ENDOCRINE SOCIETY
Hormone Science to Health

Dear Members of the Guidance Document Drafting Group,

The Endocrine Society appreciates the opportunity to comment on the draft guidance for identifying endocrine disruptors. Founded in 1916, the Endocrine Society is the world's oldest, largest, and most active organization dedicated to the understanding of hormone systems and the clinical care of patients with endocrine diseases and disorders. Our membership of over 18,000 includes researchers who are making significant contributions to the advancement of knowledge in toxicology, especially in the field of endocrine-disrupting chemicals (EDCs).

We appreciate that the draft document is a well-written, thoughtful approach towards regulation of EDCs. However, during a careful review we noted several deficiencies, and we identified amendments that would improve the effectiveness of the guidance. Our detailed, section-specific comments have been submitted separately through the public consultation form. In this letter, we describe several general principles and overarching concerns with the document.

1. The Guidance Document must have a broad scope that covers the latest science on endocrine pathways and the biology of hormones.

We are concerned that the focus of the document is on the estrogen, androgen, thyroid, and steroidogenesis (EATS) pathways described in the OECD Guidance Document (No. 150) on Standardized Test Guidelines for Evaluating Chemicals for Endocrine Disruption. We appreciate that the guidance document anticipates the development of new OECD-approved test methods, but there are many endocrine systems (e.g., metabolism and other nuclear signaling pathways) that are not captured in the OECD Conceptual Framework for Testing and Assessment of Endocrine Disrupters. Especially in the context of periodic reviews, EFSA and ECHA should consider scenarios that are not included in GD 150. We also encourage EFSA and ECHA to use available data on EDCs in non-target invertebrates.

We strongly support the use of systematic review to improve decision making; however, the Guidance Document does not use established methods of systematic review to identify EDCs. Specifically, we note that Klimisch criteria would be used to evaluate guideline studies, while other studies would be evaluated with separate reliability/relevance criteria. A systematic review should instead evaluate all studies in a transparent manner using the same validated criteria, for example using the Navigation Guide or SYRINA methods.

2. Agencies should be able to identify chemicals with endocrine disrupting properties based on realistic standards of available scientific information.

The Guidance Document provides instructions to minimize the identification of false-positives, without a similar emphasis on the important goal of reducing false-negatives. This should be

2055 L Street NW
Suite 600
Washington, DC
20036

T. 202.971.3636
F. 202.736.9705

endocrine.org

addressed by reconsidering the Mode of Action Analysis chapter, which requires detailed information to arrive at a conclusion regarding plausible links between adverse effect and endocrine activity. The guidance document is also unclear on the regulatory consequences for situations where agencies have insufficient data for evaluating a chemical. We recommend that these chemicals be classified as potential EDCs, pending further evaluation. Furthermore, there exist throughout the document elements of hazard characterization that are not useful for the identification of EDCs; these should be removed.

We have reservations about the Adverse Outcome Pathway (AOP) approach in this regulatory context. AOPs are based on linear constructs that may not represent or capture essential complexities of endocrine signaling, predicated on our current understanding of biology, which will evolve based on new scientific information. We note that AOPs were not originally developed to assist in identifying EDCs and so may not be designed to fit this unique purpose. Furthermore, the data submitted by manufacturers will not be sufficient to characterize an entire pathway or complex interactions within a system. We assert that the existence of an adverse effect within an endocrine modality is the critical feature that the public is most concerned about. This should be sufficient to assume that a chemical is an endocrine disruptor; attempts to completely characterize the biological pathway associated with the adverse event are unnecessary for a scientific hazard identification process.

3. The Guidance Document should maintain the Annex on thyroid hormone biology after careful review.

Thyroid hormone biology is critical for human health and we welcome the annex on thyroid hormone in the guidance document as an archetype for the treatment of other highly complex endocrine pathways. However, there is a lack of validated test methods that accurately capture the complexity of thyroid interference. To improve reliability, this section should be carefully reviewed, taking into account our detailed comments, and regularly updated in the future with the participation of expert endocrine scientists.

Sincerely,

Angel Nadal, PhD
Chair, EDC Advisory Group

ENDOCRINE SOCIETY
Hormone Science to Health

Line-by-line comments:

The Guidance Document must have a broad scope that covers the latest science on endocrine pathways and the biology of hormones.

- Page VII: Replace “interact” with “interfere”. If a chemical “interacts” with the endocrine system, it must produce an “effect”. This is logical inasmuch as a regulated chemical cannot “interact” with the endocrine system of some subgroups within the human population in a manner that isn’t “interference”.
- Page VIII: Replace, or remove the Klimisch-based definitions for“Relevance” and “Reliability”. The Klimisch et al. definitions are not reliable and are not consistent with elements of the guidance document. Independent definitions of these terms should be used, if at all.
- Page VIII: “Temporal Concordance”. Add a sentence explaining that most of the “KEs” in the AOPs that have been reviewed by OECD are not likely to have data associated with them – especially within the context of guideline studies.
- 190: Delete “which is generally applicable to all chemical substances” As written, the statement implies that all chemical substances ARE endocrine disruptors.
- 205-228: We recommend:
    - Changing the sentence beginning with line 205 to say “…, this guidance focuses mainly on the effects caused by estrogen, androgen, thyroid…”
    - Replace lines 218-223 with a statement stating “Additional test guidelines should also be used to substantiate identification of EDs based on updated versions of OECD GD 150.”
    - Delete the sentence beginning with “Consequently, these assays…”
    - Add a sentence stating: “If available, information on endocrine-disrupting activity in non-target invertebrates should be considered for the ED assessment.”

  This guidance should cover all endocrine pathways and should retain flexibility to include non-EATS pathways. This will be enabled by eliminating reference to the Klimisch method of evaluating study quality and using available systematic review methods. Although OECD GD 150 does not provide explicit guidance on interpretation of non-guideline endpoints, it does not logically follow that the non-guideline endpoints are not meaningful. In addition to the examples based on EATS, the guidance should also provide information on other aspects of the endocrine system that should be considered and how literature data and new in vitro information can be combined through systematic review to understand how

2055 L Street NW
Suite 600
Washington, DC
20036

T. 202.971.3636
F. 202.736.9705

endocrine.org

the chemical acts. In cases where little data exist and in the absence of MoA data, an adverse effect within an endocrine domain (e.g., reproductive and pubertal development, metabolism) should be sufficient to label a chemical as an endocrine disruptor and encourage the development of more data.

- 368-76: We recommend that this text be replaced, stating instead: "A mechanism of action is discovered and an adverse effect in an endocrine-mediated area should be considered relevant, pending additional study." It is not clear why the document distinguishes EATS-mediated parameters and EATS-sensitive parameters. Furthermore, the current in vivo guideline assays do not cover all of the possible EATS modalities, but they were not designed to do so and sensitive to but not diagnostic of EATS endpoints may indicate that other endocrine modalities may be impacted. Use of the scientific literature may provide enough evidence to determine if an adverse effect is likely endocrine mediated.
- 406: We recommend deleting the sentence beginning with "In this step…" The phrase "systematic literature methodology" implies that academic literature would exclusively be evaluated by a separate systematic review. Rather, this sentence should be replaced with a sentence stating that: "The entire body of knowledge that is obtained for that chemical should be included together in a transparent and well defined systematic review process."
- 251-3: We recommend that the term "WoE" be replaced with "systematic review." A systematic review should be conducted because it is more transparent and removes potential bias from the decision.
- 305-9: We recommend that the final sentence in this paragraph be replaced with "However, they cannot replace laboratory studies for assessing endocrine activity and adversity."
- Section 5.2: We recommend adding more information about other endocrine pathways, specifically metabolic pathways including effects on body weight regulation and insulin action/secretion. Synergistic and combinatorial exposures should also be considered for further research.

Agencies should be able to identify chemicals with endocrine disrupting properties based on realistic amounts of scientific information.

- 37-39: We recommend deleting these lines. Because a systematic review would be expected to transparently treat all data equally, these lines are unnecessary, and imply that GLP studies would be evaluated under a different set of criteria relative to non-GLP studies.

- 78: Risk characterization is outside of the scope of this document; we recommend that this line be deleted.
- 80: We recommend deleting the word “preferably” and providing further explanation of the term “Standardized methodology”.
- 266-7: We strongly recommend that systematic review be prioritized (over AOPs) in the evaluation of submitted data for a chemical. No guidance exists that clarifies the point at which a plausible link has been established for adverse effects and endocrine activity using AOPs. For example, if an MIE is missing, can a conclusion be made? How many KEs are required? Furthermore, few AOPs have been formally reviewed and published for endocrine modes of action. While AOPs may have utility as a tool to organize data, the AOP approach has not been validated for use in regulatory settings.
- 286-9: The use of expert judgement is not clear in this paragraph and should be formally and transparently delineated in the context of a systematic review process.
- 311-7: It is not clear what would be considered sufficient to indicate a plausible link between the adverse effect and endocrine disruption. Are the in vivo guideline assays not sufficient for ED identification if adverse endocrine endpoints are observed in the most recent two-gen or the extended one-gen assays? The demonstration of the plausible link should be limited to an assumption coherent with known biological principles, instead of a detailed study of ED action and mechanisms.
- 320-2: Dose-response as a criterion for hazard identification here in this paragraph is unsupported, and the document should indicate that dose-response is not absolutely required for hazard identification. See the following references arguing that dose-response is not necessarily required for hazard assessment ((Berlin consensus statement (Solecki et al. 2017)).
- 401-3: Table 4 should be accompanied by clarifying text that explains how transparency and documentation is established for decision points.
- 413-7: Please clarify what happens when no adversity based on EATS-mediated parameters is observed. Because the guidance only refers to OECD assays for EATS, how are peer-reviewed academic literature or non-EATS studies used to substantiate AOPs used? How will the evaluation proceed if no OECD study data are submitted?
- 460-8: This section is very problematic as it implies that the final determination rests fully on the outcomes of level 3 tests. If a positive ER and ERTA assay are detected, but the level 3 assays are negative, then no endocrine activity will be considered to be observed; however this may reflect a false-negative due to a potential lack of sensitive endpoints in guideline studies. If a level 2 assay or other in vitro assay is positive for endocrine, but a level 3 assay is negative, then this compound is a potential EDC, it should be clarified that

existing information is insufficient to make a determination, and additional investigation should be encouraged.

- 515-6: Remove “likely” in the sentence. If EATS-mediated parameters show adversity, it should be sufficient.
- 522 – 538, including Figure 4: Please clarify what would be considered ‘sufficient information to support the postulated MoA’? Line 531-532 states that “In the case of endocrine disruption, this sequence at least includes one endocrine mediated KE.” If an adverse effect is observed and endocrine-mediated action observed, we recommend that this be sufficient to support the definition of an EDC. “Endocrine-mediated action” cannot be restricted to perturbing a single class or system of hormones interacting with a receptor, since a single chemical or class of chemicals can interact with different endocrine pathways, and endocrine systems are often linked. Therefore, “endocrine-mediated” should specifically indicate that the adverse outcome is plausibly caused by a substance interfering with hormone action. By “hormone action”, we mean “hormone receptor activation”, recognizing that many hormones have multiple receptor isoforms including nuclear and/or membrane or other receptors that “transduce” hormone signals into cellular actions that affect development and/or physiology. It should also reflect the World Health Organization’s International Program on Chemical Safety (WHO-IPCS) definition, which encompasses all endocrine systems and effects including a) receptor-mediated effects; b) interference with endogenous ligand delivery to the receptor; and c) epigenetic effects.This should be stated in Figure 1 and 4 as the minimal information needed to fulfill sufficiency. For situations where adversity is demonstrated, and endocrine activity is documented, but a MoA isn’t obvious, chemicals should be classified as a “potential ED”.
- 538-601: What is the definition of a non-endocrine MoA if it signals an endocrine effect? Text should be added that states “Information that is not traditionally included in the OECD’s EAT endocrine mechanistic requirements may still be considered endocrine and should not be labeled ‘non-endocrine’. In these cases, the scientific literature and independent endocrine experts should be consulted to determine whether the mechanisms linked to the adverse effect should be considered to be an ‘endocrine MoA’”. To ensure transparent decision making processes, text should be added to clarify what would represent sufficient information to establish the plausibility of a MoA. For example, an adverse effect that is ‘EATS-mediated’ with little MoA evidence should easily determined as an EDC based on scientific evidence in the scientific literature. In cases where an EATS-mediated effect is observed, and data are only available for one KE, more data should be requested from the applicant or the chemical should be considered a potential EDC. In all cases, including for chemicals with activity outside of the OECD-EATS pathways, systematic review should be utilized to interpret the information available. We

approve of the text establishing that certain molecular events have the potential to affect endocrine pathways that lead to the observed adverse events. We recommend that this text be developed further, adding a sentence stating that "MoAs can be brief and qualitative in the absence of a full mechanistic data and/or when certain early molecular events are shown (e.g., ERTA, MIE or KE1)."

- 613-22: The AOP process should not be utilized here as a way to perform WoE determinations on the plausibility decision. We strongly recommend the use of systematic review as a more transparent and unbiased way to evaluate all existing data.
- 623-33: This text is very helpful and should be emphasized in the decision framework provided. It should be clear that an adverse effect observed in one of the in vivo assays is already associated by design with MoA, and there will be no need to construct an AOP or other plausibility information to justify the identification of an EDC.
- 636-40: We recommend deleting lines 636-40, this text is unnecessary given the statements in lines 623-33. As mentioned throughout our comments, systematic review should be used to comprehensively evaluate data.
- 641-7: We recommend eliminating this text. The purpose of this guidance is to establish biological plausibility for the criteria. The process for establishing and creating new AOPs should not be a requirement for these criteria. The requirement for develop AOPs in this guidance is excessively burdensome and has not been fully developed as a regulatory tool.
- Section 3.5.2: This section includes contradictory guidance; it defines the steps needed to develop an AOP for the purposes of establishing plausible links to the adverse outcome for a MoA, but also states that this approach may not be possible or feasible. This guidance sets up a data interpretation process that will lack transparency and serves the goal of developing AOPs. We assert that systematic review would be a better tool for evaluation of the available data.
- 676-81: While information for most KEs will come from the scientific literature, we are concerned that AOP pathways developed for the criteria will not have been subjected to peer-review and validated for regulatory purposes. For example, OECD requires that AOPs go through external, independent scientific review before the AOP is considered an official OECD manuscript; however, many details of this process are still being developed. The interpretation of the literature data and how it can fit into an AOP will require expert review. There is a risk that information from the literature will be mis-used unintentionally within an AOP used to determine if the criteria are supported. Regardless of the approach used, independent experts should be consulted to determine if information is correctly utilized and understood.
- 690-3: We recommend that this section be removed. Incorporation of dose-response information into the AOPs in a way that can be utilized for regulatory decisions is not

established. Especially for comparisons between in vitro assays and in vivo assays, our current understanding does not allow dose response to inform temporal concordance in most cases. Comparisons may be irrelevant or difficult to interpret because assays vary in species, dosing, and other ways.

- 699-702: We fully support the statement "Dose-response and temporal concordance can be used either within one specific study, where parameters associated with different KEs are measure, or across studies. Most often, the complete data set needed to fully address temporal concordance is not available and this should be considered in the WoE." We recommend beginning the section with this statement, and also suggest adding a sentence stating "if cases exist where KEs are comparable, then dose response and concordance can be evaluated."
- 703-15: we recommend eliminating this section. This kind of analysis using an AOP creates an unnecessarily high burden of proof. Furthermore, most of the data to establish AOPs will be from different literature studies, making dose response comparisons difficult if not impossible. Questions should first be asked to ensure that the KEs are comparable, prior to addressing dose-response. Example questions could be:
    - Do data come from different species?
    - Do experimental differences affect dose-response or temporal onset?
    - Are the endpoints evaluated comparable?
- 719-20: We recommend adding text to this statement, such that it reads "the biological knowledge of the sequence of events, if supported by the scientific literature or presumed linkages based on scientific evidence may be considered sufficient."
- 812-20: We are concerned that the WoE evaluation does not include an assessment of the science that is used to develop the AOP. A systematic review would be more appropriate here to evaluate the entire process and the information that goes into it.
- 833-5: A biologically plausible KE should not necessarily be a requirement.
- 921-2: We recommend adding a sentence that says, "reasonable uncertainties as understood by the scientific community can be accepted."

The Guidance Document should carefully review the Annex on thyroid hormone biology as a model for other complicated endocrine pathways.

- 2642: The evidence that TSH suppresses TRH "production" and serves as a negative feedback pathway within the HTP axis is weak. This "short-loop" feedback has been hypothesized but there is explicit evidence that it is not operative (Zoeller et al. 1988). We refer the authors to a recent publication by Joseph-Bravo et al. (Rev Endocr Metab Disord (2016) 17: 545-558) for more information.

- 2643 – It is not absolutely true that the HPT axis has been conserved across all vertebrates. For example, CRH (not TRH) regulates the PT axis in frogs. We recommend changing this sentence to read "The hypothalamic-pituitary-thyroid axis (HPT axis) is highly conserved across evolution in vertebrates."
- 2650-7: This paragraph includes a vague allusion to quantitative differences among species that can/should affect the relevance of animal data to humans. However, the data may still be relevant; for example, the degree to which serum T_4 must decline in humans before measurable effects are observed on (depending on timing) is probably not different from that of rodents if the data are evaluated side-by-side. We recommend that the exact quantitative differences that apply to drawing comparisons between humans and animals be specified where possible, given the critical importance of this issue.
- 2709 – 20: This paragraph should be maintained in current form, with attention paid to proper references.
- 2800 – 3: We recommend deleting the sentence starting with "Although the inclusion of a…". The addition of concurrent positive controls is required because too many factors can alter both the effect size and the sensitivity of the animal model to the positive control. Many factors can contribute to differences among apparently similar studies that cannot be corrected by historical positive controls. These include, but are not limited to, animal strain, stress, vivarium source, presence of exogenous chemicals in feed, water, bedding or caging.
- 2837 – 8: The statement beginning with "the applicability of RIA for the pups is questionable…" should either be deleted, or should be clarified to stipulate what "applicability" means, with scientific justification for the statement. It is not clear what the intended meaning of this statement is. There are many good peer-reviewed papers where measurements of thyroid hormones in rodent pups are used. The statement is not upheld by the scientific literature, and we object to the use of "personal communication" as a reference.
- 2855 – 8, Table A.1: Please include citations for this information. The size of the CV is not consistent with our understanding of the published scientific literature.
- 2870: Please include a citation for the statement ending in line 2872.
- 2908 – 12: The T_4 and T_3 assays are commercial assays with known %CV for inter- and intra-assay variance. This paragraph should be entirely rewritten, and:
    - State that in the SRO document that the %CV is within limits established for the particular commercial kit. Most of these are approximately <4% for intra-assay variation and <10% for inter-assay variance.
    - Describe how these values are established by the SRO.
    - Stipulate that only interpolated values are valid. Extrapolation from the lowest or highest standard is not valid.

- 2914: We recommend including another line stating: "A serum dilution curve should also be run to show that the assay is valid for the serum samples under investigation."
- 2928 – 32: Historical controls cannot be used to determine if serum hormone levels in treated groups are significant. We recommend replacing the sentence beginning with "However, each laboratory conducting…" with a sentence stating, "Historical controls should be consulted only as a qualitative measure of the assay reliability."

Other points to consider

- 220, 365: Change "assessment" to "identification".
- 242-5: Please clarify the rationale for separating the identification of EDCs from the determination of EDC properties. Why do these concepts need to be separated?
- 270-2, p.41 – 869-71: change "plausible" to "plausibly"
- 362-4: Readers are referred to distant tables (located several dozen pages later) to find examples of the parameters that can be attributed to the different groups outlined in lines 330-354. For quick reference, it would be very useful if several examples were listed immediately following the definition given for each of these four groups (i.e., in lines 334, 340, 347, 354).
- 409: Please define the abbreviations "RAR/DAR".
- 420-2: More guidance should be given to the assessor on what evidence is sufficient to define a compound as an EDC. To improve transparency, this text should be reworded to indicate that the assessor will drive the process and follow the guidance established in this guidance document. We recommend adding a sentence indicating that the transparency of the assessment must be such that independent scientists are in the position to evaluate the decision.
- Page 9, figure 1: The decision-making process for each of the diamonds at the bottom of the level "initial analysis of the evidence" need to be well described in the text. A positive decision that places a chemical into the "conclusion currently not possible" or "criteria not met" should be well documented with clear guidance.
- 22: We recommend that the document provide examples of sufficient evidence for relevant adverse effects.
- 241: There appears to be a typographical error in this line where there should be a reference.
- 299: In this line, we recommend changing "EAS" to "EATS".
- 366: Table 2: In the column "Observed effects", the first cell from the top should be changed to "Predicted to Inhibit iodine transport". In the column "Conclusion," several cells starting with second from the top say "Sufficient; …" Our impression is that this refers

to “Sufficient evidence”, but more detail is necessary to ensure proper interpretation. Units used in the column for “observed effects” should be labeled.

- 498: Both MoA and AOP concepts should be introduced and discussed here to ensure that readers understand what these concepts are and how they will be used.
- 602: “i.e.” should be “e.g.” here.
- 1135-6: We recommend rewriting this sentence to read “…interference with hormone receptors, their downstream signaling, their transporters, non-classical receptors, or interaction with key enzymes…”.