weight in groups 1-4, decreased in group 5-6, but still increased compared to group 1. Thus, no anti-oestrogenic effect. Also, no statistics.
17-Estradiol	Antidepressant effects, measured by immobility assay; Behaviour/male/ mice	762	2	Interference with multiple neuromotor transmitter systems, i.e. dopaminergic and serotonergic	At least 6 per group; 5 doses given 45 min before testing	Immobility in forced swim test. Decreased at 10 and 20 µg/kg and increased at 40 µg/kg.

Chemical class	NMDR effect; Organ/sex/species	Refs.	Group	MoA and other relevant text (mainly as described in the paper)	No per group; no. doses	Further details and remarks
17-Estradiol	Nocturnal activity, gene expression in preoptic area; Brain and behaviour/ female/mice	763	2	We infer that increases in behavioural arousal elicited by estrogens are mediated by changes in the levels of coupled signaling molecules. Given that treated animals have higher motor activity and lower levels of L-PGDS and A2A receptor mRNAs in sleep-active areas, these correlational findings support the hypothesis that estradiol may increase behavioural arousal by decreasing the levels of well-known sleep-inducing molecules within the preoptic region.	12 per group; 7 doses	Running wheel activity (RWA) in ovarietectomized mice increased in groups 2 and 3 and decreased in groups 4-7 compared to group 3. However, RWA in group 7 is increased compared to control. Placed in group 2 mainly because ovarietectomized mice were used.
Corticosterone	Spatial memory errors; Behaviour/male/rats	764	2	The elevated corticosterone levels needed to occur in conjunction with a behavioural stress state for corticosterone-related memory impairments to be expressed.	13 per group; 3 and 5 doses	Radial arm water maze. Lower or higher than normal levels of corticosterone caused increased number of errors.
Corticosterone	Contextual fear conditioning; Behaviour/male/rats	767	2	??	10 per group; 5 doses	The results of Experiment 1 indicate that administration of corticosterone after memory training enhances consolidation of CFC in a dose-dependent manner. The enhancing effect has essentially an inverted U-function, i.e. only statistically significant at dose 4.
Corticosterone	Locomotor activity; Behaviour/male/ captive Adelie penguins	768	3			Adelie penquins - relevance here? Anyway, cannot find data or figure with NMDR.
T4	Bone growth; Tibia/male/rats with induced hypothyroidism	771	3	May be due to effect on body weight or general toxicity or a specific effect of T4	10 per group; 5 doses	Made hypothyroid by methimazol. Increase in epiphysial growth at 2, 8 and 32 µg/kg, decrease at 64 µg/kg. Similar profile for body weight. Only the right montonic part of the dose-response is likely to be seen in animals with normal T4 levels
Oxytocin	Memory retention; Behaviour/male/ mice	773	2	Neuromodulator role in the CNS.	10 or 15 per group; 6 groups	1 sc. dose immediately after training, adult mice, receptor antagonist AOT induced a dose-dependent inverted U-shaped increase in retention performance. Relevance for ED uncertain.

Chemical class	NMDR effect; Organ/sex/species	Refs.	Group	MoA and other relevant text (mainly as described in the paper)	No per group; no. doses	Further details and remarks
Dopamine	Memory; Brain/both/rhesus monkey	775	2		3-5 per group? 4 doses for each substance	Selective dopamine D1 receptor full agonists A77636 and SKF81297 were examined in aged monkeys for effects on the working memory. Low doses improved performance although higher doses impaired or had no effect on performance. The relevance for ED is uncertain.
Pharmaceuticals						
DES	Sex ratio, neonatal body weight, other neonatal development/both/ Mice	777	2		6-10 per group; 6 (7) doses	Animals at highest dose could not give birth. % males per litter decreased at the two lowest doses, but similar to controls at 3 higher doses. Tendency to more pups per litter at lowest dose and fewer pups per litter at highest dose. Birth weight, PD 2 and PD 6 weigth decreased at lowest dose and increased at 2nd highest or highest dose, however, statistics may not have considered litter size.
DES	Adult prostate weight; Male/mice	689	1	Potential mechanisms mediating a decrease in prostate weight in response to supraphysiological doses of estrogen include receptor down-regulation and the capacity for estradiol (and possibly other estrogenic chemicals) to bind to receptors for other steroids, such as androgen receptors, resulting in antagonistic effects mediated via other receptor systems.	6-8 litters, 1/litter; 7 doses	Increased prostate weight at dose 3-5, decreased at dose 7. Body weight used as covariate.
DES	Uterine weight; Female/mice	761	2		At least 5 per group; 5 doses	Dose-related increase in relative uterine weight in groups 1-3, decreased in group 4-5, but still increased compared to control. Thus, no anti-oestrogenic effect. Also, no statistics

Chemical class	NMDR effect; Organ/sex/species	Refs.	Group	MoA and other relevant text (mainly as described in the paper)	No per group; no. doses	Further details and remarks
DES	Morphological parameters; Mammary gland/male and female/ mice	779	2	Potential mechanisms mediating the reduction in mammary gland growth at high doses of DES may include receptor downregulation and the capacity for oestrogens to bind to receptors for other hormones – androgen or glucocorticoid receptors, resulting in antagonistic effects mediated via other receptor systems in response to supraphysiological doses of oestrogens. However, the possibility that the effects of high doses of DES are toxic cannot be ruled out.	10-16 immature per group; 7 doses	Young (PD 18) or adults ovarietectomized. The percentage area of the mammary fat pad occupied by mammary epithelial structures progressively increased by DES from dose 0.01 µg/day. The maximum effective dose of DES was 0.1 µg/day both in young intact and adult OV-X females. However, high dose of DES (10 µg/day had the opposite effect (inverted-U-shaped dose–response curve): mammary size decreased to control levels. Body weights were recorded but are not reported. Group 2, because toxicity at the high dose cannot be excluded.
Estradiol benzoate	Dorsal prostate weight, body weight; Male/rats	780	2	The overall dose response of prostate sizes on PNDs 35 and 90 was monotonic; In contrast to absolute weights, relative prostate weights (normalized to BW) on PND 35 showed a nonmonotonic dose response. Thus, the MoA may be via effect on body weight, e.g. toxicity.	SD rats:?? F233:6 per group and 4 doses	On PND 35, there was an increase in prostate weights of SD rats treated with low doses of EB and a decrease in prostate weights of SD rats treated with high doses. The low-dose effect was entirely abolished by PND 90, and only high-dose suppression of organ sizes was found. The transient nature of the effect in low-dose animals suggests an advancement of puberty as the cause for increased reproductive organ weights on PND 35. F344 rats were more sensitive than SD rats to the suppressive effects of high doses of neonatal EB on PND 90. Despite this heightened responsiveness in the F344 rats, a low-dose estrogenic effect on adult prostate weights was not observed. Thus, in the rat model a sustained effect at low doses of natural estrogens is not present in the prostate glands.
Tamoxifen	Uterine weight; Female/mice	761	2	E	At least 5 per group; 11 doses	Dose-releated increase in relative uterine weight in groups 1-5, decreased in group 6-11, but still increased compared to control. Thus, no anti-oestrogenic effect. Also, no statistics.

Chemical class	NMDR effect; Organ/sex/species	Refs.	Group	MoA and other relevant text (mainly as described in the paper)	No per group; no. doses	Further details and remarks
Plastics						
BPA	Fertility; Reproductive axis /female/mice	316	2		18-21 dams per group; 4 doses	Continous breeeding. Mainly significant effect at 25 µg/kg, i.e. highest dose. No sign of NMDR. However, also some effect on cumulative number of pups at 25 ng/kg, but not at 50 ng/kg. May be NMDR but could also be random
BPA	Reproductive behaviours; Behaviour/male/ rats	785	3		2-3 dams (12 male offspring) per group; 5 doses	Not corrected for litter effects and very few litters. NMDR for some endpoint of male mating behaviour, but not others.
BPA	Timing of vaginal opening, tissue organization of uterus; Reproductive axis/female/mice	577	2		Control 48 per group, BPA 5-17 per group; 8 doses	Inverse U-shaped curves for vaginal opening, significantly earlier at lowest and highest dose of 0,1 and 100 mg/kg, respectively. Those 2 groups had n= 5 and 6, respectively.
DEHP	Aromatase activity; Hypothalamus/ male/rats	788	2		11-12 litters per group; 11 doses	Brain HPOA: Decreased aromatase activity at 0,1 and 4 mg/kg, increased at 15, 45, 135 and 405 mg/kg in male offspring. Effect may not be adverse, so group 2.
DEHP	Timing of puberty; Reproductive axis /male/rats	789	1	These data suggest that elevated serum testosteron levels contributed to precocious preputial separation in the rats that were exposed to the low-dose DEHP.	10 per group; 4 doses	PD21-48, The age of preputial separation was 41.5 + 0.1 days postpartum in controls (vehicle). The 10 mg/kg DEHP dose advanced pubertal onset significantly to 39.7 + 0.1 days postpartum, whereas the 750 mg/kg DEHP dose delayed pubertal onset to 46.3+ 0.1 days postpartum. Similar picture for bw, seminal vesicle weight and serum T.
DEHP	Body weight at birth, vaginal opening, and first estrous; Female/rats	790	3		11-16 litters per group; 11 doses	No effects on birth weight. Pup weight PD1 in pups for necropsy varies, but N is smaller and litter effect most likely not considered. Vaginal opening: only delayed at the highest doses 15-405 mg/kg. First oestrus: no significant effects. Thus no signs of NMDR.

Chemical class	NMDR effect; Organ/sex/species	Refs.	Group	MoA and other relevant text (mainly as described in the paper)	No per group; no. doses	Further details and remarks
DEHP	Seminal vesicle weight, epididymal weight, testicular expression of steroidogenesis genes; Male/rats	791	2		4 per group; 4 doses	PD21-35. Absolute epididymis weight decreased at 10 mg/kg, but not at 500 and 750 mg/kg. Absolute seminal vesicle weight decreased at 10 and 750 mg/kg, but not 500 mg/kg. Small group size, i.e. 4 per group. Group 2 mainly because of small group size.
Detergents, surfactants						
Semicarbazide	Timing of preputial separation, serum DHT; Male/rats	796	2	Unbalance of steroid metabolism, interaction with ER and interference with CNS function at hypothalamic level. Delayed preputial separation at 140 mg/kg may be related to lower weight, i.e. be caused by general toxicity.	5 per group; 4 doses	Earlier preputial separation at 40 and 75 mg/kg, later at 140 mg/kg. DHT serum levels were significantly decreased at 40 and 75 mg/kg, but not at 140 mg/kg. Group 2 mainly because the effect at highest dose may be due to general toxicity.
UV filters						
Octyl methoxy-cinnamate	Activity, memory; Behaviour/both/rats	800	2	Decreased T4 and testosterone may be involved.	11-18 litters per group; 4 doses	Activity: increased in males only at 750 mg/kg at 17 weeks of age, but not at 9 weeks. Authors state that it may be a chance finding. Radial arm maze: decreased number of errors at 500 and 1000 mg/kg, but not at 750 mg/kg.
Aromatic hydrocarbons						
Toluene	Locomotor activity; Behaviour/male/rats	801	3	Dopamine-dependent, toluene at the highest dose (1200 mg/kg) attenuated the spontaneous motor movement and produced behavioural signs of intoxication including ataxia, which apparently interfered with forward locomotion, thereby resulting in fewer total photocell interruptions.	5-6 per group; 6 doses	Dose-dependent increase in activity from 600-900 mg/kg, some decrease at 1200 mg/kg. Not surprising as the animals become sedated. Group 3 due to toxicity at highest dose.

Chemical class	NMDR effect; Organ/sex/species	Refs.	Group	MoA and other relevant text (mainly as described in the paper)	No per group; no. doses	Further details and remarks
Phytoestrogens						
Genistein	Aggressive, defensive behaviours; Behaviour/male/ mice	811	1	Not associated with reduced testes size or decreased production of testosterone. May be combination of mechanisms related to sex steroid production and action. Or related to decreased maternal food intake and decreased pup weight during the lactation period, e.g. prenatal programming.	9-14 litters, 1 male per litter per group; 3 doses	Decreased body weight during lactation, mainly at 5 mg/kg. Not related to litter size as this was smallest at 5 mg/kg. Shorter AGD on PD21 at 5 mg/kg, but no effect on relative AGD. Increased defensive behaviour at 5 mg/kg, but not at 300 mg/kg. No effect on aggressive behaviour, but overall decreased aggression score, thus indication of demasculinization. No effect on mating behaviour, but not an optimal model.
Phytochemicals						
Phlorizin	Memory retention; Behaviour/male/ mice	814	2	Competitive inhibitor of glucose transport from blood to brain. Acting as a ‘‘glucose-like substance’’ although the mechanism(s) of this enhancement is unknown. Inverted U-shape is usual with numerous other memory-modulating treatments.	10 or 15 per group; 4 doses	Increased memory retention at 30 µg/kg only. Did not increase the retention latencies of mice that had not received a foot shock during training. Relevance for ED uncertain.
Herbicides						
Commercial mixture with mecoprop, 2,4-dichlorophenoxyacetic acid and dicamba	Number of implantation sites, number of live births; Female/mice	818	2	If the observation had been a true hormetic response, we would have expected an increase in litter size at the lower doses and not a decrease. In rodents, uterine receptivity to embryos is modulated by ovarian estrogen and progesterone. It is tempting to propose that some sort of endocrine modulation is mediating the effects, however, this proposal is speculative at this point.	31-63 litters per group from several studies. 4 doses	Litter size and implantation sites were significantly affected by dose, but resorptions were not significantly affected. Implantation sites and litter size in the very low and low doses both differed significantly from their control and high doses, respectively. The response varied from season to season. Group 2 due to uncertainties related to seasonal variation.

Chemical class	NMDR effect; Organ/sex/species	Refs.	Group	MoA and other relevant text (mainly as described in the paper)	No per group; no. doses	Further details and remarks
Simazine	Estrous cyclicity; Reproductive axis/female/rat	819	3	The significant decrease in cycle numbers in the first study with 25 and 100 mg/kg proved not to be significantly different with a longer dosing regimen.With such few cycles to examine as in study 1, we cannot determine whether the chemical is perturbing long-term cyclicity or if the delay in onset of VO is causing the temporary acyclicity, which is commonly observed after VO.	10 per group; 5 doses on PD 21-42 and 6 doses on PD 21-62	Delayed vaginal opening from group 3 (monotonic). Estrous cyclicity was generally affected at highest dose. For one of 3 endpoints, i.e. number of cycles, an effect was also found in group 3 (25 mg/kg) in the first study. The second study show a similar picture, but the lower number of cycles is not statistically significant. Group 3 as NMDR not seen with longer dosing and thus better data.
Insecticides						
DDT	Number of pups, sex ratios, neonatal body weight, male anogenital distance; Mice	777	2		6-10 per group; 6 doses	Number of pups per litter, sex ratio and pup bw: some signs of NMDR, but may be due to random variation.
Methoxychlor	Number of pups, anogenital distance (males and females), neurobehaviours (males and females); Mice	777	2	Changes in numbers of androgen receptors? Prostaglandin?	6-10 per group; 6 doses	Number of pups per litter increased in group 3 only. AGD decreased in both males and females in 2nd highest dose group, but increased at highest dose. Cliff avoidance latency increased at 2nd highest on PD 2, but not at PD 5. Righting reflex decreased at lowest dose on PD 2, but not PD 5.
Chlorpyrifos	Body weight; Male/rats	821	3		9-10 per group; 4 doses	Male body weight gain appears as showing NMDR with largest increase in the middle of the dose-response, but there are no statistics related to this and the effect is quite limited, i.e. max. 108% of control weight.

Appendix 3 – Sensitivity for finding threshold-like doses, based on power analysis

The power of a statistical test is the probability that the test will reject the null hypothesis when the null hypothesis is false (i.e. the probability of not committing a Type II error, hence the probability of not making a false negative decision on whether to reject a null hypothesis). In other words, power is the probability of finding a difference that does exist.

The probability of a Type II error occurring is referred to as the false negative rate (β). Therefore power is equal to $1 - \beta$, which is also known as the sensitivity.

Most researchers assess the power of their tests using 0.80 as a standard for adequacy which means that the probability for a false negative is less than 0.2. This convention implies a four-to-one trade-off between the probability of a Type II error and a Type I error, when 0.05 is selected as the value for statistical significance.

Power analysis can be used to calculate the minimum sample size required so that one can be reasonably likely to detect an effect of a given size. Power analysis can also be used to calculate the minimum effect size that is likely to be detected in a study using a given sample size.

A number of endpoints relevant for EDs provide quantal data, i.e. they are binary results providing a yes/no answer. Examples of quantal endpoints include malformations of reproductive organs (e.g. hypospadias), fertility index, pregnancy index etc. Results of histopathological evaluation may be reported either as yes/no answers or as distribution among scores from e.g. 0-3. Nipple retention is a yes/no endpoint if it is expressed as the number of males with or without nipple, but this endpoint can also be semi-quantitative, if the number of nipples is recorded (i.e. from 0 to 12).

Assessment of quantal endpoints is generally expected to lead to a lower power than assessment of continuous endpoint. To explore this, we have calculated the effect size needed for finding significant effect, i.e. $p < 0.05$, for yes/no endpoints (Table 2) and continuous endpoints (Table 3). This was done for studies with 8 litters per group, because that is the group sizes expected in the OECD TG 421/422 Reproductive Toxicity Screening study which is used for REACH testing at the tonnage level of 10 tpa. In addition it was done with 20 litters per group as this is the expected number per group in the OECD TG 416 and OECD TG 443. As the evaluation of some endpoints may be done in more than one offspring per litter, the calculations for the quantal endpoints also illustrate the effect sizes needed when 2 or 5 offspring per litter is assessed. However, the correct effects sizes needed for 2 or 5 animals per litter are likely to be higher than the ones shown as our calculation is based on the offspring as the statistical unit. To be correct, the calculations should be based on the litter as the statistical unit, i.e. the method should have corrected for litter effects. This was unfortunately not possible for us as there to our knowledge are no available easily used statistical programs for that purpose for quantal data.

Table 2. Effect sizes for quantal endpoints needed for p value < 0.05 in one-tailed Fisher Exact test*

Litters per group	Pups per litter	Group	No. with effect	No. without effect	Effect size	Increase of effect size compared to historical control of 0.3% for hypospadias
20	1	Control	0	20	**0%**	
20	1	Exposed	5	15	**25%**	8333%
20	2	Control	0	40	**0%**	
20	2	Exposed	5	35	**13%**	4167%
20	5	Control	0	100	**0%**	
20	5	Exposed	5	95	**5%**	1667%
8	1	Control	0	8	**0%**	
8	1	Exposed	4	4	**50%**	16667%
8	2	Control	0	16	**0%**	
8	2	Exposed	5	11	**31%**	10417%
8	5	Control	0	40	**0%**	
8	5	Exposed	5	35	**13%**	4167%

*The statistics used when more than one male pup per litter is included is based on using the pup as the statistical unit. Generally, the litter is considered as the correct statistical unit in developmental toxicity studies and using this approach will in most cases lead to even higher effect sizes than those shown in the table.

Quantal data

The results in table 2 show that for having a statistically significant effect with 20 litters per group the frequency of effect in the exposed group has to be 25% with 1 male per litter, 13% with 2 males per litter and 5% with 5 males per litter. With 8 litters per group the frequency of effect in the exposed group has to be 50% with 1 male per litter, 31% with 2 males per litter and 13% with 5 males per litter.

The frequency of hypospadias in humans is around 1 of 300, i.e. 0.3% (Moore 1983). Based on our historical control values for male external genital malformations, incl. hypospadias in Wistar rats, we actually find a similar frequency of 0.32% in rats (1 of 308). The effect sizes needed for finding a significant effect compared to this historical control frequency showed that the increase of the frequency has to be very large, i.e. 17-170 fold depending on the number of litters per group and the number of males studies per litter (Table 2).

We have in our calculations assumed that there were no offspring with hypospadias observed in the control group and this will with a historical control value for hypospadias around 0.3% be the majority of cases. However, in some few cases one hypospadias may occur in the control group and the effect sizes needed for finding a significant effect will become around 50% higher (data not shown) and the increase compared to the historical control value for hypospadias will be similarly increased to around 25-250 fold (data not shown). This very low sensitivity for detecting significant effects on rare adverse outcomes is generally recognized for malformations. Thus, the occurrence of

a few similar rare malformations such as hypospadias may generally be considered toxicologically relevant although the finding is not statistically significant.

For other quantal endpoints such as histopathology, fertility index, pregnancy index etc. one or a few cases may occur in the control group and there will often only be one data point per litter. Thus for these endpoints, the effect size needed for finding a significant effect may often be around 50% higher than those in table 3, i.e. range from 25-37% with 20 animals per group and from 50-75% with 8 animals per group. This very low sensitivity is often not considered when evaluating such endpoints.

Overall, the results illustrate that the effect size for a quantal effect has to be high in studies with 20 litters per group i.e. 25-37% and even higher with only 8 litters per group, i.e. 50-75%. This implies that the sensitivity for detecting quantal effects is very low and that effects sizes of human relevance may be present at the NOAEL.

Continuous effect data

The statistical power for detecting significant effects on continuous endpoints depends on the group size and the coefficient of variation (CV) in the control group. Generally, 80% or higher power is regarded as sufficient. The CV is a normalized measure of dispersion of a probability distribution. It is also known as the variation coefficient. The CV is also sometimes known as relative standard deviation (RSD), which is expressed as a percentage, i.e. it is calculated as the sample standard deviation divided by the sample mean and multiplied by 100. The results of power calculations based on different groups sizes and CVs using the shareware program C*3.1.3 are shown in table 3.

For AGD, the CV is normally around 5- 7% and an effects size of ca. 4-7% will in most cases be statistically significant in studies with N=20 per group, whereas the effect size has to be 7-11% if only 8 litters per group are studied. For endpoints with higher CV's such as 10% or 12%, the effects sizes also have to be higher, i.e. 9-11% for 20 litters per group and 15-18% for 8 litters per group. These results illustrate that the effect sizes needed for detecting effects with N=8 is approximately 1.6-1.8 times higher than when N=20.

Continuous effect data are generally expected to be more sensitive than quantal data and this is also found here as the effect sizes needed for continuous data range from 4-18%, whereas the effect sizes for quantal data in the previous section was found to range from 25-75%.

Table 3. Effect sizes for continuous data needed for power > 79% and with p<0.05

CV	N=20	N=8	Ratio
4.5%	4%	7%	1.8
7.1%	7%	11%	1.6
10.0%	9%	15%	1.7
12.0%	11%	18%	1.6

CV = Coefficient of variation (standard deviation/group mean*100 for control group)