## Manufacturing doubt about endocrine disrupter science

# A rebuttal of industry-sponsored critical comments on the UNEP/WHO report "State of the Science of Endocrine Disrupting Chemicals 2012"

Bergman, Åke; Becher, Georg; Blumberg, Bruce; Bjerregaard, Poul; Bornman, Riana; Brandt, Ingvar; Casey, Stephanie C.; Frouin, Heloise; Giudice, Linda C.; Heindel, Jerrold J.; Iguchi, Taisen; Jobling, Susan; Kidd, Karen A.; Kortenkamp, Andreas; Lind, P. Monica; Muir, Derek; Ochieng, Roseline; Ropstad, Erik; Ross, Peter S.; Skakkebæk, Niels Erik; Toppari, Jorma; Vandenberg, Laura N.; Woodruff, Tracey J.; Zoeller, R. Thomas

*Published in:*
Regulatory Toxicology and Pharmacology

*DOI:*
10.1016/j.yrtph.2015.07.026

*Publication date:*
2015

*Document version*
Final published version

Regulatory Toxicology and Pharmacology 73 (2015) 1007–1017

Contents lists available at ScienceDirect

Regulatory Toxicology and Pharmacology

journal homepage: www.elsevier.com/locate/yrtph

Commentary

# Manufacturing doubt about endocrine disrupter science – A rebuttal of industry-sponsored critical comments on the UNEP/WHO report "State of the Science of Endocrine Disrupting Chemicals 2012"

Åke Bergman [a, *], Georg Becher [b], Bruce Blumberg [c], Poul Bjerregaard [d], Riana Bornman [e], Ingvar Brandt [f], Stephanie C. Casey [c], Heloise Frouin [g], Linda C. Giudice [h], Jerrold J. Heindel [i], Taisen Iguchi [j], Susan Jobling [k], Karen A. Kidd [l], Andreas Kortenkamp [k], P. Monica Lind [f], Derek Muir [m], Roseline Ochieng [n], Erik Ropstad [o], Peter S. Ross [g], Niels Erik Skakkebaek [p], Jorma Toppari [q], Laura N. Vandenberg [r], Tracey J. Woodruff [h], R. Thomas Zoeller [r]

[a] Swedish Toxicology Sciences Research Center (Swetox), Södertälje, Sweden
[b] Norwegian Institute of Public Health, Oslo, Norway
[c] University of California, Irvine, USA
[d] University of Southern Denmark, Odense, Denmark
[e] School of Health Systems and Public Health, University of Pretoria, Pretoria, South Africa
[f] Uppsala University, Uppsala, Sweden
[g] Vancouver Aquarium Marine Science Center, Vancouver, Canada
[h] University of California, San Francisco, USA
[i] National Institute of Environmental Health Sciences, Research Triangle Park, USA
[j] National Institute for Basic Biology, Okazaki, Japan
[k] Brunel University London, Uxbridge, UK
[l] University of New Brunswick, Saint John, New Brunswick, Canada
[m] Environment Canada, Burlington, Canada
[n] Aga Khan University Hospital, Nairobi, Kenya
[o] Norwegian University of Life Sciences, Oslo, Norway
[p] Copenhagen University Hospital, University of Copenhagen, Copenhagen, Denmark
[q] University of Turku, Turku, Finland
[r] University of Massachusetts Amherst, Amherst, USA

ARTICLE INFO

*Article history:*
Received 20 July 2015
Received in revised form
28 July 2015
Accepted 29 July 2015
Available online 31 July 2015

*Keywords:*
Endocrine disruption
EDCs
Endocrine disruptors

ABSTRACT

We present a detailed response to the critique of "State of the Science of Endocrine Disrupting Chemicals 2012" (UNEP/WHO, 2013) by financial stakeholders, authored by Lamb et al. (2014). Lamb et al.'s claim that UNEP/WHO (2013) does not provide a balanced perspective on endocrine disruption is based on incomplete and misleading quoting of the report through omission of qualifying statements and inaccurate description of study objectives, results and conclusions. Lamb et al. define extremely narrow standards for synthesizing evidence which are then used to dismiss the UNEP/WHO 2013 report as flawed. We show that Lamb et al. misuse conceptual frameworks for assessing causality, especially the Bradford–Hill criteria, by ignoring the fundamental problems that exist with inferring causality from empirical observations. We conclude that Lamb et al.'s attempt of deconstructing the UNEP/WHO (2013) report is not particularly erudite and that their critique is not intended to be convincing to the scientific community, but to confuse the scientific data. Consequently, it promotes misinterpretation of the UNEP/WHO (2013) report by non-specialists, bureaucrats, politicians and other decision makers not intimately familiar with the topic of endocrine disruption and therefore susceptible to false generalizations of bias and subjectivity.



* Corresponding author.
*E-mail address:* ake.bergman@swetox.se (Å. Bergman).

http://dx.doi.org/10.1016/j.yrtph.2015.07.026

## 1. Introduction

In 2013, the United Nations Environmental Programme (UNEP) and the World Health Organization (WHO) published a report "State of the Science of Endocrine Disrupting Chemicals – 2012" (UNEP/WHO, 2013) and a companion report "State of the Science of Endocrine Disrupting Chemicals 2012 – Summary for Decision-Makers" (UNEP/WHO, 2013b). The reports were prepared in the context of United Nations (UN) activities in recognition of human health and biodiversity, to take stock of new findings, to identify key concerns and future needs. The two UNEP/WHO reports were specifically developed in connection with the Strategic Approach to International Chemicals Management (SAICM), a UN policy framework established by the UN International Conference on Chemicals Management (ICCM). The objective of SAICM is to achieve the sound management of chemicals throughout their life cycle so that, by 2020, chemicals are used and produced in ways that minimize significant adverse effects on human health and the environment. On the basis of the evidence assembled by UNEP/WHO (2013), ICCM decided in September 2012 to include endocrine disrupting chemicals (EDCs) as an emerging issue under SAICM.

As scientists with research expertise in the field of endocrine disruption, we were charged by UNEP and WHO officials in a face-to-face meeting to prepare the UNEP/WHO (2013) report which was intended as an update of the earlier WHO-IPCS report on the subject (IPCS 2002) and to which some of us had also contributed. The new report was reviewed by 23 endocrine disruption experts from 12 countries, with additional extensive critical reviews conducted by UNEP and WHO staff.

We served in a personal capacity, and not as representatives of any organization, government or industry, to prepare the two reports (UNEP/WHO, 2013a; b). We signed Declaration of Interest statements, and no conflicts of interest were identified. The development and publication of the report was supported by funds provided to UNEP from the Norwegian government, the Swedish Environment Ministry, the Swedish Research Council (FORMAS) and the Swedish Environmental Protection Agency, with additional support to WHO from the United States National Institute of Environmental Health Sciences (NIEHS).

The UNEP/WHO report has been criticized by a group of authors (Lamb et al., 2014) who received funding support from several industrial organizations with financial stakes in chemical production, including American Chemistry Council (ACC), CropLife America (CLA), CropLife Canada (CLS), CropLife International (CLI), European Chemical Industry Council (Cefic), and European Crop Protection Association (ECPA).

In their critique, Lamb et al. maintain that UNEP/WHO (2013) does not provide a balanced perspective on endocrine disruption, because scientific evidence was cited in a biased way, with a preference toward studies reporting exposure-disease associations, and without evaluations of the quality of underlying studies. They conclude that the report is neither a state-of-the-science review, nor an update of the earlier WHO/IPCS (2002) report. Lamb et al. (2014) support these conclusions by expressing several specific concerns relating directly to the content and narrative of UNEP/WHO (2013): 1) failure to apply a systematic framework for identifying, reviewing and evaluating data, 2) adoption of an unduly informal approach to assessing causation from EDCs, 3) reliance on disease trends to suggest associations with EDCs, and 4) disregard for the role of exposure, dose and potency in endocrine disruption. In the Lamb critique, the earlier WHO/IPCS (2002) report, to which two of the authors (Foster, Van Der Kraak) contributed, is held up as a positive example.

The overall claim of Lamb et al. (2014) that, "Overall, the 2012 report does not provide a balanced perspective, nor does it accurately reflect the state of the science on endocrine disruption." prompted us to analyze their critique and to respond in detail. We approached this analysis openly. However, we discovered that the conclusion of Lamb et al. (2014) was based on many distortions, inaccuracies, false generalizations, non-scientific argumentation, and erroneous claims. These are illustrated below.

## 2. Results

### 2.1. Does the odd missed reference constitute bias?

Throughout their paper, Lamb et al. claim that we have selectively cited literature without discussion of contradictory studies, and have failed to consider alternative causes of effects, thus giving the reader the impression that the weight of evidence for endocrine disruption is stronger than is justified by the scientific data. Importantly, to substantiate these general allegations, they resort to a maneuver that can only be characterized as misdirection of the reader. Specifically, they selected topic areas from the UNEP/WHO report where the evidence for associations between chemical exposures and endocrine disruption is not very strong (and where this was clearly stated in our report, for example adrenocortical hyperplasia in seals, prostate cancer, poor semen quality), sought out a few studies that were not cited, and make the sweeping generalization that these omissions "call[s] into question the integrity of decisions at all levels of the 2012 report".

This kind of critique might make sense in the context of strongly contested issues, but is misleading when directed at topics where no strong claims are made regarding links between exposure and effect, and where failure to include one or two additional studies would not change the conclusion.

Accusations of bias and imbalance would have been justified, if we had consistently ignored evidence pointing in a specific direction. However, a fair assessment of this question is only possible by engaging scientifically with the substance, content and conclusions of our report, something Lamb et al. avoid doing. Bias cannot be inferred from the odd missed reference.

Instead of engaging with the scientific content of UNEP/WHO (2013), Lamb et al. shift attention to methodological issues concerned with preconceived ideas of what might constitute a state of the science review, and approaches to assessing causation and weight of evidence. This tactic creates the false impression of bias, imbalance and subjectivity. As we will discuss below (see Section 2.8), the tactics used by Lamb et al. have striking similarities with those employed by the tobacco industry to undermine attempts of introducing standardized packaging for cigarettes. These approaches were elucidated recently by Ulucanlar et al. (2014).

### 2.2. UNEP/WHO (2012) – not a state of the science assessment, nor an update of the WHO/IPCS (2002) report?

During the 10 years that passed since publication of the WHO 2002 report, endocrine disruption research has deepened and expanded our understanding of the endocrine system and the role of chemicals in disrupting aspects of its functioning. A brief glance at the range of new topics covered in UNEP/WHO (2013), and not dealt with in the IPCS 2002 report, highlights the progress made. While the 2002 report considered mainly male reproductive health, the 2013 report documented the progress made in female reproductive health. In 2002, there was relatively little knowledge of prostate, thyroid and ovarian cancer in endocrine disruption, but since then, our insight has grown, as documented in the 2013 report. The last ten years have seen an improvement in our knowledge of metabolic disorders and endocrine disruption, and

this was taken up as a new topic in the 2013 report. While in 2002 the focus of wildlife effects was on fish and a small number of mammals, this has expanded to invertebrates, amphibians, reptiles and birds, all documented in the 2013 report. There have also been significant developments in cross-cutting issues, such as mixture effects, which were not mentioned at all in the 2002 report. The sheer explosion of knowledge can also be gleaned from comparisons of the length of the two reports and the number of papers cited: more than 2200 references and 280 pages in the 2013 report, versus 1400 references and 170 pages in WHO/IPCS (2002). The main conclusions and advances in knowledge since 2002 are presented both in the "Executive summary" of the main report (UNEP/WHO, 2013) and in chapter 13 of the Summary for Decision-Makers (UNEP/WHO, 2013a), confirming the extensive progress in EDC related research over the last 10–15 years.

Not surprisingly, we were in a position to reach more refined and definitive conclusions than was possible in 2002. Yet, Lamb et al. do not highlight the intense research activity that was reviewed in the 2013 report, but claim that the 2013 report is neither a state-of-the-science assessment, nor an update of the earlier 2002 report, "but rather a reworking of that earlier report" and support this claim with the assertion that the 2013 report "reviews much of the same information cited in the 2002 report by frequently citing literature from the year 2001 or earlier".

While it is correct that we have cited literature from 2001 and earlier, this was necessary to have complete coverage of the evidence and to place scientific developments into the appropriate context. But to infer from this that the 2013 report is a mere "reworking", with the implication that we covered little new evidence, is a gross distortion of the facts. More than 75% of the citations (ca. 1650 citations) are publications that appeared from 2002 to 2013. Lamb et al. go even further when they write "The fact that the 2012 report reaches more definitive conclusions **based on the same data** emphasizes a reliance on subjective decision making and less stringent criteria for evaluating potential causal relationships compared to the earlier 2002 report." (emphasis added). This creates the absurd impression that there has been little scientific progress since 2002, and that, because we reached more definitive conclusions than the 2002 report "based on the same data", we must be biased, subjective and imbalanced. These claims are without any basis.

A recurring pattern of argumentation in Lamb's critique is to construct arbitrary standards, highlight areas where the UNEP/WHO (2013) report purportedly falls short of these arbitrary standards, and then move to the sweeping generalization that the report must be fundamentally flawed. This tactic is only viable by avoiding any engagement with the scientific substance of the report. It is by this approach that Lamb et al. (2014) come to the surprising conclusion that UNEP/WHO (2013) is not a state-of-the-science report. Specifically, this conclusion is based on several elements:

Lamb et al. note that "The report never defines what might be meant by "state of the science" nor discusses what such an assessment should cover and characterize." They move on to fill this "gap" by defining a state-of-the-science review as one that "should have a defined scope with a systematic approach to the collection and review of data, and a clear methodology for the integration and assessment of these data."

Lamb et al. then criticize that a systematic approach to collecting and reviewing data was not used in UNEP/WHO (2013). However, this exposes a contradiction, which Lamb et al. do nothing to resolve. On the one hand, they point out that the literature "is extensive and beyond the scope of either the 2002 or 2012 report" and concede that "neither report could be expected to undertake complete reviews for even a small subset of chemicals". On the other hand, they assert that "a systematic methodology would have ensured that a representative spectrum of the available literature was captured in review", but fail to define how identification of "representative" literature should be done in a systematic, objective fashion. They ignore that any choice made will always be open to the criticism of lacking a systematic approach to literature selection. Indeed, the same criticism can be leveled against the WHO/IPCS (2002) report which used exactly the same approach as UNEP/WHO (2013), but which Lamb et al. hold up as the example to emulate. They fail to explain why they exempt the WHO/IPCS (2002) document from that criticism.

Indeed, the flexible standards of a state-of-the-science review are illustrated in a document written in 2007 by two of the authors critiquing the UNEP/WHO (2013) report, Hentz and Lamb. They published a report for the Weinberg Group, entitled "2007 Update: State of the Science and Policy for Endocrine Disruption", dated May 29, 2007, in which they summarize the scientific evidence and policy actions relating to endocrine disruption on 14 pages, citing only 21 references and without a systematic approach to collecting and analyzing the literature. (This document was downloaded in 2011, but is no longer available on the www. On request, we are happy to provide the document to anyone interested.)

Regarding a "clear method for the integration and assessment of data", Lamb et al. claim that "the 2002 report attempted to integrate information on exposure, toxicological testing (including dose-dependence of effects), the ability of putative disruptors to interfere with endocrine-mediated control, and patterns of appearance of possibly endocrine-related effects in populations. In contrast, the 2012 report discusses each of these elements independently and specifically declines to consider how these aspects can be brought together to assess whether there are real and current endocrine disruption problems or how well an integrated view of the scientific evidence can answer that question."

This claim is false. The 2002 report did not integrate exposure information and toxicological data. Such integration is a matter for toxicological risk assessment, which was considered out of scope by WHO and UNEP. Examination of both documents reveals that this was applied equally to the UNEP/WHO (2013) and the WHO/IPCS (2002) reports. Accordingly, the WHO/IPCS (2002) report discussed exposure data in a separate chapter, in the same way as UNEP/WHO (2013).

### 2.3. *Weight-of-evidence approaches and frameworks for identifying, reviewing and evaluating data*

Lamb et al. criticize the UNEP/WHO (2013) report for not adopting a weight-of-evidence approach or a systematic framework for identifying, reviewing and evaluating data. At the heart of their critique is the assertion that "the lack of a systematic approach to assess causation for specific chemicals and associated health outcomes resulted in conclusions that were predisposed to the identification of potential EDCs". This statement is based on several naïve assumptions, the first of which is that a systematic approach to assessing causation exists and is uncontroversial for endocrine disrupting chemicals. Secondly, the statement assumes that the application of a systematic approach completely protects against bias by obviating the need for scientific judgments, when in fact scientific judgments are part of the process of systematic reviews. Finally, it is assumed that such "standardized approaches" will always yield the same "objective" assessment outcome. In this and the following section we show that these ideas are illusory and are based on gross simplifications of the debates about causality that have taken place in the epidemiological and science philosophy literature.

First, Lamb et al. ignore that a universally accepted weight-of-

evidence method for endocrine disrupters does not yet exist, and that it will be a considerable challenge to develop such an approach. The methods that currently exist, e.g., for carcinogens under the auspices of WHO IPCS (Sonich-Mullin et al., 2001; Boobis et al., 2006) cannot readily be applied to endocrine disrupters, mainly because an approach useful for EDCs will have to deal with the issues of adversity and mode-of-action at the same time, which is currently without precedent. A more detailed discussion of this topic can be found in Kortenkamp et al. (2011).

It is important to stress that weight-of-evidence approaches were also not used consistently throughout the WHO/IPCS (2002) report, counter to the impression Lamb et al. create.

In fact, the authors of the 2002 report took a much more nuanced stance than insinuated by Lamb et al. The 2002 report acknowledged the difficulty of the task of creating objective and unbiased assessments of whether endocrine disrupting chemicals may have adverse effects. It was recognized that the challenges of this task arise from the vast number of studies conducted, the improbability that a single study can provide all the necessary information to link an exposure scenario to a particular health outcome in wildlife or humans, and the diverse circumstances (e.g., varied experimental conditions, numerous endpoints) from which data have been generated.

To deal with these difficulties, the authors of the 2002 report proposed a framework for assessing the relationship between exposure to endocrine disrupters and altered health outcomes. This framework is a variation on the Bradford–Hill criteria for assessing causality in epidemiology. The approach taken in WHO/IPCS (2002) was a tentative and cautious one in which the framework was tried out in a series of case studies in a separate chapter at the end of that report. It was introduced at the final meeting of the authors of the 2002 report, with the intention of assessing the applicability of this approach (cf. Tables 7.1 and 7.2, in WHO/IPCS (2002)). This is very different from a consistent application of a framework for assessing causality, and becomes evident from the articulation of several caveats in WHO/IPCS (2002): "This structured, framework approach acknowledges that 1) there are a number of scientific uncertainties, 2) a degree of scientific judgment is involved, and 3) assessments are likely to change as additional information becomes available. (…) Also, it should be noted that these assessments are qualitative determinations of the current overall state of the science. They are not quantitative risk assessments that relate specific exposure situations to probabilities of adverse effects."

Notable is the acknowledgment by the authorship of the WHO/IPCS (2002) report of the need for expert scientific judgment, as well as the delineation from quantitative risk assessment, which Lamb et al. fail to state. Thus, scientific judgment comes to bear at all stages of the process, including hypothesis formulation and evaluation against Bradford–Hill's criteria. Critical to this is the framing of a hypothesis linking an outcome of concern to a putative stressor, and this influences the outcome of the assessment most strongly. If the hypothesis is formulated too narrowly, the entire assessment may lose relevance, or even become erroneous. With the benefit of hindsight from 10 years' worth of research after 2002, which led to the discovery of numerous chemicals as endocrine disrupters, certain case study hypotheses in WHO/IPCS (2002) will today be judged as artificially restricted, and with a rather limiting focus on persistent organic pollutants such as PCBs, DDT or polychlorinated dibenzo-*p*-dioxins and dibenzofurans. For example, the 2002 report considered a case study of breast cancer, with the following hypothesis: "Increased incidences of breast cancer are caused by exposure to organochlorine chemicals (e.g., PCBs, DDT, and metabolites) possessing estrogenic activity." By application of the Bradford–Hill criteria of temporality, strength of association, consistency and biological plausibility, the 2002 report concluded that "the evidence is weak in support of the hypothesis that exposure to PCBs, DDT, and other organochlorines contribute to increased risk based on the lack of consistency of the results, weak associations, and questions of biological plausibility". This may well be so, even today. However, today we know that a more meaningful assessment of that question would focus not only on PCBs, DDT and other organochlorines, but also take account of additional agents revealed as xenoestrogens, as well as the issue of combination effects from simultaneous exposures to a large number of chemical and non-chemical risk factors including genetics.

Thus, the use of criteria for assessing causality does not per se safeguard against the temptation to buttress pet theories. As we stated in the UNEP/WHO (2013) report, "it is important to use a systematic and transparent approach to evaluating the scientific evidence about the relationship between environmental exposure and health effects". Such approaches were developed over the last 20 years from clinical science and have helped to reduce bias while integrating expert judgment, but their application to endocrine disrupting chemicals is still in their infancy (NTP, 2015). We explicitly recommended that "Efforts are needed to develop systematic and transparent approaches to identifying, evaluating and synthesizing the scientific evidence for endocrine disrupters that consider the science of endocrine action."

It is misguided to think that the application of a framework in itself protects against bias. Protection against bias is only possible by applying rigorous scientific judgment for which a multitude of approaches exists. In much the same way as in WHO/IPCS (2002), the approach used in UNEP/WHO (2013) was that of a narrative review, a method Lamb et al. construe as allowing for "a selective presentation of information without a critical review of the data." Here, Lamb et al.'s shortfalls are clear: First they suggest that a narrative review per se predisposes to selective presentation of data, and secondly, they neglect to apply the same criticism to the WHO/IPCS (2002) report.

### *2.4. The debate about the usefulness of Bradford–Hill's criteria for assessing causality*

As disapprovingly noted by Lamb et al., the UNEP/WHO (2013) report did not employ the framework for assessing causality that was suggested in WHO/IPCS (2002). They write: "Most critically, the lack of a formal framework or **standardized** approach to evaluate the data on specific chemicals and the potential causal association with adverse outcomes via an endocrine-mediated MOA is a significant shortcoming in the UNEP/WHO 2012 review (emphasis added)." The stance here is that scientific judgments derived from procedures not adhering to a process that complies with Lamb et al.'s ideas of causality must be subjective: "In the absence of a formal assessment of causation, subjective inference is relied on to suggest causation." Does this view hold any water?

We were not insensitive to the issues articulated by Lamb et al. and present below a detailed discussion of the problems associated with the Bradford–Hill criteria as a tool for judging causality (UNEP/WHO, 2013). These problems were recognized by Bradford–Hill himself (Hill, 1965) and described by us in UNEP/WHO (2013), but are consistently overlooked, not only by Lamb et al. but also in the WHO/IPCS (2002) report.

Bradford–Hill pointed out that the question of causality should not be discussed in isolation, separated from the context in which decisions have to be made whether to act on the available evidence, e.g., by adopting preventative measures. Thus, evaluating evidence is only one part of the decision making process, which must take account of many other elements important in preventing exposures. He observed that "it almost inevitably leads us to introduce **differential standards** before we convict. Thus on relatively slight

evidence we might decide to restrict the use of a drug for early-morning sickness in pregnant women. If we are wrong in deducing causation from association no great harm will be done. The good lady and the pharmaceutical industry will doubtless survive. On fair evidence we might take action on what appears to be an occupational hazard, e.g. we might change from a probably carcinogenic oil to a non-carcinogenic oil in a limited environment and without too much injustice if we are wrong. But we should need very strong evidence before we made people burn a fuel in their homes that they do not like or stop smoking the cigarettes and eating the fats and sugar that they do like (emphasis added)."

Thus, even when the outcome of an assessment employing the criteria suggests that the overall strength of association might be weak, this does not automatically exonerate from a consideration of protective measures. Lamb et al. are remiss in acknowledging that the goal of the toxicological and epidemiological sciences is not to provide assessments as an end in themselves, but to explore and evaluate conditions that offer disease prevention and public health initiatives. This cannot be realized by mechanical application of criteria for causation in the style of checklists. It is well known among epidemiologists that the implementation of preventative measures that eliminate factors which only weakly contribute to risks may reap disproportionately large benefits in terms of public health protection. An example would be excessive salt intake and its (weak) association with hypertension. Regulating salt content in foods could shift entire population distributions of blood pressure downwards, thereby decreasing the prevalence of moderate and severe hypertension (Szklo and Nieto, 2007). Lamb et al. ignore these insights.

However, the problems with using Bradford–Hill's criteria as a tool for inferring causality go much deeper, and their uncritical use trivializes the fundamental problems that exist with inferring causality from empirical observations. Indeed, Bradford–Hill himself (who never used the word "criteria" and instead chose the term "viewpoints") was fully aware of these problems when he stated that "none of my nine viewpoints can bring indisputable evidence for or against the cause and effect hypothesis and none can be required as a sine qua non…what they can do is help us to make up our minds on the answer to the fundamental question – is there another way of explaining the set of facts before us".

Lamb et al. ignore the debate of Bradford–Hill's viewpoints and the issue of assessing causality that has taken place in epidemiology (summarized by, e.g., Rothman and Greenland, 2005), in which epidemiologists have strongly refuted the existence of absolute criteria for assessing causality. It is a widely held view among experimental scientists that proof of causality is impossible in epidemiology, with the implication that the situation in the experimental sciences is better. However, David Hume's insight that proof is impossible in all empirical sciences equally applies to the experimental sciences, and Rothman and Greenland cite the cold fusion debacle in support of this notion. Numerous additional examples could be used to illustrate this.

Rothman and Greenland (2005) offered a specific critique of the utility of Bradford–Hill's criteria, and a discussion tailored to issues relevant to endocrine disruption was presented by Zoeller et al. (2014). Here, we summarize these points with reference to the modified criteria used in the WHO/IPCS (2002) case studies:

The criterion of *temporality* explores whether the putative risk factor precedes the occurrence of an effect in time. Insofar as any inference of causation must stipulate that a cause has to act before the effect can materialize, this criterion may appear clear-cut. However, there are practical difficulties, as the temporal relationship between exposure to an EDC and a specific endocrine-mediated adverse outcome may be quite complex. The classic example is that of diethylstilbestrol (DES) exposure during fetal life and the production of reproductive tract cancer 20 years later, long after DES exposure had ceased. This relationship was observed because pregnant women were prescribed DES and because there were specific records of exposure. This will not likely be the case for most non-accidental exposures to EDCs. Thus, "temporality" may be important, but it may be a concurrent relationship, i.e. exposures that are measured at the same time as the health endpoints being assessed (Zoeller et al., 2014). Furthermore, as Rothman and Greenland (2005) pointed out, the observation of a reverse time order between cause and effect (i.e., exposure and health outcome) is not evidence against the hypothesis that the putative cause can lead to the effect. Observations in which the cause followed the appearance of an effect merely show that the putative cause cannot have caused the effect in the instances that were considered. However, the generalization that the hypothesized cause is therefore not linked to the effect is fallacious.

The criterion of *strength of association* expresses the sentiment that strong associations are more likely to be causal than weak associations, which are often explained by undetected bias. However, Bradford–Hill himself did not rule out that a weak association might suggest causality, and Rothman and Greenland (2005) cite the weak association between smoking and cardiovascular disease as an example. Despite the weakness of the association, no one today would doubt that cardiovascular disease is causally linked with smoking. The reason for the weakness of the relationship is that cardiovascular disease is rather common and can be linked to other factors, and the same applies to many endocrine disorders. Conversely, there are examples of strong, but non-causal associations, and Rothman and Greenland (2005) point to the strong relation between Down syndrome and birth rank which is confounded by maternal age. These examples show that strong associations between putative risk factors and diseases do not necessarily indicate causality. The mere fact that an observed association is weak says nothing in terms of absence of causality.

The authors of WHO/IPCS (2002) have subsumed Bradford–-Hill's original viewpoint of *specificity* under that of *strength of association*. The criterion of *specificity* requires that a putative cause leads to a single, recognizable effect. The example used by Bradford–Hill (Hill, 1965) was that of nickel refiners of South Wales who suffered from a high incidence of rare cancers of the lung or nose. The specificity of this relationship could be used as evidence of causation. However, Bradford–Hill cautioned about making too much of the specificity of the relationship and concluded that, "In short, if specificity exists we may be able to draw conclusions without hesitation; if it is not apparent, we are not thereby necessarily left sitting irresolutely on the fence." As it is unrealistic to expect that a cause of a specific effect should lack other effects, the criterion of specificity is of limited usefulness in assessing causality. This point is particularly relevant to endocrine disruption where putative causes must be evaluated carefully because hormone systems are involved in a great many processes which are life-stage specific. For example, androgens play an important role in the development of the male reproductive system in the fetus, but in the adult, androgens are related to different processes in men and women. Transient hypothyroidism in the adult can lead to weight gain that is reversible but has long lasting effects on brain function when occurring in fetal and neonatal life (Zoeller et al., 2014).

*Consistency* expresses the notion that multiple studies should demonstrate the same relationships between exposure and outcome. However, as pointed out by Rothman and Greenland (2005), lack of consistency does not rule out causality, because some effects may be produced by their causes only under specific circumstances or in specific regions of the world. This point is particularly relevant for endocrine disruption considering the life-

stage specificity and geographical variations of many effects (e.g., semen quality, see Section 2.5.1 below).

Bradford–Hill insisted that "it will be helpful" if the causation we suspect is *biologically plausible*, but he was careful not to demand this in absolute terms. He cautioned that the association observed may be one new to science or medicine and therefore should not be dismissed too light-heartedly as too odd. Likewise for EDCs, biological plausibility will strengthen our confidence in the causal nature of relationships of interest. Moreover, our knowledge of hormone actions will likely drive us to evaluate specific relationships. However, there is a great deal we have yet to learn about the endocrine system, and the demand for complete knowledge of the endocrine mechanism mediating a relationship of interest is not achievable (Zoeller et al., 2014). There are numerous examples where explanations, which to contemporary observers appeared as lacking biological plausibility, turned out to be the correct ones. Accordingly, Maclure (1985) maintained that biological plausibility should be applied not to the outcome of a study, as suggested in WHO/IPCS (2002), but to the hypotheses under investigation, and pointed out "The more the data are contrary to prevailing thought, the more informative they are".

Finally, *evidence of recovery* (originally termed "experiment" by Bradford–Hill) expresses the idea that occasionally, confidence in a conclusion of causality could be strengthened by changing elements of the environments and observing a predicted change. For example, as a result of altered work conditions leading to a reduction in dust exposures, the health of workers should improve. Bradford–Hill did not include animal or biochemical experiments in this "viewpoint". At a practical level, the main problem with this "criterion" is the lack of relevant data, and indeed none of the case studies examined in WHO/IPCS (2002) had relevant data to examine this aspect. Furthermore, evidence of recovery is not a criterion to assess outcomes and to infer causality, but rather is a test of the underlying hypothesis, as was pointed out by Rothman and Greenland (2005). In most cases, such tests are simply not available, but even if they are, there is always room for alternative explanations, often rendering such tests inconclusive in the sense of establishing causality. An apt example of this is the discussion of the recovery in mollusk populations in the wake of reductions of exposure to TBT. Lamb et al. express doubts that this recovery is an indication that TBT was the cause of the declines and marshal other factors as possible explanations. In this case, the laboratory evidence linking cause and effect (in terms of TBT-induced imposex and sterility as a highly sensitive and chemical-specific phenomenon) was irrefutable and it was this observation that led to rapid control of TBT, not the evidence of population recovery that ensued (Vos et al. 2000).

If there are no absolute criteria for assessing causality, is it still possible to assess the validity of an experiment or a study? Rothman and Greenland (2005) argue: "What is required is much more than the application of a list of criteria. Instead, one must apply thorough criticism, with the goal of obtaining a quantified evaluation of the total error that afflicts the study. This type of assessment is not one that can be done easily by someone who lacks the skills and training of a scientist familiar with the subject matter and the scientific methods that were employed. Neither can it be applied readily by judges in court, nor by scientists who either lack the requisite knowledge or who do not take the time to penetrate the work."

It was therefore appropriate and fully justified to abstain from an uncritical use of Bradford–Hill's criteria in the UNEP/WHO (2013) report. In no way did this decision expose the report to bias and imbalance. Nevertheless, we recognize the importance of reaching a consensus about how to "weigh" results of epidemiological and toxicological studies and acknowledge the need for developing such approaches tailored to the specific demands of EDCs. This is discussed in more detail in Zoeller et al. (2014).

#### *2.4.1. Response to specific examples chosen by Lamb et al. to illustrate that lack of application of Bradford–Hill's criteria leads to subjective judgment*

The way in which Lamb et al. used specific examples from the UNEP/WHO (2013) report to substantiate their contention that our decision not to employ Bradford–Hill's viewpoints has led to subjective inference, can be used to illustrate the flaw in the concept that application of a framework for assessing causality per se will resolve the issues surrounding scientific judgment and thus achieve "objectivity" by "standard" application of assessment criteria. Lamb et al. propose that such a procedure, if only it were adopted universally and liked by everybody, will always produce the same conclusions. In this system, assessment outcomes that produce deviating conclusions can only be the result of misapplication of the assessment framework, and therefore have to be deemed subjective. As we show below, this ignores that assessment outcomes depend strongly on the way in which questions for assessment are framed. However, hypothesis formulation cannot be judged in categories of "objectivity".

For example, when discussing prostate cancer and possible associations with endocrine disrupters, Lamb et al. take issue with the statement in UNEP/WHO (2013) that there is sufficient evidence for associations between pesticide exposures and prostate cancer among pesticide applicators. While not denying that such associations exist, Lamb et al. argue that before we can "convict" (to use Bradford–Hill's phrase) it is necessary to show that prostate cancer is attributable to alterations in endocrine function in humans as a result of exposure to putative endocrine disrupters. While the sufferers of prostate cancer may not care much whether they contracted the disease through an endocrine disrupting mechanism, a hypothetical producer of pesticides may well be interested in resolving this rather specific question, in the interest of protecting themselves against recriminations that arise from the specifics of chemical regulation of endocrine disrupters. However, while the question of associations of pesticide exposure with prostate cancer can be resolved with currently available scientific methods, "proof" that an individual cancer has arisen via alterations of endocrine function cannot be provided with methods currently at our disposal, and probably never will be. Consequently, a study question of whether pesticide exposures increase prostate cancer risks and that this occurs through an endocrine mode of action is very likely to yield inconclusive results. However, if the study question is whether pesticide exposures increase the risk of prostate cancer, the assessment will result in a better strength of evidence. Thus, how the question is asked may matter more than the approach used to answer it. This confusion about approach and study question is promulgated by Lamb et al. But it is open to debate whether the question, framed in the way Lamb et al. advocate, has any meaning from the perspective of public health protection, and whether that assessment can claim any "objectivity".

UNEP/WHO (2013) presented a more balanced discussion of the issue. While highlighting the observed associations between pesticide exposures and prostate cancer, the report pointed out that the nature of the pesticides involved remains obscure, and that it therefore remains unclear whether the cancers arose through an endocrine-mediated mechanism. However, this lack of clarity does not weaken the conclusion that the evidence linking pesticide exposure among applicators to prostate cancer can be deemed sufficient.

The criticism formulated by Lamb et al. shows that the elusive goal of objectivity cannot be achieved just by application of an

assessment framework based on Bradford–Hill's viewpoints in the style of a checklist, because a great deal is determined by hypothesis formulation. However, as we have seen, hypothesis formulation itself cannot claim objectivity – hypotheses are framed in a context of specific interests and are subject to scientific judgment. The commercial interests supporting Lamb et al. will likely always prefer articulation of hypotheses in the narrowest possible way and such narrow hypotheses will, through application of "standard" procedures, almost always return the standard verdict "inconclusive", with the consequence of abstaining from recommendations of emission-reducing measures. From the viewpoint of public health protection, however, it is important to formulate more meaningful hypotheses that take account of multiple exposures and multi-causality.

### 2.5. Response to the specific examples of alleged imbalance and bias highlighted by Lamb et al.

In the sections below we give specific responses to the examples Lamb et al. (2014) selected from the UNEP/WHO (2013) report to show "why and how the UNEP/WHO 2012 report is not an update of the WHO-IPCS 2002 report…" and how bias and subjectivity allegedly guided our analysis and our conclusions.

#### 2.5.1. Semen quality

Lamb et al. highlight a "discrepancy" between the conclusions in UNEP/WHO (2013) and WHO/IPCS (2002) about semen quality which they find "difficult to explain" considering that "it does not appear that the evidence for changes in sperm quality differ (sic) from that reported in the WHO-IPCS (2002) report." They conclude: "When an objective, structured, and transparent weight-of-the-evidence analysis reaches one conclusion and a subjective analysis concludes the opposite, logic dictates the driving force for such a difference stems from the methodology and bias inherent in a subjective analysis."

However, the logic of this argumentation collapses when the alleged discrepancy between the conclusions of the "objective, structured and transparent" WHO/IPCS (2002) and our "subjective analysis" in UNEP/WHO (2013) can be shown not to exist.

Two key issues surround the debate about semen quality: First, have there been deteriorations of semen quality, and secondly, is there any evidence of associations with endocrine disrupting chemicals?

The WHO/IPCS (2002) report concluded that "Viewed as a whole, several of the published reports support the hypothesis that there are time-related decreases in semen quality at least within some regions, as reflected in sperm concentration and, where measured, sperm motility and morphology, but do not support the hypothesis that the decline is worldwide." By analyzing the available studies using Bradford–Hill's viewpoint of consistency, the 2002 report listed ten studies showing a decline, six studies with evidence for improvement and eight studies where no change was found. It was clear already in 2002, that changes in semen quality strongly depended on location, among other factors, and that a world-wide decline was not supported by the data.

Lamb et al. disregard that, like the WHO/IPCS (2002) report, UNEP/WHO (2013) highlighted the geographical variations in semen quality that were apparent also in the new studies conducted after 2002. The 2013 report made no claims about worldwide declining trends. However, Lamb et al. base their allegation of bias and selective citation on the fact that we did not cite four studies of semen quality which did not report semen quality deteriorations. They mislead the reader by obscuring that citation of these studies would not have altered our conclusion that a worldwide decline does not exist, but that there are regions with alarming declines. Citation of these studies would therefore not have added anything to our report.

Importantly therefore, the undisputed fact of geographical variations cannot distract from observations of alarming declines in semen quality in some regions, and this was emphasized in UNEP/WHO (2013). Lamb et al. confuse the issue by criticizing the 2013 report for focusing on the decreased semen quality, thereby "ignoring the variability in sperm quality reported around the world". This misses the point entirely and is an example of misuse of Bradford–Hill's viewpoint of consistency. As emphasized earlier, a lack of consistency does not rule out causality, and cannot be used to argue away that some populations have experienced declines in semen quality.

Lamb et al. further make the false allegation that UNEP/WHO (2013) did not discuss limitations and biases of semen quality studies, when in fact our report pointed out that "The reason for the controversy may partly be explained by…differences in methods for semen analysis and variation in results within individuals…".

Regarding the issue of associations of declines in semen quality with exposures to endocrine disrupting chemicals, Lamb et al. claim that UNEP/WHO (2013) used observations of suboptimal or poor semen quality in some parts of the world to "suggest that these **perceived** trends are the consequence of exposure to endocrine disruptors with no strong evidence to support this claim" (emphasis added). To pause on the effrontery of re-interpreting data of observed semen quality declines as "perceived" trends would indeed miss the point: UNEP/WHO (2013) clearly states "A role for EDCs has been hypothesized, but to date there are no clear data except for some rare cases of environmental or occupational accidents." This conclusion is not different from WHO/IPCS (2002), but this fact is omitted by Lamb et al.

The 2013 report highlights the difficulties in unraveling the possible contribution of endocrine disrupting chemicals that stem from prenatal exposures. The few studies of possible associations with chemicals have all focused on exposures experienced in adulthood, and may therefore have missed periods of heightened sensitivity in fetal or neonatal life, another point entirely disregarded by Lamb et al.

Regarding semen quality, the main conclusion of the UNEP/WHO (2013) report is that semen quality among 20–40% of young men in several European countries is in the sub fertile range. This conclusion is based on new studies conducted after 2002, but does not disagree with the trends already discernible in WHO/IPCS (2002). The 2013 report also concludes that "there is, however, very little direct evidence for a role of endocrine disrupting chemicals in causing low semen quality in men following developmental exposures", quite similar to the stance taken in WHO/IPCS (2002).

Thus, what Lamb et al. declare as a "discrepancy" between the conclusions of the two reports is nothing but exaggeration and misdirection. Their claim that UNEP/WHO (2013) has "*concluded the opposite*" and that "the driving force for such a difference stems from the methodology and bias inherent in a subjective analysis [of UNEP/WHO, 2013]" is baseless and the result of an extraordinary distortion of the facts by misquoting the evidence.

#### 2.5.2. Adrenal disorders

The way in which UNEP/WHO (2013) dealt with the issue of population declines and adrenocortical hyperplasia in Baltic seals and its possible association with persistent organic pollutants is cited by Lamb et al. as a specific example of "selective citation of literature without discussion of contradictory studies and the failure to consider alternative causes of reported effects" which gives the reader "the impression that the weight of evidence is stronger than is justified by the available scientific data". They go on to claim that this "calls into question the integrity of decisions at all levels of

the 2012 report".

In their attempt to substantiate these serious allegations, Lamb et al. misrepresent both the WHO/IPCS (2002) and the UNEP/WHO (2013) report.

They write that "it was stated in the 2002 report that although adrenal effects in wildlife were associated with dichloro-diphenyldichloroethane (DDD), DDT, and PCBs, the involvement of these compounds in the cause of these disorders was uncertain."

However, in contrast to the portrayal by Lamb et al., the 2002 report actually stated that "Although overhunting and habitat destruction may have been contributing factors, it is generally accepted that persistent pollutants, which adversely affected the reproductive performance of the females, resulted in the decline in seal numbers." The issue was also chosen as one of the case studies for evaluating causality, and the WHO/IPCS (2002) came to the conclusion that the overall evidence that persistent organic pollu-tants including PCBs contribute to reproductive toxicity in Baltic seal through an endocrine mechanism is moderate, but not uncertain as claimed by Lamb et al. In the overall assessment the 2002 report states: "Relative to the outcome of concern [reproductive failure in Baltic seals] there is considerable evidence that the reproductive success of Baltic seal populations has been impacted and that these exposures have altered adrenal gland function in members of the exposed population. However, because the link between altered adrenal function and reproductive impairment has not been clearly established, the overall evidence of an EDC-related mode of action is **moderate** (emphasis added)."

Since 2002, new evidence concerning adrenocortical hyperplasia in Baltic seals and a suite of alterations characteristic for Cushing disease in humans has come to light, disregarded by Lamb et al., but discussed in UNEP/WHO (2013). Nevertheless, the position adopted in UNEP/WHO (2013) is much more nuanced than Lamb et al. suggest. The 2013 report emphasizes that "the exact mechanisms behind the Cushing-like condition remain unclear. Likewise, the individual compounds producing these adrenal lesions remain unknown. Although it cannot be ruled out that a component of stress was playing a role, it seems likely that persistent exposure to organohalogens were involved." Thus, Lamb et al.'s claim of selective citation of literature without consideration of alternative explanations in the UNEP/WHO report is without substance.

### 2.5.3. Endometriosis

The way in which UNEP/WHO (2013) dealt with the topic of endometriosis is also held up by Lamb et al. as an example of selective citation of evidence. More specifically, Lamb et al. stress that "the UNEP/WHO 2012 discussion on endometriosis is not an update of the WHO-IPCS 2002 report as it mainly consists of a re-review of the information evaluated in 2002". They go on to say "It is not clear what, if anything, has changed regarding the state of the science in terms of cause and effect for endocrine disruption and endometriosis since the WHO-IPCS report". These claims are false and a marked distortion of the content of the 2013 report.

UNEP/WHO (2013) deals with the topic by first giving an explanation of the natural history of endometriosis, with new information about the increasing disease incidence, and an update about risk factors, based on no fewer than 14 publications that appeared after 2002 and which, therefore, were not included in WHO/IPCS (2002).

UNEP/WHO (2013) then proceeds to a discussion of hormonal mechanisms implicated in endometriosis, with the relatively new hypothesis that the origin of this disorder might be in fetal life. The involvement of the immune system and the necessity of estrogens for disease progression are also considered. Nine references with a publication date after 2002 are cited to substantiate this discussion.

The 2013 report then turns to describing animal models of endometriosis. In relation to this discussion Lamb et al. claim that we failed to note that many of the studies linking TCDD to endometriosis employed high doses of TCDD. This criticism is also misleading, since in fact we wrote that "this includes a study of twenty rhesus monkeys dosed and followed for 15 years, which reported an increase in incidence and severity with higher dioxin exposures".

The section on animal models details four publications that appeared after 2002. These studies provide evidence for the involvement of epigenetic changes in endometriosis and in support of the idea that such changes can be induced by in utero exposure to chemicals.

Finally, the UNEP/WHO (2013) considers epidemiological evidence for the involvement of endocrine disrupters in endometriosis and highlights that most available studies focused on adults, with the implication that periods of heightened sensitivity (e.g., exposures in fetal life) might have been missed. This section of the report mentioned seven studies that did not report any associations between endometriosis and environmental exposures, of which five appeared after 2002. There were, however, associations with PCBs and six studies were cited in support of this idea, of which four appeared after 2002. The 2013 report also mentioned the associations between endometriosis and phthalate esters (nine studies, of which seven appeared after 2002) and one study reporting a link with cadmium. Only one study was reported that dealt with fetal or early life exposures to DES. We also highlighted that there is a strong genetic component in endometriosis and that certain genotypes might be particularly susceptible to the disease.

Considering these details, it is difficult to recapitulate how Lamb et al. (2014) could arrive at the conclusion that "it is not clear what, if anything, has changed regarding the state of the science in terms of cause and effect for endocrine disruption and endometriosis since the WHO-IPCS report" and that "this illustrates that the UNEP/WHO 2012 report is not an update of the WHO-IPCS 2002 report".

### 2.6. What do temporal trends in diseases show?

The UNEP/WHO (2013) report recorded incidence trends of many endocrine-relevant disorders and emphasized that worldwide, there has been a failure to adequately address the underlying environmental causes of these trends. We highlighted that the increase in disease incidence and prevalence has occurred too fast for genetic causes to provide plausible explanations, and concluded that environmental factors in a wider sense must be at play. These factors include diet, exercise, lifestyle factors and chemical exposures. Lamb et al. claim our report did not acknowledge that "much of the time, the environmental causes of the diseases being discussed are not chemicals exposures." This claim is false. We have acknowledged non-chemical risk factors on numerous occasions, but due to the topic of our report, had to focus on the issue of chemical exposures.

In a familiar pattern, Lamb et al. then allege that "through selective citation in the UNEP/WHO 2012 report, an impression is created that certain diseases have an increasing incidence or prevalence." In support of their claim, they point out that several papers were cited in our report to indicate rising prevalence of hypospadias, but that we did not cite Fisch et al. (2010) who stated that "a review of the epidemiological data on this issue amassed to date clearly demonstrates that the bulk of evidence refutes claims for an increase in hypospadias rates." That paper was indeed not cited, and for good reason: Fisch et al. is an analysis that explores the absurd hypothesis that hypospadia incidences have risen in every country. They succeeded in demolishing that hypothesis, only to conclude that, consequently, rising incidence of hypospadias is

not occurring anywhere. We ensured that papers of such low standards were excluded from consideration.

In their discussion of wildlife population declines Lamb et al. (2014) attempt to cast doubt on the effects of TBT on gastropods and of DDT/DDE on bird egg shells. Concerning the effects of TBT on gastropods, Lamb et al. (2014) dismiss the statement of Jörundsdottir et al. (2005) that the impacts on gastropods around larger harbors are presumably associated with the continued use of TBT, on the grounds that no measurements of TBT concentrations were made. However, in a more recent paper, the same research group (Gudmundsdottir et al., 2011) finds a strong correlation between organic tin concentrations and imposex in the dogwhelk *Nucella lapillus*. Morton (2009) is cited by Lamb et al. (2014) in support of the idea that "due to lack of confirmatory chemical data, the changes in population size, structure, and reproduction herein reported upon for *N. lapillus* cannot be correlated positively with changes in the ambient TBT levels". Here, Lamb et al. (2014) quote out of context and omit the continuation of Morton's sentence "… but they can and are correlated with freedom from imposex. This is the first time such a dramatic recovery from imposex, following the banning of TBT, has been documented".

Lamb et al. (2014) emphasize the lack of sensitivity of a number of gastropod species to TBT-induced imposex; however, this is misleading. The fact that one species may be less sensitive is irrelevant to an assessment of the effects in a more sensitive species. Moreover, this emphasis on species that appear insensitive to TBT distracts from the fact well known to ecotoxicologists that imposex has been reported in over 170 species of neogastropod mollusks to date; hardly a small number.

Lamb et al. (2014) assert that "the WHO-UNEP 2013 report does not note the lack of agreement among researchers on the mechanism for induction of effects" in mollusks. This is false; page 46 in the WHO-UNEP (2013) report says "Although the effects of TBT on the reproductive system of female gastropods are well-established, the underlying mechanisms are not yet understood". Following this statement, four suggested mechanisms – which do not mutually exclude each other – are described.

Discussing DDT and its effects on bird populations, Lamb et al. (2014) reference Henny et al. (2010) to state that "a number of confounding factors are likely to have affected the recovery of Osprey populations". Several species of predatory and fish-eating birds were severely affected by the exposure to DDT/DDE and it is true that the recovery of some of these bird species in some geographical regions was aided by various kinds of management (i.e., supplying nest sites). However, this does not change the fact that exposure to DDT/DDE caused the population declines in the first place and that the decreasing environmental concentrations of DDT after the ban was a requirement for the re-establishment of the populations. As with the case of TBT, it was experimental data gleaned from a number of laboratory studies that clearly confirmed the reality of the epidemiological link between DDT/DDE exposure and the effects in birds and not only the recovery of the bird populations.

Lamb et al. (2014) also use the example of ospreys and decreasing DDE concentrations in their general attempt to discredit the scientific credibility of the UNEP/WHO 2013 report. Lamb et al. (2014) call it a mischaracterized citation that we have drawn a best fit line through the values on DDE concentrations taken from Henny et al. (2010) because such a line was not drawn by the original authors. However, a simple regression on Henny et al.'s (2010) data clearly shows a decreasing trend ($p = 0.003$; $r^2 = 0.91$). In the scientific literature, there are numerous examples of similar decreases of DDT/DDE concentrations in various organisms after the ban of DDT (some of them presented in figure 3.18 and 3.40 in the UNEP/WHO, 2013 report).

Lamb et al. (2014) repeatedly praise the WHO/ICPS (2002) report for its approach and quality. For the overall assessment of the effects of TBT and DDT/DDE, it is noteworthy that in the 2002 report (WHO/IPCS, 2002) (to which two of the co-authors of Lamb et al. (2014), Foster and Van Der Kraak, contributed) the overall strength of the evidence was assessed as 'strong' for the association between exposure to TBT and the induction of imposex in marine gastropods (both for the hypothesis and for the EDC mechanism). For the association between eggshell thinning in colonial water birds and DDE, the hypothesis and evidence for EDC mechanism were assessed as 'strong' and 'moderate', respectively. It is unclear what information led Foster and Van Der Kraak to cast doubt on the UNEP/WHO 2013 assessments in the Lamb et al. critique, since they do not differ from the assessments made in WHO/IPCS (2002).

### 2.7. The debate about dose–response relationships and potency

Lamb et al. discuss features of dose–response relationships in the context of the characteristics of the endocrine system. They criticize our perceived failure to discuss that "the endocrine system is specifically designed to respond to environmental fluctuations and such homeostatic responses generally are considered normal, adaptive, and necessary as long as they are transient and within the normal homeostatic range". As already discussed by Zoeller et al. (2014), this criticism misses the point of endocrine disruption entirely. By focusing on the role of the endocrine system in maintaining homeostasis, Lamb et al. ignore that the endocrine system also has the role of programming development and fail to discuss that disruption of the programming functions of the endocrine system during specific windows of susceptibility can have irreversible consequences. They misconstrue our report when they say that "the report suggests that at vulnerable developmental stages potency may not be very relevant". Potency is of course relevant, but it cannot be the sole and decisive criterion when it comes to regulating endocrine disrupters. Timing of exposure, irreversibility of effect and other criteria also have to be considered (Kortenkamp et al., 2011).

Lamb et al. also present a summary of the debates surrounding the topic of non-monotonic dose–response relationships, but criticize the review by Vandenberg et al. (2012) rather than the UNEP/WHO (2013) report, confusing the positions articulated in Vandenberg with those in our report.

### 2.8. Similarities with misrepresentations of scientific evidence used by the tobacco industry to block contemporary tobacco regulation

The techniques used by Lamb et al. to discredit UNEP/WHO (2013) have striking similarities to those currently employed by the tobacco industry in the ongoing debate about tobacco regulation through standardized packaging intended to remove all brand imagery and text. An analysis conducted by Ulucanlar et al. (2014) of tobacco industry interventions shows how these are based on a number of techniques designed to misrepresent the scientific evidence. Many of these techniques were also used by Lamb et al. (2014).

At the most basic level there is misleading quoting of evidence through inaccurate reporting of study objectives, methods and findings. This involves what is referred to as the "tweezers" method of partially and selectively quoting the original source, thereby omitting qualifying information (Ulucanlar et al., 2014). Lamb et al., knowingly or not, have used this technique most frequently in relation to the WHO/IPCS 2002 report, by distorting the methods used in that report and by omitting qualifying statements particularly in relation to the use of Bradford–Hill's viewpoints. Other examples of the "tweezers" method can be found in the way Lamb

et al. frame the debate about TBT and DDE and their effects on wildlife.

A second tactic identified by Ulucanlar et al. (2014) in tobacco industry submissions and also widely observed in the Lamb et al. critique can be called "mimicked scientific critique". It involves the detailed inspection of scientific reports for methodological rigor with the aim of rejecting an entire body of evidence as flawed or biased. In reaching that conclusion, scientific methods of critique appear to be used, but only close examination of the approach reveals that these critiques are embedded in an essentially unscientific paradigm. As identified by Ulucanlar, these interventions are "mimicked" versions of scientific critique because they seek methodological perfection by judging studies against unrealistic and perfectionist criteria, and insist on methodological uniformity, among others.

In the preceding sections we have identified several examples where Lamb et al. judged scientific studies against unrealistic criteria, such as in the discussion of TBT and DDE and wildlife effects or in the framing of study questions in epidemiology in terms of very narrow hypotheses ("prostate cancer as a result of endocrine mechanisms").

As pointed out by Ulucanlar et al. (2014), this tactic works at a deeper level than the "tweezers" method, because its identification and critique requires a good epistemological understanding of science, specialized expertise, knowledge of research traditions and methodologies and deep skills in interpretative analysis, all of which are unlikely to be found among non-specialists and decision makers in the civil service or similar bureaucracies.

In dismissing UNEP/WHO (2013) as flawed and biased on the grounds that a narrative synthesis of the evidence was used, rather than a more quantitative meta-analysis, Lamb et al. insisted on methodological uniformity whereby extremely narrowly defined standards are used to dismiss any review that does not fit their preferred approach. Narrative synthesis, as used in UNEP/WHO (2013) and, counter to the claims of Lamb et al., also in WHO/IPCS (2002), is an established method with authoritative guidelines available for its conduct (Centre for Review and Dissemination, 2009). In insisting on methodological uniformity, Lamb et al. display a fundamental failure to understand the requirement for methodological pluralism in science.

### *2.9. State of the Science of Endocrine Disrupting Chemicals 2012 – summary for decision-makers*

Lamb et al. state that "The relationship between the 2012 main report and the Summary for Decision-Makers is confusing at best. Based on the title of this document, one might presume that this document is a summary of – or at least based on the analysis of – the main report. But a closer look reveals that the Summary is actually characterized as "another product" of the process." The Summary for Decision-Makers is indeed not an executive summary of the main report. An executive summary is contained in the main report. Instead the "Summary for Decision-Makers" was written to help non-scientific experts and especially decision makers, to understand the science reported in UNEP/WHO (2013). This practice is commonly applied by intergovernmental organizations, a detail overlooked by Lamb et al. (2014). Lamb et al. (2014) also fail to see that the scientific substance of the Summary for Decision-Makers does not differ from that of the UNEP/WHO (2013) report, despite the stylistic differences they misleadingly emphasize.

## 3. Conclusions

We conclude that the criticism of the UNEP/WHO (2013) report presented by Lamb et al. is without basis. It creates the false impression of scientific controversy and does not engage with the scientific substance of our report. As we have shown, Lamb et al.'s attempt of deconstructing the UNEP/WHO (2013) report is not particularly erudite scientifically. It appears that the critique is not intended to be persuasive to the scientific community, but is designed to speak to bureaucrats, politicians and other decision makers not intimately familiar with the topic of endocrine disruption and therefore susceptible to false generalizations of bias and subjectivity. It is important to recognize that while we drafted this response, evidence has come to light of sustained and concerted lobbying efforts by American and European chemical industries aimed at delaying European Commission efforts to implement regulations for endocrine disrupters (Horel, 2015), especially in the context of the ongoing negotiations of the Transatlantic Trade and Investment Partnership (TTIP) between the USA and the EU. Importantly, many of these same lobbying organizations have funded the critique by Lamb et al. (2014). The alignment of the conclusions of Lamb et al. with the goal of the lobbying efforts of their funders certainly creates the appearance of a conflict of interest. Moreover, Lamb et al.'s lack of focus on scientific issues and the similarity in their type of criticisms with those developed by the Tobacco Industry make it difficult to conclude anything but that the funders recruited writers known to have industry-friendly themes in their historical writing to "critique" the 2013 UNEP/WHO report.

Finally, the concept of "controversy" in science must be viewed appropriately, as articulated well by Blaise Pascal in the 1600s (Auden and Kronenberger, 1966): "Contradiction is not an indication of false, nor is lack of contradiction a sign of truth."

## Conflicts of interest

As in our efforts to produce the 2013 UNEP/WHO report, the authors were not supported by any outside stakeholders in writing this response. There are therefore no conflicts of interests as coauthors of the present article, and all coauthors have declared this in writing.

## Disclaimer

The views expressed in the present article are the collective views of the authors and are not necessarily the views of their organizations.

## Acknowledgments

We, the authors, have prepared this response to the Lamb et al. (2014) article without any specific funding but in our position as experts on EDCs and as authors of the UNEP/WHO 2013 reports.

## Transparency document

Transparency document related to this article can be found online at http://dx.doi.org/10.1016/j.yrtph.2015.07.026.

## References

Auden, Kronenberger, 1966. The Viking Book of Aphorisms. New York.
Boobis, A.R., Cohen, S.M., Dellarco, V., McGregor, D., Meek, M.E., Vickers, C., et al., 2006. IPCS framework for analyzing the relevance of a cancer mode of action for humans. Crit. Rev. Toxicol. 36 (10), 781–792.
Centres for Review and Dissemination, 2009. Systematic Reviews: CRD's Guidance for Undertaking Reviews in Healthcare. University of York, York. Available: http://www.york.ac.uk/inst/crd/pdf/Systematic_Reviews.pdf.
Fisch, H., Hyun, G., Hensle, T.W., 2010. Rising hypospadias rates: disproving a myth. J. Pediatr. Urol. 6, 37–39.
Gudmundsdottir, L.O., Ho, K.K.Y., Lam, J.C.W., Svavarsson, J., Leung, K.M.Y., 2011.

Long-term temporal trends (1992–2008) of imposex status associated with organotin contamination in the dogwhelk *Nucella lapillus* along the Icelandic coast. Mar. Pollut. Bull. 63, 500–507.

Henny, C.J., Grove, R.A., Kaiser, J.L., Johnson, B.L., 2010. North American osprey populations and contaminants: historic and contemporary perspectives. J. Toxicol. Environ. Health Part B Crit. Rev. 13, 579–603.

Hill, A.B., 1965. The environment and disease: association or causation? Proc. R. Soc. Med. 58, 295–300.

Horel, S., 2015. A Toxic Affair – How the Chemical Lobby Blocked Action on Endocrine Disrupting Chemicals. Available: http://corporateeurope.org/food-and-agriculture/2015/05/toxic-affair-how-chemical-lobby-blocked-action-hormone-disrupting.

Jorundsdottir, K., Svavarsson, J., Leung, K.M.Y., 2005. Imposex levels in the dogwhelk *Nucella lapillus* (L.)-continuing improvement at high latitudes. Mar. Pollut. Bull. 51, 744–749.

Kortenkamp, A., Martin, O., Faust, M., Evans, R., McKinlay, R., Orton, F., Rosivatz, E., 2011. State of the Art Assessment of Endocrine Disrupters. Final Report. [Online] Available at: http://ec.europa.eu/environment/endocrine/documents/4_SOTA%20EDC%20Final%20Report%20V3%206%20Feb%2012.pdf.

Lamb IV, J.C., Boffetta, P., Foster, W.C., Goodman, J.E., Hentz, K.L., Rhomberg, L.R., Staveley, J., Swaen, G., Van Der Kraak, G., Williams, A.L., 2014. Critical comments on the WHO-UNEP State of the Science of Endocrine Disrupting Chemicals 2012. Regul. Toxicol. Pharmacol. 69, 22–40.

Maclure, M., 1985. Popperian refutation in epidemiology. Am. J. Epidemiol. 121, 343–350.

Morton, B., 2009. Recovery from imposex by a population of the dogwhelk, *Nucella lapillus* (Gastropoda: Caenogastropoda), on the southeastern coast of England since May 2004: a 52-month study. Mar. Pollut. Bull. 58, 1530–1538.

NTP, 2015. National Toxicology Program. Handbook for Conducting a Literature-based Health Assessment Using Office of Health Assessment and Translation (OHAT) Approach for Systematic Review and Evidence Integration. January 9, 2015 release. Available at: http://ntp.niehs.nih.gov/go/38673.

Rothman, K.J., Greenland, S., 2005. Causation and causal inference in epidemiology. Am. J. Public Health 95 (S1), S144–S150.

Sonich-Mullin, C., Fielder, R., Wiltse, J., Baetcke, K., Dempsey, J., Fenner-Crisp, R., et al., 2001. IPCS conceptual framework for evaluating a mode of action for chemical carcinogenesis. Regul. Toxicol. Pharmacol. 34, 146–152.

Szklo, M., Nieto, F.J., 2007. Epidemiology Beyond the Basics. Jones and Bartlett Publishers, ISBN 0-7637-2927-2.

Ulucanlar, S., Fooks, G.J., Hatchard, J.L., Gilmore, A.B., 2014. Representation and misrepresentation of scientific evidence in contemporary tobacco regulation: a review of tobacco industry submissions to the UK Government consultation on standardised packaging. PLoS Med. 11 (3), e1001629.

UNEP/WHO, 2013a. World Health Organization, United Nations Environment Programme (WHO-UNEP). In: Bergman, A., Heindel, J.J., Jobling, S., Kidd, K.A., Zoeller, R.T. (Eds.), State of the Science of Endocrine Disrupting Chemicals. Available at: http://www.who.int/ceh/publications/endocrine/en/index.html.

UNEP/WHO, 2013b. World Health Organization, United Nations Environment Programme (WHO-UNEP). In: Bergman, A., Heindel, J.J., Jobling, S., Kidd, K.A., Zoeller, R.T. (Eds.), State of the Science of Endocrine Disrupting Chemicals. Available at: http://www.who.int/ceh/publications/endocrine/en/index.html.

Vandenberg, L.N., Colborn, T., Hayes, T.B., Heindel, J.J., Jacobs Jr., D.R., Lee, D.H., Shioda, T., Soto, A.M., vom Saal, F.S., Welshons, W.V., Zoeller, R.T., Myers, J.P., 2012. Hormones and endocrine-disrupting chemicals: low-dose effects and nonmonotonic dose responses. Endocr. Rev. 33 (3), 378–455.

Vos, J.G., Dybing, E., Greim, H.A., Ladefoged, O., Lambré, C., Tarazona, J.V., Brandt, I., Vethaak, A.D., 2000. Health effects of endocrine-disrupting chemicals on wildlife, with special reference to the European situation. Crit. Rev. Toxicol. 30, 71–133.

WHO/IPCS, 2002. World Health Organization, International Programme on Chemical Safety (WHOIPCS). In: Damstra, T., Barlow, S., Bergman, A., Kavlock, R., Van Der Kraak, G. (Eds.), Global Assessment of the State-of-the-science of Endocrine Disrupters. World Health Organization, Geneva. Available at: http://www.who.int/ipcs/publications/new_issues/endocrine_disruptors/en/.

Zoeller, R.T., Bergman, A., Becher, G., Bjerregaard, P., et al., 2014 Dec 22. A path forward in the debate over health impacts of endocrine disrupting chemicals. Environ. Health 14, 118. http://dx.doi.org/10.1186/1476-069X-13-118.