

President of the EU Commission Mr Jean-Claude Juncker
EU Commissioner Mr Vytenis Andriukaitis
European Commission
Directorate General for Health and Food Safety
B – 1049
Brussels, Belgium

16 November 2016

Open letter in response to the redrafted criteria for identification and regulation of endocrine disrupting chemicals, under the PPP and Biocides Regulations

Dear President Juncker and Commissioner Andriukaitis,

We are writing to you as scientists conducting research into endocrine disrupting chemicals (EDCs) and systematic review methods for chemical risk assessment, in order to voice our concerns about the redrafted criteria for identification and regulation of EDCs under the PPP and Biocides Regulations, and to contribute our perspective on the challenge of aggregating scientific evidence in the process of identifying EDCs.

We welcome the additional detail in the redraft of the EDC criteria but do not believe it goes far enough to address the issues a number of us described in a letter dated 6 July this year, while raising others. In particular, we are concerned about the following:

1. An unclear fit between the requirement that a substance “may cause adverse effects” and the implication throughout the document that only chemical substances which are known to cause adverse outcomes via alteration of the function of the endocrine system will be classified as EDCs.
2. That the new draft has replaced what was unequivocally a too-high burden of proof with an ambiguous burden of proof.
3. The addition of new detail which explains the regulatory implementation of weight-of-evidence assessment and systematic review as it pertains to EDCs is welcome, but it does not capture best practice in evidence synthesis and integration, and is not an operable approach to making full and fair use of the existing evidence to identify endocrine disruptors.
4. The retention of a two-tier hierarchy of evidence, of “internationally agreed study protocols” as against “other relevant scientific data”, further prevents implementation of a fair and operable evidence integration methodology.
5. That the move from “negligible exposure” to “negligible risk” is being justified as a scientific matter when in fact it seems a political one, the implementation of which requiring a different regulatory process than the one currently being followed.

Below we provide detailed comments and suggested wording to resolve tensions and challenges in the redrafted proposal. Overall, in response to the redraft of the criteria, we recommend the following:

- a. Unambiguous allowance for regulatory identification of a chemical substance as an EDC when the level of proof is lower than “known”.

- b. A requirement that best practices in finding, appraising, synthesising and integrating evidence are used when assessing whether or not a chemical should be classified as an EDC, with systematic and/or weight-of-evidence approaches to be applied where feasible and appropriate.
- c. In delivering a full and fair assessment of the relevant data, ensuring that all evidence is assessed on merit without prior privileging of certain study types.
- d. The introduction of a hierarchy of categories for EDCs, with clear, unambiguous criteria distinguishing “known” from e.g. “probable”, “possible” or “not classifiable”, to describe the results of the assessment of the evidence.
- e. The definition of clear and unambiguous standards for strength of evidence within each of the three individual components of the EDC definition, and criteria for integrating these individual judgements into a final conclusion about the extent to which the evidence indicates that a chemical substance is an EDC.

We appreciate this is a lengthy letter, but since our initial correspondence both the proposals for the EDC criteria and our own thinking have advanced such that a more detailed response can be made.

If it would be of assistance, and there is time before any further redrafting, we would like to request another meeting so we can articulate our concerns in more detail, based on our experience of developing scientific guidance for the identification and classification of EDCs, and to discuss how they might be resolved through further redrafting of the regulatory proposal.

We look forward to hearing your response.

Yours sincerely,

Mr Paul Whaley*. Lancaster Environment Centre, Lancaster University, Lancaster, UK.

Dr Marlene Ågerstrand. Department of Environmental Science and Analytical Chemistry, Stockholm University, Stockholm, Sweden.

Professor Ake Bergman. Department of Environmental Science and Analytical Chemistry, Stockholm University, Stockholm, Sweden.

Professor Lisa Bero. Charles Perkins Centre, University of Sydney, Sydney, Australia.

Dr Anna Beronius. Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden.

Professor Carl-Gustaf Bornehag. Department of Health Sciences, Karlstad University, Karlstad, Sweden. Icahn School of Medicine at Mount Sinai, New York City, USA.

Professor Ian Cotgreave. Swedish Toxicology Sciences Research Center (Swetox), Karolinska Institutet, Södertälje, Sweden.

Mr David Gee. Institute of Environment, Health and Societies, Brunel University London, Uxbridge, United Kingdom.

Dr Crispin Halsall. Lancaster Environment Centre, Lancaster University, Lancaster, UK.

Professor Malcolm Macleod. Centre for Clinical Brain Sciences, University of Edinburgh, Scotland, UK.

Dr Olwenn Martin. Institute of Environment, Health and Societies, Brunel University London, Uxbridge, United Kingdom.

Professor Christina Ruden. Department of Environmental Science and Analytical Chemistry, Stockholm University, Stockholm, Sweden.

Professor Martin Scheringer. RECETOX, Masaryk University Brno, Czech Republic. Institute for Chemical and Bioengineering, ETH Zürich, Zürich, Switzerland.

Professor Laura Vandenberg. Department of Environmental Health Sciences, University of Massachusetts Amherst School of Public Health & Health Sciences, Amherst, MA, USA

Professor Tracey Woodruff. School of Medicine, Program on Reproductive Health and the Environment, University of California, San Francisco, Oakland, CA, USA. (added 7 July)

**Address for correspondence: Mr Paul Whaley, 45 Trafalgar Road, Lancaster, LA1 4DB, UK.*

Specific comments on the redrafted text

“May cause adverse effects”

Several times in the annex, there is text which appears to describe EDCs as chemical substances which “may cause adverse effects”. While this could be a positive step forward in the development of the criteria, potentially allowing regulatory control of compounds which raise concerns but are not definitively proven to be EDCs, it appears the phrase is only questionably consistent with wording throughout the rest of the annex which could in fact imply a high burden of proof on demonstrating that a chemical substance is an EDC. While there is an appearance of lowering of burden of proof, it is not clear if in fact the criteria will be interpreted in this way.

The ED Criteria

The redrafted criteria do seem to be moving closer to what is needed; however, there is problematic wording which is ambiguous both in overall burden of proof for classification as an EDC (which may in at least one instance be excessively high), and seems to require different strength of evidence for each individual criterion, making it difficult to see how the criteria can be interpreted consistently.

Criterion #1 “adverse effect”: The phrase “it [the substance] shows an adverse effect” is semantically peculiar, because substances do not themselves show adverse effects; rather, it is relevant scientific evidence of adverse effects, interpreted by experts, which potentially shows an adverse effect, depending on how strong that evidence is. Presumably, this is why previous wording emphasised “it is known to cause”. While that was too high a level of proof, at least the phrase captured the outcome of the scientific assessment; in contrast, “shows” is ambiguous, with conditions for fulfilment left undefined. While this may be intended to be softer than the “is known” phrasing it replaces, it is not clear with what level of proof “shows” is intended to correlate (e.g. is it “presumed”?), nor how “shows” can be the outcome of a weight-of-evidence assessment.

Criterion #2 “endocrine mode of action”: The level of proof required for satisfying this criterion is not defined, and it is not clearly stated whether “it has an endocrine mode of action” is something which needs to e.g. be known, or presumed, or suspected, in order for a chemical substance to potentially be classified as an EDC. Furthermore, the term “mode-of-action” is not defined and often confused with “mechanism-of-action” in much of the scientific literature. It could also be interpreted as presupposing that evidence for a mode-of-action needs to be observed at the cellular level.

However, according to EFSA's own analysis, scientific evidence informing an endocrine mode-of-action, corresponding to level 2/3 of the OECD conceptual framework, is not routinely required and generally lacking (EFSA, 2015). It is therefore unclear as to what evidence ought to be used for the assessment of this criterion and what level of proof would be required to fulfil this criterion.

Criterion #3 “consequence of”: The use of the phrase “consequence of” is ambiguous in terms of implied level of proof, being interpretable in at least two ways. If “consequence of” is to be interpreted in a looser sense of “mediated by” alteration of the function of the endocrine system, then it seems the phrase is ambiguous in terms of required level of proof; if the intent is that the adverse effect is known to be caused by the endocrine mode of action, then the implied burden of proof is very high and not consistent with the intent to identify as EDCs substances which “may” cause adverse effects.

Recommendation: Rather than stating or implying anything about required level of proof for each individual criterion, we suggest applying a global condition which describes the standard of proof to be applied across all three criteria. This could be articulated, in simplified form, as follows:

“An active substance, safener or synergist shall be considered as having endocrine disrupting properties that may cause an adverse effect in humans if ... there is sufficient evidence of:

- *an adverse effect in an intact organism or its progeny [etc.]*
- *alteration in the functioning of the endocrine system, and*
- *the adverse effect being mediated by the alteration in function of the endocrine system”*

What counts as “sufficient evidence” can then be unambiguously defined, and the extent to which it exists can be determined by the weight-of-evidence process. While previous drafts suggest a preference for defining this as “known”, we believe that a weight-of-evidence process which yields at least a judgment of “known” or “presumed” endocrine disruptor (or suitably unambiguous, equivalent language) would be most appropriate for regulatory classification of a compound as an EDC, and would be consistent with the intent to classify as EDCs chemical substances which “may” cause adverse effects.

Note that the issue of whether the criteria are fulfilled or not is unlikely to be clear and binary (i.e. simply fulfilled or not) because weight-of-evidence and systematic review methods normally produce a statement of *the extent* to which expert reviewers believe the criteria can be considered fulfilled given the available evidence. Only on rare occasions will there be a clear-cut conclusion as to whether or not they are fulfilled. Requiring such a clear-cut conclusion before classifying a compound as an EDC would, we believe, result in a large number of compounds in need of risk management measures evading regulatory control.

Indeed, logically it is unclear how the regulation is supposed to capture compounds which “may” cause adverse effects as endocrine disruptors, if the regulation will only classify compounds as EDCs when adverse effects consequent to endocrine activity are “known” to be taking place.

We therefore believe that having several categories of classification of endocrine disruptor is a sensible approach to transparent, equivalent codification of the strength of the evidence from weight-of-evidence assessment and consequent level of regulatory priority accorded to a compound.

Handling of “all available relevant scientific data”

While there is a welcome increase in detail on how evidence is to be assessed, there remain a number of inconsistencies and confusions in the articulation of the use of weight-of-evidence and systematic review methods which render the processes as described inoperable.

Handling of “all available relevant scientific data”. The way this requirement is articulated potentially yields a two-tier hierarchy of evidence via the separation of (a) “scientific data generated in accordance with internationally agreed study protocols [etc.]”, as against (b) “other relevant scientific data”.

Firstly, this is logically incoherent, because by definition there can be no “other relevant” data in addition to all relevant data. Secondly, wording such as “in particular” appears to accord greater importance to certain types of evidence (e.g. “protocols listed in the Commission Communications”). This presents an ambiguous hierarchy of information which goes against the apparent intent of ensuring all data is fully taken into account in the assessment.

It should be noted that it is not actually necessary to pre-specify a hierarchy of evidence because if, on systematic assessment, the “internationally agreed” protocols provide the strongest evidence then they will carry most weight in the analysis without having to be accorded it in advance.

Recommendation: Wording such as “in particular” should be dropped, in favour of unambiguous articulation of the need to make full use of all the relevant evidence, and assess the evidence on its own merits. To eliminate the implied hierarchy of evidence, combine clauses (a) and (b): “all available scientific data, found and selected using systematic search and inclusion methods, to include evidence generated in accordance with internationally agreed study protocols [etc.] AND scientific data generated using other study methods.”

Weight-of-evidence and systematic review

Weight-of-evidence and systematic review. The relationship between weight-of-evidence and systematic review methods is still unclear, with systematic review reserved for “other relevant scientific data selected applying a systematic review methodology”.

This phrase does not make sense, because systematic review is not a method merely for selecting data. In fact, weight-of-evidence and systematic review methods have much in common, both sharing the objective of finding, appraising and synthesising existing evidence (hence why the concept of using systematic review methods to only *select* evidence is nonsense); the difference is in the specific methods used, in particular the explicit focus of systematic review on techniques which seek to minimise risk of bias in the results of the evidence assessment process.

While more detail has been given in the way in which the weight-of-evidence process is to be conducted, the characterisation is fundamentally problematic. This is not unexpected, given that weight-of-evidence methods have in general been found to be under-defined and inconsistently articulated (Ågerstrand & Beronius 2016), and that as a phrase “weight-of-evidence” has been described by the US National Academy of Sciences as “too vague” and “of little scientific use” (US National Resource Council 2014).

The reference in the criteria to the “quality, reliability, reproducibility and consistency” of evidence does define some aspects of the appraisal of a body of evidence, but the terms are not defined, and they only provide partial coverage of what needs to be taken into account in determining the quality of a body of evidence. For example, the Navigation Guide (Woodruff and Sutton 2014) and National Toxicology Program Office of Health Assessment and Translation (OHAT) approaches to systematic

review (Rooney et al. 2014), have adapted the GRADE approach used in Cochrane systematic reviews of healthcare interventions (Morgan et al. 2016) to systematically take into account the following features of the evidence base when determining confidence in the results of a systematic review:

- Risk of bias across the evidence base
- Consistency of the evidence
- Precision of the evidence
- Risk of publication bias
- Directness of the evidence base
- Plausible confounding
- Magnitude of effect
- Dose-response relationship

This provides more comprehensive coverage of concepts only alluded to in the current draft. Evidence is considered stronger the better it performs in each of these categories, with strong evidence being broadly defined as that which is highly unlikely to be overturned by a new study (because e.g. it would have to be very large and show an effect in the opposite direction to that already being observed).

That said, we acknowledge that there is not yet consensus on how best to apply systematic review methods in the context of identifying EDCs (Whaley et al. 2016), and that this is a matter of ongoing research, such as in the development of the SYRINA methodology which many of the undersigned have been involved in developing (Vandenberg et al. 2016). Pending changes to weight-of-evidence methodologies and the development and implementation of systematic review methods, we believe it would be better to stipulate that best practice must be followed, but avoid going into unnecessary detail which could commit agencies and review committees to using methods which could become superseded in the near future.

Recommendations: Given current moves to systematise the assessment of EDC evidence and extend systematic review methods into chemical hazard and risk assessment, and uncertainty about the future direction of weight-of-evidence methods, it does not appear to us that pinning the criteria to weight-of-evidence methods is appropriate. Instead, we suggest that reference is made to best practices in evidence gathering, appraisal and integration, assuring that all evidence is appraised fully and fairly. The text should leave open the choices as to which specific methods to use, depending on context. Wording could be something like:

“An assessment of the available, relevant scientific evidence is conducted by applying best practices for finding, appraising, synthesising and integrating all the relevant evidence for assessing ED potential, to determine the extent to which criteria 1-3 are fulfilled”.

This allows best practice to be used in assessing the evidence, without any presuppositions needing to be made about appropriateness of systematic review or weight-of-evidence methods.

[Negligible exposure vs. negligible risk](#)

This is less of a purely scientific matter, but we wanted to echo concerns raised elsewhere that the move from “negligible exposure” to “negligible risk” in the proposed criteria is questionable. We note that the regulation argues this is a scientific matter. This is surely not the case: whether society wishes to manage risk of harm from chemical substances via hazard- or risk-based approaches strikes us as a values-based decision, not as one which can be straightforwardly determined by

scientific research. If the Commission believes risk assessment has advanced to the point that there can be sufficient confidence in the ability to quantify risks to health posed by EDCs, such that one argument in favour of a hazard-based approach is put to bed, then it should present this reasoning to Parliament and a democratic consensus on whether or not we should proceed in this manner can be reached. While it may well be the opinion of EFSA that EDCs can be adequately risk-assessed, it is not clear whether this is a consensus view, nor that such a consensus has been demonstrated via due political process.

Recommendation: Retain language describing exclusion of contact with humans, remove references to “negligible risk”, replace with “negligible exposure”.

References

Ågerstrand, M., & Beronius, A. (2016). Weight-of-evidence evaluation and systematic review in EU chemical risk assessment: Foundation is laid but guidance is needed. *Environment International*, 92-93, 590–596. <http://doi.org/10.1016/j.envint.2015.10.008>

Morgan, R. L., Thayer, K. A., Bero, L., Bruce, N., Falck-Ytter, Y., Ghersi, D., ... Schünemann, H. J. (2016). GRADE: Assessing the quality of evidence in environmental and occupational health. *Environment International*, 92-93, 1–6. <http://doi.org/10.1016/j.envint.2016.01.004>

Rooney, A. A., Boyles, A. L., Wolfe, M. S., Bucher, J. R., & Thayer, K. A. (2014). Systematic Review and Evidence Integration for Literature-Based Environmental Health Science Assessments. *Environmental Health Perspectives*. <http://doi.org/10.1289/ehp.1307972>

US National Resource Council. (2014). Review of EPA’s Integrated Risk Information System (IRIS) Process. The National Academies Press. Retrieved from http://www.nap.edu/openbook.php?record_id=18764

Vandenberg, L. N., Ågerstrand, M., Beronius, A., Beausoleil, C., Bergman, Å., Bero, L. A., ... Welshons, W. (2016). A proposed framework for the systematic review and integrated assessment (SYRINA) of endocrine disrupting chemicals. *Environmental Health*, 15(1), 74. <http://doi.org/10.1186/s12940-016-0156-6>

Whaley, P., Halsall, C., Ågerstrand, M., Aiassa, E., Benford, D., Bilotta, G., ... Taylor, D. (2016). Implementing systematic review techniques in chemical risk assessment : Challenges , opportunities and recommendations. *Environment International*. <http://doi.org/10.1016/j.envint.2015.11.002>

Woodruff, T. J., & Sutton, P. (2014). The Navigation Guide Systematic Review Methodology: A Rigorous and Transparent Method for Translating Environmental Health Science into Better Health Outcomes. *Environmental Health Perspectives*. <http://doi.org/10.1289/ehp.1307175>