



European Centre for Ecotoxicology and Toxicology of Chemicals (ECETOC)

European Chemical Industry Council (CEFIC)

European Crop Protection Association (ECPA)

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

*prepared by European Food Safety Authority (EFSA) and European Chemicals Agency (ECHA) staff
with support from the Joint Research Centre (JRC)*

Brussels, July 2018

Introduction

Since the publication of an outline of a Guidance for the identification of endocrine disruptors (EDs) on 20 December 2016 (ECHA and EFSA, 2016), ECETOC, CEFIC and ECPA commented on all stages of the development of the present Guidance. In March 2017, ECETOC published a Technical Report *Seven Steps for the Identification of Endocrine Disrupting Properties* (ECETOC 7SI-ED; ECETOC, 2017) that follows the structure of the ECHA and EFSA (2016) outline. Representatives of CEFIC and ECPA continuously participated in the relevant ECHA and EFSA *ad hoc* working groups, and during all public consultations on the draft versions of the Guidance, ECETOC, CEFIC, and ECPA jointly submitted detailed formal comments.

We welcome that the final Guidance is considerably improved as compared to the earlier draft versions. Nevertheless, a number of major concerns remain. We believe the Guidance encompasses profound inconsistencies and does not stand in line with the plant protection products (PPP) and biocidal products (BP) legislations, the legal text of the ED criteria (Commission, 2017), or the current 'state of the science' understanding of endocrine biology.

The Guidance does not reflect the legal text of the ED criteria (Commission, 2017):

The legal text requires that all elements of the ED criteria are met. Therefore, where any one of these criteria are not met, the only legal conclusion can be that the substance under investigation is not an ED.

Specifically, we have the following seven major concerns, explored in further detail below:

- The concept of 'EATS-mediated' parameters as assuming an endocrine mode-of-action (MoA) by default contradicts the ED criteria (major concern no. 1).
- The definition of a minimum data set to identify if a substance is an ED encourages unnecessary vertebrate animal testing. Further, it contradicts the concepts of the weight-of-evidence (WoE) evaluation and the use of all available data as per legal PPP or BP data requirements (major concerns no. 2-3).
- The level of proof to conclude that a substance is an ED is very low. In stark contrast, it is almost impossible to demonstrate that an effect results from a non-endocrine MoA. This is inconsistent with the ED criteria that do not require demonstration of a non-endocrine MoA (major concerns no. 4-5).
- Evaluation of population relevance (and non-relevance) is a critical component of the ED criteria for non-target vertebrates, but it is entirely dismissed in the Guidance (major concern no. 6).
- Appendix A of the Guidance concerning thyroid disruption neither represents the state-of-the-art nor does it provide workable guidance on how to assess thyroid-related effects in the context of the legal text of the ED criteria (major concern no. 7).

On account of these inconsistencies, the Guidance has a high probability of leading to the incorrect classification of substances as EDs. Further, it is likely to introduce a 'tick-box' testing approach leading to unnecessary vertebrate animal testing. This also contradicts the WoE approach clearly expressed in the scope of the Guidance and the legal PPP or BP data requirements.

The legal identification of an ED is to be based upon '*all available relevant scientific data*' (Commission, 2017). The Guidance is not consistent with this requirement since it suggests that 'EATS-mediated adverse' effects are sufficient to conclude on endocrine disruption. Such assumptions contradict the state of the science of endocrine biology. Therefore, they are invalid unless there is evidence to demonstrate that the adverse effect and the endocrine activity are biologically plausibly linked by a specific endocrine MoA. An ED cannot be '*presumed*' to be identified in the absence of evidence (*cf.*, e.g., p. IX¹, definition for ED criteria, that contradicts the legal text of the ED criteria).

Similarly, if all available relevant scientific data are considered not sufficient to meet the ED criteria, this should lead to the conclusion that the substance is not an ED. The implementing Regulations allow for additional data/studies to be required following consultations between regulators, risk assessors (national competent authorities) and applicants. Nevertheless, it would be disproportionate (and further contradict animal welfare requirements) to expect every applicant to update a previously submitted dossier, which complied with data requirements at the time.

In summary, we are concerned that the Guidance has general inconsistencies and lacks consideration of the state of the science of endocrine biology. Therefore, it does not fit the purpose of identification of endocrine disrupting properties as defined in the legal text of the ED criteria (Commission, 2017).

1. The term '(o)estrogen, androgen, thyroid, steroidogenic (EATS)-mediated' is confounded with the identification of adversity and mistaken assumption that such parameters exclusively identify an endocrine MoA. This results in an unjustifiably low threshold for the identification of EDs.

The flawed concept of 'EATS-mediated' parameters and the way the term is applied (i.e. the assumption of a pre-established link between adversity and an endocrine modality) are out of line with the current understanding of endocrine biology or with the legally implemented three-element ED criteria. This results in an unjustifiably low threshold for identification of endocrine disrupting properties. The distinction between 'EATS-mediated' and 'sensitive to, but not diagnostic of EATS' (STBNDO-EATS) parameters is arbitrary and not founded on the state of the science of endocrine biology. The Guidance itself acknowledges that, while the 'EATS-mediated' parameters may often be plausibly linked to (indicative of) endocrine activity, this is not exclusively the case (*cf.* p. IX, definition for 'EATS-mediated': '*in the absence of other explanations*'). Consequently, the so-called 'EATS-mediated' parameters are also only sensitive to, but not diagnostic of EATS, i.e. STBNDO-EATS. Consistent with the legal text of the ED criteria, there is as spectrum of STBNDO-EATS parameters that have a varying range of plausibility depending on all the data available.

While the definition for 'EATS-mediated' does include the qualification '*in the absence of other explanations*' (p. IX), there is no guidance for how to establish the 'presence' of such other explanations. Moreover, it is stated (p. 41; Section 3.5.2; Sub-section 'EATS-mediated' adversity):

...when an adverse effect is 'EATS-mediated' the biologically plausible link is already pre-established in the absence of information proving the contrary (i.e. a fully developed non-ED MoA).

¹ Note: all page numberings refer to the text of the Guidance.

In practice, this will most likely constitute an unsurpassable hurdle. While we see the technical benefit of not needing to develop a detailed MoA in all cases, introducing the concept of a pre-established link between an 'EATS-mediated' effect and an endocrine activity is contrary to the MoA analysis approach. It will be highly subjective to 'prove' that such a pre-established link exists, and this will undoubtedly lead to very conservative conclusions.

The assessment strategy is based on the three conditions stipulated in the ED criteria (adversity, endocrine activity, and a biologically plausible link between the two) and on the grouping of the parameters as described above. The 'EATS-mediated' parameters listed in the OECD GD 150 drive the assessment strategy because, by providing evidence for both endocrine activity and the resulting potentially adverse effects, they are considered indicative of an endocrine MoA. (p. 9, Section 3.1.3, 1st para.; similarly, p. 31; Section 3.4)

This statement, implying that 'EATS-mediated' parameters exclusively identify an endocrine MoA, is a misinterpretation of OECD Guidance Document (GD) No. 150. OECD GD 150 lists those effects which should be seen if a substance has an EATS-mediated activity. If such effects are not seen, then the substance does not have the particular EATS-mediated activity. However, if these effects are seen, then an EATS-mediated activity is only one potential cause for their occurrence, whereas the substance might also have elicited the effect via a non-endocrine MoA. This becomes clear from the overall contents of the OECD GD 150, as the following example taken from the table on page 101-102 of the OECD GD 150² shows. Referring to the extended one-generation reproductive toxicity study (EOGRTS; OECD Test Guideline (TG) 443), this table lists "changes in oestrus cyclicity, decreased age at vaginal opening" and "histopathologic changes in vagina, uterus (+ cervix), ovaries, testis, epididymis, prostate, seminal vesicles and coagulating glands" as endpoints for agonistic oestrogen-mediated activity. Clearly, all of these effects can also be caused by non-endocrine mechanisms, such as systemic toxicity.

The assumption that 'EATS-mediated' parameters are evidence for both adversity and an endocrine MoA and represent a 'pre-established link' between adversity and endocrine MoA implies that the ED criteria apply by default based on partial information. This contradicts the legal text of the ED criteria, and it is out of line with current scientific knowledge. The assessment strategy described on p. 9, (*cf.* quote above from Section 3.1.3, 1st para.) is likely to result in significant over-categorisation of compounds as EDs, and the requirement to establish a 'fully developed non-ED MoA' (*cf.* quote above from p. 41; Section 3.5.2; Sub-section 'EATS-mediated' adversity) will compound this over-categorisation and is likely to result in increased vertebrate animal testing. Where evidence of any one of the elements of the legal text of the ED criteria is absent, it can only be concluded that the ED criteria are not fulfilled so that an ED has not been identified.

Finally, we welcome that the Guidance explicitly states that one 'EATS-mediated' parameter in isolation does not suffice to identify an ED (p. 19, Section 3.3.1; p. 22, Section 3.3.1; p. 31-32, Section 3.4.1). In following up on the example of jointly evaluating hypospadias, anogenital distance and nipple retention (p. 42, Section 3.5.2, Sub-section 'EATS-mediated' adversity), an isolated change in one parameter is insufficient to conclude 'EATS-mediated' adversity. For example, if anogenital distance was decreased in male rats but there were no other changes relating to an anti-androgenic MoA, this would be insufficient to conclude A-mediated adversity.

² Revision of OECD GD 150; draft version V3 of December 2017, available at: https://www.oecd.org/chemicalsafety/testing/GD150_2017%20v3%2006122017b_clean.pdf.

2. The Guidance gives the impression of introducing new data requirements. However, these are set in the respective substance-specific legislation. The legal text of the ED criteria mandates a WoE evaluation using the available data. It is beyond the remit of the Guidance to alter legislative requirements.

The Guidance repeatedly and explicitly states that data need to be generated. This is beyond the remit of the Guidance as it would require amendments to the relevant data requirements for PPPs and BPs (e.g., Commission (2013)). It also contravenes the concept of a WoE evaluation of all the available evidence.

Exemplary references where the Guidance states that data need to be generated:

- p. 37, Section 3.4.4.2, Sub-section Scenario 2a, iii: *If the endocrine activity has not been sufficiently investigated (see section 3.4.2), it is necessary to generate further information.*
- p. 42, Sub-section 'Adversity based on STBNDO-EATS parameters': *in these cases, it is likely that further empirical data will need to be generated e.g. level 3, 4 and/or 5 on the substance under evaluation to demonstrate the link between the observed adverse effect and an endocrine MoA.*
- p. 44, Section 3.5.3, 2nd para.: *On a case-by-case basis, when adversity is indicated by 'EATS-mediated' parameters, and the conclusion on the biological plausibility for the link between adverse effects and endocrine activity for the postulated MoA is challenged by the applicant (...) further data must be generated, in order to substantiate the alternative non-endocrine MoAs.*

In the context of the BP legislation, it is expected that many non-active substances are relatively data-poor. The BP Regulation Guidance (Competent Authorities, 2018) foresees that detailed evaluation should only be conducted if there are significant indications that a non-active substance may have endocrine disrupting properties based on the existing knowledge and the available scientific information. However, the Guidance provides no specifications on what these 'indications' should be. Thus, the interpretation amongst applicants and various evaluating authorities will most likely vary. We are concerned that significant potentially unjustified testing requirements will be routinely triggered for non-active substances. This directly contradicts statements in the Guidance regarding the prevention of unnecessary animal testing (p. 32, Section 3.4.3, 1st para.). We agree that the applicant and risk assessor should cooperatively agree on sufficient information in order to avoid unnecessary animal testing. The ECHA ED Expert Group is advised to play a key role, providing oversight and advice regarding the need for additional data generation. Unfortunately, the Guidance unnecessarily limits the opportunity for alignment negotiation amongst risk assessors and applicants on the conclusion of sufficient evidence to achieve the 'not met' criteria (Section 3.1: p. 7; Sections 3.4.1 and 3.4.2). Flexibility is required to explicitly facilitate this option.

As mandated in the legal text of the ED criteria, all available data are considered in a WoE evaluation. This should also apply to data from *in vitro* assays (p. 33, Section 3.4.3). Typically, *in vitro* assays are intended to exhibit high sensitivity but often lack specificity. The use of 'all available relevant scientific data' in the ED criteria is understood to allow the inclusion of data from tools available today and those approaches expected to emerge in due course, e.g. population models for non-target organism assessment or further *in vitro* mechanistic assays (EATS or non-EATS). The consideration of available albeit atypical data would normally be captured in the WoE and its (non-)validity justified by the applicant / risk assessor. This is similar to the US EPA 'OSRI' approach ('other scientifically relevant information' approach; Bishop and Willett, (2014)). Notwithstanding, an option to generate data at a higher level of the OECD Conceptual Framework (CF) to evaluate a signal arising from an *in vitro* assay should always be available.

In summary, any need to perform supplementary studies shall be discussed with the national competent authorities (*cf.* e.g., the data requirements for PPPs (Commission, 2013)). The Guidance is simply

supportive information and not legally binding. Proscriptive statements around the requirements to conduct, or not to conduct, additional testing as included in the Guidance undermine discourse around testing needs between applicants and the national competent authorities based upon the available substance-specific databases. This is likely to result in compounds being miscategorised as EDs, and in increased vertebrate animal testing.

3. There is a disconnect between the concept of a WoE evaluation and the focus on a select number of specific apical higher-tier studies. Attempting to meet the definition of ‘sufficient data’ as provided in the Guidance runs the risk of leading to a huge amount of animal testing with no or only minor improvement in human health or environmental safety.

We welcome that both OECD CF Level 5 mammalian studies (OECD TG 416 and 443) are considered adequate for the assessment of endocrine disrupting potential in mammals (p. 31, Section 3.4.1, 2nd para.). The statement in this same paragraph that both the fish full life-cycle test (FFLCT; with endocrine endpoints) and the medaka extended one generation reproductive toxicity test (MEOGRT) are acceptable is also welcomed. This is especially relevant since the animal-intensive MEOGRT is very difficult to conduct and poorly validated, as shown by a recent analysis of the published validation studies (Salinas and Weltje, 2018).

However, we are concerned that the Guidance stipulates a minimum battery of tests (and parameters to be considered) to provide sufficient evidence for lack of ‘EATS-mediated’ adversity (Section 3.4.1, p. 31) or endocrine activity (Section 3.4.2, p. 32), i.e. to allow the conclusion that the ED criteria are not met. In this regard, it also contradicts animal welfare requirements that the provisions for the EOGRTS include a ‘by default’ F₂ generation.

Generally, the sufficiency of any dataset should never be assessed applying a ‘tick-box’ approach. Instead, it should be assessed on a case-by-case basis applying a WoE evaluation based on expert judgement.

Similarly, there should be no general requirement that a sufficient dataset can only be achieved if studies have been performed according to the latest version of a specific TG. The Guidance implies that new testing may be required if reproductive toxicity studies have been performed in accordance with outdated versions (p. 14, Section 3.2.1.1, 2nd para.):

In the context of the ED hazard identification, new studies should be carried out according to the latest version of the corresponding TGs to be considered fully relevant. This is of particular importance since in recent years a number of TGs have been revised to include additional parameters which are relevant for identification of ED properties (e.g. the latest version of OECD TG 416 (OECD, 2001b)).

Further, the Guidance seems to preserve the option to repeat sub-acute, sub-chronic or chronic tests which have been conducted based on earlier versions of the OECD TGs, even in the absence of specific concerns (p. 31, Section 3.4.1). This is not justified and stands in conflict with animal welfare requirements. Even if existing studies conducted according to outdated versions of TGs have deficiencies compared to studies conducted according to the most recent versions, such deficiencies may be overcome by considering endpoints from other studies in the WoE evaluation. Otherwise, a precedent is set for repeat studies to be generated every time a TG is revised.

Specific to the EOGRTS (OECD TG 443), this test method was designed to reduce animal use. Requiring a ‘by default’ F₂ generation undermines the purpose of this study design and contravenes the concept of a WoE evaluation or the legal mandate to minimise vertebrate animal testing. The default inclusion of the F₂ cohort in the EOGRTS is not the standard design in any jurisdiction or any substance-specific EU legislation. Therefore, very few compounds are likely to have this information.

In summary, prescriptive statements specifying the data which are (and are not) required to identify an ED undermines the mandated WoE concept and substance-specific discussions around testing needs between applicants and the national competent authorities. This is likely to result in substances being miscategorised as EDs, and in unjustified increased vertebrate animal testing.

4. The Guidance requires an unreasonable burden-of-proof to demonstrate that a substance is not an ED. This contradicts the legal text of the ED criteria where available data are used to determine if a substance is an ED.

The definition of ED criteria in the Guidance (p. IX) that these criteria allow “*for the identification of both known and presumed endocrine disrupting substances*” contradicts the legal text of the ED criteria that prescribes that these criteria are used to identify if a substance has endocrine disrupting properties.

The Guidance has a focus on showing that a substance is not an ED (e.g., p. 31, Section 3.4, last para.; p. 32; Section 3.4.2, end of 1st para.) and sets an unreasonable burden-of-proof to demonstrate that this is the case. In this regard, the burden-of-proof is undefined and potentially impossibly high for the determination of MoAs demonstrating that endocrine disrupting effects are secondary to other toxicities (p. 21, Section 3.3.1.1., bullet point 2).

The low level of evidence required to conclude that a substance meets the ED criteria (*cf.* major concern no. 1) stands in striking contrast to the approach that has been introduced for alternative non-endocrine MoAs (p. 43, Sub-section ‘Alternative non-endocrine MoA’):

In cases where an applicant considers to postulate an alternative non-endocrine MoA for adverse effects based on ‘EATS-mediated’ parameters, the level of empirical support and biological plausibility would need to be very strong to demonstrate that the alternative MoA was the more likely explanation of the adverse effects observed. In such cases a comparative MoA analysis will need to be applied when postulating and substantiating an alternative non-endocrine MoA (Meek et al., 2014b) (see Section 3.5.2). Such an alternative non-endocrine MoA may be postulated where the potentially endocrine-related adverse effects are considered secondary to other non-endocrine related toxic effects (see Section 3.3.1).

While this paragraph does not provide much guidance how the evaluation for alternative non-endocrine MoAs should be conducted and documented, in the subsequent Section it is stated that further data must be generated if the applicant challenges an endocrine MoA (p. 44, Section 3.5.3, 2nd para.):

On a case-by-case basis, when adversity is indicated by ‘EATS-mediated’ parameters, and the conclusion on the biological plausibility for the link between adverse effects and endocrine activity for the postulated MoA is challenged by the applicant (...) further data must be generated, in order to substantiate the alternative non-endocrine MoAs (p. 44, Section 3.5.3).

In contrast to this statement, an effect on an ‘EATS-mediated’ parameter in the presence of evidence showing no effect in a complete battery of EATS mechanistic studies is clear evidence that an EATS MoA is not operant and that the ED criteria are not met.

...where an endocrine MoA is considered not to be relevant for humans, absence of other/concomitant endocrine MoAs leading to the same adverse effect in humans should also be excluded. (p. 48, Section 3.5.4.4, 2nd para.)

The underlined part of this statement appears to imply that the applicant should investigate all other possible MoAs which may lead to the same effect in humans. Due to the complexity of the endocrine system and its many interactions and cross-talk, this will most likely result in the need to conduct a broad

spectrum of studies to investigate other possible relevant MoAs - without ever coming into a position to demonstrate that all have been covered and claim with absolute certainty than any other MoA can definitively be ruled out. As such, the quoted statement is unhelpful and raises questions around data adequacy and burden-of-proof that are not addressed in the Guidance. A 'sufficient' dataset showing no evidence of endocrine activity should be adequate information to conclude that a substance does not meet the ED criteria.

Moreover, it is stated (p. 41, Sub-section 'EATS-mediated' adversity):

When an adverse effect is 'EATS-mediated' the biologically plausible link is already pre-established in the absence of information proving the contrary (i.e. a fully developed non-ED MoA).

A WoE assessment of all available data is the heart of the ED assessment (*cf.* major concern no. 1), and it should not be 'over-ruled' with such pre-established assumptions.

In contrast to the very high burden-of-proof required in the Guidance to demonstrate that a substance is not an ED, when assessing the link between the adverse effect and the endocrine activity, the Guidance states that it is sufficient to demonstrate only one ED-related key event in order to identify an ED, and not the full set of key events or key event relationships (p. 39, 1st paragraph below the grey box). Further, if multiple (non-endocrine and endocrine) MoAs have been postulated, and the available evidence does not allow deciding on a specific pathway, the Guidance states that the substance should be considered an ED if E, A or S are highly likely (p. 43, 1st para.).

These exemplary statements from the Guidance further highlight the low level of evidence required to conclude that a substance meets the ED criteria stands as compared to the almost unsurpassable burden-of-proof that has been introduced to demonstrate non-endocrine MoAs.

The Guidance states that the identification of biological plausibility carries more weight than empirical observation (p. 39, grey box). Such a 'by default' statement has a high probability of reducing the reproducibility of assessments by focusing primarily on whatever is thought to be biologically plausible, at the expense of other relevant Bradford-Hill criteria, including a lack of empirical support, that may suggest otherwise. Such an approach confounds scientific assessment by magnifying the impact of uncertainties. To prevent an inconsistent application of the World Health Organisation/International Programme on Chemical Safety (WHO/IPCS) MoA/human relevance framework (Meek et al., 2014), the steps of this framework should be strictly adhered to using all available data.

We strongly disagree with the point on human relevance in the Guidance where it is indicated that as the goal of the Guidance is hazard identification, rather than risk assessment, qualitative species differences, but not quantitative differences, should be the focus of human relevance assessments (p. 48, Section 3.5.4.4, 1st para.). The established science for assessment of the human relevance of toxic effects clearly states that an understanding of both qualitative and quantitative species differences is critical when evaluating human relevance (Meek et al., 2014). The key event relationships between two key events almost universally requires a threshold to have been exceeded for one key event prior to progression to the key event in the pathway (Boobis et al., 2006). In accordance with the state of the science in this area, both quantitative and qualitative considerations should be evaluated when examining the human relevance of an effect in a laboratory animal species.

5. The guidance on weighing lines of evidence is not coherent. This runs the risk of resulting in irreproducible and contestable conclusions on the presence or absence of an endocrine disrupting property.

The guidance on weighing lines of evidence is inconsistent in the different Sub-sections to 3.5.4 (beginning on p. 44). Three-scale ratings are specified to evaluate biological plausibility, empirical support and essentiality as ‘strong - moderate - weak’ as compared to the derivation of a binary ‘yes / no’ conclusion on the presence of an ED (p. 52, Section 3.6).

Further, biological plausibility can only be assessed as ‘weak’ if the relationship between key events is ‘not understood’ (p. 45, Section 3.5.4.1, final bullet point). This is not meaningful. It should be possible to assign ‘weak’ if such relationships are most likely absent. Finally, Sections 3.5.5 (‘Extent of support for the overall assessment of the MoA analysis’; p. 48) and 3.5.6 (‘Conclusion on the MoA analysis’; p. 49) provide no guidance for how to integrate the different strong - moderate - and weak evidences for biological plausibility, empirical support and essentiality to derive a conclusion on MoA analysis or the presence of an ED.

6. The Guidance does not describe how population relevance of effects should be assessed in non-target species. Instead it resorts to a circular reasoning that the definition for adversity considers population relevance.

The Guidance fails to provide guidance on how an applicant can assess the population relevance of adverse effects in non-target vertebrates. We believe this is contradictory to the legal text of the ED criteria and the scientific basis of the definition of an ED in non-target organisms. There is no guidance on how population modelling and field studies may be used in such an assessment. Further, the Guidance misquotes several citations that are presented in support of why population modelling cannot be used (Marty et al., 2017; Matthiessen et al., 2018; Mintram et al., 2018); p. 17, 1st para.) when in fact these references support the use of population modelling to address the population relevance question in endocrine assessments.

The identification of an ED should leave the option for sound scientifically reasoned assessments based on population modelling and/or field studies to be presented to and evaluated by competent authorities.

The Guidance states that population modelling can be used to support an approval and

should be considered as part of the information for ED hazard identification (p. 14, Section 3.2, last para.).

However, this is then contradicted, and confusing statements concerning population modelling have been introduced into the Sub-section ‘Field studies, monitoring, and population modelling’ (p. 16):

...however, generally these models are more suitable for risk assessment purpose.

There is scientific evidence clearly showing that this is not necessarily the case. Population models are independent from the context in which they can be used. For example, they can be used to describe the effect of various stressors, such as a predator or, alternatively, a chemical. Thus, population models can operate without accounting for exposure (i.e., for hazard-based assessment). Therefore, models can predict population trajectories by using variations in the reproduction rate regardless of the cause of such variations. Stating that “*models are more suitable for risk assessment purpose*” does not make sense.

In conclusion, while the mentioned tools are considered promising, they currently cannot be used to dismiss the population relevance of an adverse effect without further investigating the link between

the effects observed in laboratory test and the population dynamics (Marty et al., 2017; Matthiessen et al., 2018; Mintram et al., 2018). (p. 17, 1st para.).

While the meaning or intention behind this conclusion are unclear, it is not supported by the cited references that instead support the use of modelling for ED population relevance assessment. Population models, however, do not replace the need to establish a causal link between adverse effects observed in a laboratory test and the drivers of population dynamics.

Further ambiguity has been introduced by the statement on p. 20 (Section 3.3.1, final 2 paragraphs):

When assembling and assessing the line of evidence, any available epidemiological studies should be considered as supportive evidence for the evaluation of whether an ED is likely to have adverse effects for humans. However, they cannot be used to override or dismiss evidence of adversity found in laboratory studies, nor can they replace laboratory studies. Similarly, when assembling the lines of evidence for non-target organisms any field and monitoring studies and population modelling can be considered as supportive evidence.

It is not clear what the word 'similarly' implies in the final sentence. Which conclusion is drawn if the field, monitoring, or population modelling evidence supports the view that there is no effect? Taking into account the overall WoE of the available data, it should be possible to draw a conclusion from these types of studies that a substance is not an ED for non-target vertebrates or that the effects are not relevant at the population level.

Instead it is stated (p. 22, 1st full para.):

Therefore, the relevance of such effects at the population level should be assumed when determining the adversity in the absence of appropriate scientific data demonstrating non-relevance.

Taking into account the above statement from p. 20, Section 3.3.1, this is a self-fulfilling prophecy, since the Guidance assumes that there cannot actually be any appropriate scientific data demonstrating non-relevance because all alternatives have been ruled out.

Finally, on p. 22 (Section 3.3.1.4, last para.), it is described that thyroid histopathological findings without impairment of growth/development and/or reproduction observed in the rat are likely not relevant at population level. It is unclear if this implies that the additional rat developmental neurotoxicity study should be always conducted to determine the adversity for non-target mammals. The Guidance seems to indicate that any change in thyroid histology is considered adverse for humans. Many substances on the market may cause thyroidal histopathological effects, and a qualitative difference of liver enzyme induction between humans and rodents may be difficult to prove (*cf.* major concern no. 7).

7. Appendix A of the Guidance singles-out concerns for thyroid toxicity and the role of liver enzyme activation without a clear rationale for this focus. This Appendix is inaccurate and impractical; it does not reflect the state of the science and provides no useful guidance to support MoA evaluations following observed thyroid effects.

The Guidance, and specifically, Appendix A, does not reflect the state of the science of thyroid toxicity. Dubious statements around human relevance of thyroid toxicities in rodents are made without considering differences in species sensitivity. The Guidance proposes a testing scheme which does not describe essential elements of the methodology for thyroid MoA testing, or how the data generated from this testing should be evaluated to reach a conclusion on whether the substance under consideration meets the legal definition of an ED. Whilst the human hypothalamic–pituitary–thyroid (HPT) axis is qualitatively similar to that of laboratory animals, significant quantitative differences exist. Even though

Appendix A briefly mentions that quantitative differences in species responsiveness should be considered, it does not explain what would constitute a sufficient quantitative difference to establish human non-relevance.

In addition to this overall concern with respect to Appendix A, we have the following specific concerns:

P. 110: We disagree that the EC 2017 workshop (EC, DTU, Brunel University, 2017) is an appropriate reference for the ‘*current understanding of thyroid physiology and toxicology*’. This is well beyond the scope of this workshop, whose objectives which were:

To address and discuss interpretations of experimental laboratory studies, wildlife field data as well as human epidemiological data in relation to the identification of thyroid disrupting substances; and to identify ways forward in addressing potential gaps in the test methods in relation to identification of thyroid disrupting substances (EC, DTU, Brunel University, 2017).

P. 110-111: In the three bullet points regarding the interpretation of data from experimental animals, general assumptions (‘*would pose a hazard*’, ‘*would still present a potential concern*’) are set as foundation for the interpretation of the available data. Such general assumptions contradict any meaningful science-based WoE approach that requires that the available evidence is interpreted on a case-by-case basis. Further, the statement of the 3rd bullet point (p. 111):

In the absence of substance-specific data which provide proof of the contrary, humans and rodents are considered to be equally sensitive to thyroid-disruption (including cases where liver enzyme induction is responsible for increased thyroid hormone (TH) clearance).

is inconsistent with well-established understanding of the differences in species sensitivity to TH perturbation, and it is contrary to how assessments have been conducted for many years with no real justification for a change in approach. Indeed, this statement is also inconsistent with an earlier statement in the Appendix A (p. 110, 3rd para. of Background):

Although the HPT axis and the basic physiological processes regulating TH synthesis and release are qualitatively similar across species, there are, however, quantitative species-specific differences (Janssen et al., 2017).

P. 111: The Section ‘Investigation of increase in thyroid hormone metabolism in the liver’ is likely to lead to an increased expectation of ‘routine’ data exploring of the role of liver in contributing to any changes in thyroid parameters (hormones or histopathology). Further, it is stated (p. 111, bullet point 2):

Comparative studies of enzyme activity induced by the test substance in liver in vitro systems should be measured in both the relevant test species (e.g. rat, mouse and dog) and humans.

This recommendation seems premature since enzyme induction is difficult to replicate in *in vitro* cell assays (variability between human donors), especially for the phase II inactivating enzymes (Soars et al., 2004; Li et al., 1999; Ritter et al., 1999). These data may not appropriately address the question of human relevance. To assess test substance-induced liver enzyme activity *in vitro* may be difficult because there are currently no validated TGs for this purpose. The Guidance does not address the following questions: What are the target enzymes to be investigated? What are the criteria for determination? A concrete approach to investigate species differences is required including toxicokinetics considerations. The data proposed to establish whether enzyme induction is responsible for thyroid effects seems to exceed the requirements to demonstrate this MoA. Nevertheless, it will remain difficult to prove that effects secondary to liver enzyme induction are not endocrine disrupting due to the lack of detail in Appendix A.

A decrease in T4 (total or free) in the absence of adverse histological changes should act as a trigger for further studies. It is known from the broad knowledge of biology (e.g. human clinical experience

and epidemiological data) that a drop in T4 results in impaired pre- and post-natal neurological development (Alshehri et al., 2015). (p. 111-112)

We disagree that the totality of the evidence demonstrates that a drop in T4 alone results in impaired neurological development. In fact, in the cited paper (Alshehri et al., 2015) the influence of transthyretin on T4 delivery is reviewed. Similarly, Morreale de Escobar et al. (1992) demonstrated that the foetus is relatively resilient to fluctuations in the supply of T4, as fetal brain T3 homeostasis tends to be maintained due to several regulatory mechanisms. Apparently, the Alshehri paper has been misunderstood in the Guidance. Other than what is summarized in Appendix A, a greater understanding of the point of departure (or degree of change needed) for T4 in the maternal animal and/or fetal plasma (and postnatal animal) that results in decreased active hormone (T3) in the brain is needed to determine if T4 changes in the absence of histopathology should act as a trigger for further studies.

It is still unclear which alternative MoA(s) should be excluded for compounds suspected to alter thyroid parameters through induction of hepatic enzymes. Appendix A recognizes that there are currently insufficient tools to fully address all the potential MoAs included in the thyroid modality and states that a *'reasonable effort is anticipated'* (p. 111, bullet point 3) but this is not well explained. A reasonable testing and assessment strategy remains to be developed.

In summary, the current Appendix A does not reflect the state-of-the-science of thyroid toxicology, and it is likely to result in the generation of uninterpretable data in vertebrate animals without facilitating decision-making with respect to thyroid toxicity.

References

Note: All websites were accessed on 25 June 2018.

Alshehri B et al. (2015). The diversity of mechanisms influenced by transthyretin in neurobiology: development, disease and endocrine disruption. *J Neuroendocrinol* 27:303-323.

Bishop PL, Willett CE (2014). The use and acceptance of Other Scientifically Relevant Information (OSRI) in the U.S. Environmental Protection Agency (EPA) Endocrine Disruptor Screening Program. *Birth Defects Res B Dev Reprod Toxicol* 101:3-22.

Boobis AR et al. (2006). IPCS framework for analyzing the relevance of a cancer mode of action for humans. *Crit Rev Toxicol* 36:781-92.

Commission (2013). Commission Regulation (EU) No. 284/2013 of 1 March 2013 setting out the data requirements for plant protection products, in accordance with Regulation (EC) No 1107/2009 of the European Parliament and of the Council concerning the placing of plant protection products on the market. *OJ L* 93, 85-152.

Commission (2017). Commission Delegated Regulation (EU) 2017/2100 of 4 September 2017 setting out scientific criteria for the determination of endocrine-disrupting properties pursuant to Regulation (EU) No 528/2012 of the European Parliament and Council. *OJ L* 301(60), 1-12.

Competent Authorities (2018). Note agreed by Member States' Competent Authorities for biocidal products. The implementation of scientific criteria for the determination of endocrine-disrupting properties in the context of biocidal product authorization. CA-March18-Doc.7.3.b-final

EC, DTU, Brunel University (2017). European Commission, Directorate-General for the Environment, DTU National Food Institute Denmark; Brunel University London. Supporting the organisation of a workshop on thyroid disruption –Final Report (Framework Contract ENV.A.3/FRA/2014/0029 on implementation of the Community strategy on Endocrine Disrupters). In. Publications Office of the European Union, Luxembourg. Available online: <https://doi.org/10.2779/921523>.

ECETOC (2017). ECETOC Seven Steps for the Identification of Endocrine Disrupting Properties (ECETOC 7SI-ED). ECETOC TR 130, Belgium, March 2017; available at: <http://www.ecetoc.org/wp-content/uploads/2017/03/ECETOC-TR-130-7SI-ED.pdf>.

ECHA and EFSA (2016). Outline of draft guidance document for the implementation of the hazard-based criteria to identify endocrine disruptors, published 20 December 2016; available at: https://www.efsa.europa.eu/sites/default/files/documents/161220_ed_guidance_outline.pdf.

Kleinstreuer NC et al. (2017). Development and validation of a computational model for androgen receptor activity. *Chem Res Toxicol* 30:946-964.

Li AP et al. (1999). Effects of cytochrome P450 inducers on 17 α -ethinyloestradiol (EE₂) conjugation by primary human hepatocytes. *Br J Clin Pharmacol* 48:733–742.

Meek ME et al. (2014). New developments in the evolution and application of the WHO/IPCS framework on mode of action/species concordance analysis. *J Appl Toxicol* 34:1-18.

Morreale de Escobar G et al. (1992). Homeostasis of brain T3 in rat fetuses and their mothers: effects of thyroid status and iodine deficiency. *Acta Med Austriaca*: 19 Suppl 1:110-6.

Ritter JK et al. (1999). Expression and inducibility of the human bilirubin UDP-glucuronosyltransferase UGT1A1 in liver and cultured primary hepatocytes: evidence for both genetic and environmental influences. *Hepatology* 30:476–484.

Salinas ER, Weltje L (2018). The unforeseen consequences for animal welfare of the OECD TG 240 (MEOGRT) biological validity criteria. Presentation at the SETAC Europe 28th Annual Meeting, 13-17 May 2018, Rome, Italy.

Soars MG et al. (2004). An assessment of udp-glucuronosyltransferase induction using primary human hepatocytes. *Drug Metab Dispos* 32:140-148.