Critical comments on the WHO-UNEP State of the Science of Endocrine Disrupting Chemicals – 2012

James C. Lamb IV a,†, Paolo Boffetta b, Warren G. Foster c, Julie E. Goodman d, Karyn L. Hentz a, Lorenz R. Rhomberg d, Jane Staveley a, Gerard Swaen e, Glen Van Der Kraak f, Amy L. Williams a

a Exponent, 1800 Diagonal Road, Suite #500, Alexandria, VA 22314, USA
b The Tisch Cancer Institute and Institute for Translational Epidemiology, Icahn School of Medicine at Mount Sinai, 17 East 102 Street Floor West Tower, 5th Floor Room 5-142, New York, NY 10029, USA
c Department of Obstetrics & Gynecology, McMaster University, 1280 Main Street West, Hamilton, ON L8S 4K1, Canada
d Gradient, 20 University Road, Cambridge, MA 02138, USA
e Exponent, The Lenz, 1st Floor, Hornbeam Park, Harrogate HG2 8RE, UK
f Department of Integrative Biology, University of Guelph, Guelph, ON N1G 2W1, Canada

ABSTRACT

Early in 2013, the World Health Organization (WHO) released a 2012 update to the 2002 State of the Science of Endocrine Disrupting Chemicals. Several significant concerns have been identified that raise questions about conclusions reached in this report regarding endocrine disruption. First, the report is not a state-of-the-science review and does not follow the 2002 WHO recommended weight-of-evidence approach. Second, endocrine disruption is often presumed to occur based on exposure or a potential mechanism despite a lack of evidence to show that chemicals are causally established as endocrine disruptors. Additionally, causation is often inferred by the presentation of a series of unrelated facts, which collectively do not demonstrate causation. Third, trends in disease incidence or prevalence are discussed without regard to known causes or risk factors; endocrine disruption is implicated as the reason for such trends in the absence of evidence. Fourth, dose and potency are ignored for most chemicals discussed. Finally, controversial topics (i.e., low dose effects, non-monotonic dose response) are presented in a one-sided manner and these topics are important to understanding endocrine disruption. Overall, the 2012 report does not provide a balanced perspective, nor does it accurately reflect the state of the science on endocrine disruption.

1. Introduction

In 2002, the World Health Organization (WHO), in collaboration with the United Nations Environment Programme (UNEP), published what is presented as an “update” to the 2002 WHO-UNEP, 2012b report: State of the Science of Endocrine Disrupting Chemicals – 2012 Summary for Decision-Makers (WHO-UNEP, 2012b). The WHO-IPCS 2002 report was not an assessment of particular agents or risks, but set out to summarize the prevailing state of scientific knowledge – what was known, what was uncertain, and what the prospects were for resolving the uncertainties with...
further research and data collection regarding endocrine disruption. The report described patterns in natural human and animal populations that were considered possible manifestations of endocrine disruption and assessed the basis for evaluating whether these patterns should be regarded as real and robust, whether explanations for them other than endocrine disruption could be possible, and what might be the state of toxicological evidence for attributing them to an interference with endocrine-mediated control by environmental chemicals at prevailing environmental concentrations. Issues under debate were described forthrightly along with the nature and extent of evidence available to support the differing points of view. Importantly, the assessment was notable not only for its product, but also its process. A large and widely representative set of international experts, including those with a variety of views, articulated and employed a weight-of-evidence methodology to integrate various kinds and lines of evidence and to gauge how, and how well, the collective evidence supported conclusions, which were then extensively reviewed. The aim was to be appropriately circumspect, yet earnestly probing – that is, neither to be alarmist, focusing only on feared possibilities, nor to be complacent and dismissive of concerns that had yet to be adequately supported scientifically. The WHO-IPCS 2002 assessment largely succeeded in these aims and it won wide acceptance and respect as an objective picture of what science had to say (and the limits as to what current knowledge allowed it to say) about the possibilities, prevalence, and magnitude of impacts of environmental chemicals on natural populations through interaction with endocrine systems.

Unfortunately, the 2012 report falls well short of the standard set by the earlier 2002 assessment, both in its openness and objectiveness of process and as a substantial evaluation of current scientific knowledge and thinking on the issues. Whereas the 2002 assessment was produced by consensus among a large set of scientists spanning the range of views on the matter, the 2012 report was produced by a more limited set of authors. The 2002 report articulated and used a weight-of-evidence evaluation process and, while the 2012 report criticizes that process, it does not replace it with anything else, relying instead on an unexplained “best professional judgment.” 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.

The present paper identifies several concerns regarding the WHO-UNEP 2012 report. Namely, the report fails to present an objective assessment of the current state of the science of endocrine disruption and does not, in fact, serve to update the 2002 assessment. Instead, the 2012 report seeks to replace the earlier assessment with a much less thoroughly reasoned evaluation that stresses possibilities of concern rather than an assessment of evidence about whether those possibilities result in real human health or environmental problems. An underlying concern with the report is the presentation of evidence in a manner that infers that the information demonstrates endocrine disruption without full consideration of alternative explanations for the observed effects. This is partially achieved by the imprecise use of key terms or concepts. For example, throughout the report there are sections titled “Epidemiological evidence for EDCs [endocrine disrupting chemicals] causing [insert health effect under discussion, e.g., early puberty].” This title gives the reader an impression that evidence will be presented on chemicals that cause that particular effect, when these sections should have more appropriately been characterized as a discussion of EDCs associated with these effects. Section 2.4 provides other examples and more detail on the use of inference to imply rather than show that EDCs are causally associated with certain effects. The following Table 1 provides a summary of key terms, with their definitions, as used in this paper.

The observations, comments and criticisms of the WHO-UNEP 2012 report are provided with further discussion and examples below. The aim of this critique is not to reevaluate these points, nor to conduct a comprehensive assessment of the issue of endocrine disruption or the 2012 report. Rather this critique illustrates, with specific examples, where the 2012 report has made statements that claim or imply a finding about endocrine disruption as a cause of actual effects, but have not been supported by a balanced and thorough evaluation of the pertinent evidence. New data and new understanding of the endocrine activity of chemicals have been developed since 2002 and using this information to build on the 2002 analysis is worthwhile – but, the WHO-UNEP 2012 report does not achieve this goal. This paper focuses on the limitations of the WHO-UNEP 2012 report and outlines specific concerns that include: the inconsistency between the Summary for Decision-Makers and the main report; the lack of a transparent and systematic framework for identifying, reviewing, and evaluating data; the failure to update the 2002 WHO-IPCS report as stated; the informal approach to assessing causation from endocrine disrupting chemicals (EDCs); the reliance on disease trends to suggest associations with EDCs; and ignoring the role of exposure, dose, and potency in endocrine disruption. Each of these key concerns is presented below in the Discussion section and includes specific examples of limitations in the WHO-UNEP 2012 report.

2. Discussion

2.1. Companion report: summary for decision-makers

In addition to the State of the Science of Endocrine Disrupting Chemicals – 2012 main report, a second publication, State of the Science of Endocrine Disrupting Chemicals 2012 Summary for Decision-Makers was simultaneously released (WHO-UNEP, 2012b). 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. In some cases, the Summary does present an overview of key findings from the main report, but there are many parts of the Summary which include conclusions and assertions not reflected in the main report. Indeed, some conclusions are matters not mentioned at all in the main report. It is very important to draw this distinction and make clear that the Summary for Decision-Makers is not truly representative of the content of the main report. Thus, there are even more shortcomings in the Summary for Decision-Makers than in the 2012 report itself.

The Summary for Decision-Makers presents a broader scope in discussions and the statements are presented as more definitive conclusions compared to those in the main report. For example, a list of diseases are presented in Figure 5 of the Summary and described as being induced by endocrine disrupting chemicals (EDCs); however, no references are provided in the Summary to support these inferences and insufficient data are presented in the main report itself to show that these diseases are in fact induced or caused by any EDC. The lack of references in the Summary or even cross-references to particular sections in the main report makes it difficult for any reader to find the basis for many of these
general statements or to gauge their degree of scientific support. Efforts to simplify the information for decision makers or lay people have resulted in a failure to appropriately characterize or present the existing data gaps and uncertainties. Consequently, the Summary overstated the strength of its conclusions. For example, the Summary stated that EDC exposures are linked with a variety of neurobehavioral diseases or disorders, of which dyslexia is included, but in the main report, the only data provided on dyslexia relate to trends in prevalence and no information is given on an association, let alone causation, with any EDC. Additionally, the Summary approached the state of the science from the same perspective as the main report, and consequently, has many of the same limitations related to: the lack of a framework for evaluating data and conducting a weight of evidence review, the use of inference to imply causal relationships between chemicals and diseases, the reliance on disease trends to suggest associations with EDCs, and ignoring the role of dose and potency in endocrine disruption.

2.2. State of the science?

How the WHO-UNEP 2012 report is to be considered an assessment of the state of the science is unclear to the reader. The report never defines what might be meant by “state of the science” nor discusses what such an assessment should cover and characterize. A state-of-the-science review 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. Several factors need to be considered in the process of integrating and interpreting data, particularly when evaluating the relevance of experimental animal studies to human health or wildlife. To that end, one of the key concerns is how dose–responses observed in experimental animal studies compare to the exposures potentially experienced by humans or wildlife. Other factors that also should be considered in the integration of data include: the quality of the available data, the consistency of the results, the presence of bias and confounding factors that may influence the findings, and the identification of data gaps. When interpreting data, the complete spectrum of findings should be considered and any controversy or debate on the issues at hand should be presented. As discussed in greater detail below, given the undefined scope and lack of a structured methodology for integrating and assessing the weight of evidence in the WHO-UNEP 2012 report, the state of the science was not evaluated in a consistent, objective manner. In some cases, the weaknesses of the data are acknowledged, such as the lack of data on EDC exposure and ovarian cancer; in other cases, important studies are ignored that counter those cited in the report.

The literature published on the potential endocrine activity of specific chemicals is extensive and beyond the scope of either the 2002 or 2012 reports. Neither report could be expected to undertake complete reviews for even a small subset of chemicals, but a systematic methodology would have ensured that a representative spectrum of the available literature was captured in the review. Publications selected for discussion in the report were not systematically assessed for their quality; nor were the results placed in a broader perspective. The only description of the literature retrieval process indicates that emphasis was placed on literature published after 2000 through March 2012. Despite the claimed focus on literature available in the last ten years or so, a substantial number of citations relied upon in the 2012 report are ones that were previously cited in the 2002 report or were published prior to the turn of the century. The conclusion that the WHO-UNEP 2012 report is not an actual update to WHO-IPCS 2002 report, but rather, a reworking of that earlier report is further supported by the 2012 report’s strong reliance on older citations. This is discussed more fully in Section 2.3 of this paper.

A particular concern regarding the data collection process is that there seems to be a strong preference toward citing studies that report an association with exposure and omitting those studies that do not support such associations. For example, studies that showed some association between levels of pesticides in the environment and adverse effects on frogs are cited (e.g., Hayes et al., 2003; McDaniel et al., 2008), while studies to the contrary (e.g., Du Preez et al., 2009; Kloas et al., 2009; Murphy et al., 2006; Skelly et al., 2010; Smith et al., 2005; Spolyarich et al., 2011) are not mentioned. The same is true of the experimental studies cited for thyroid effects from polychlorinated biphenyls (PCBs). While twelve studies were cited in the discussion, half of these studies were from one research laboratory and thus do not represent independent verification of findings in separate laboratories. Other studies investigating PCB effects on the thyroid were not included in the review (e.g., Kato et al., 2004, 2010; Martin and Klaassen, 2010). Another example
of the selective citation of literature in the 2012 report is the failure to cite many of the experimental animal studies on bisphenol A (BPA) that were conducted with larger numbers of animals and dose groups under Good Laboratory Practices (GLP) (e.g., Stump et al., 2010; Tyl et al., 2002, 2008). As recommended by Conrad and Becker (2011a,b), all well conducted laboratory studies, both GLP and non-GLP, should be considered in a review in order to provide a comprehensive understanding of the mode of action (MOA), hazards, and risks of a chemical. It is not expected that all of the available data could be critically reviewed in the state-of-the-science report, but greater objectivity and balance would have helped the reader better understand the controversies and elucidate the key uncertainties that exist in the field of endocrine disruption.

A discussion of divergent results in the literature is essential to determine why different results were generated, particularly since one of the hallmarks of science is reproducibility. If only one of several laboratories is able to generate a particular outcome, it suggests that the outcome is not very robust or that there may be something unique in how the results were obtained and thus they may not be reliable. An additional reason for discussing divergent results is to make the reader aware of controversy that is fundamental to the scientific process, rather than give a false sense of agreement. Only an informed individual can make decisions about the impact of data discrepancies in the use of these data. Finally, by identifying studies with divergent findings, important gaps in understanding are characterized and potential research priorities and behavior scores in female children – but no assessment of phthalates was conducted in this earlier study. Finally, Figure 2.31 (WHO-UNEP, 2012a, p. 182) presents data on concentrations of DDE (dichlorodiphenyldichloroethylene, a metabolite of dichlorodiphenyltrichloroethane [DDT]) in osprey eggs, with a straight line drawn through the points to depict a clear decline over time. This figure is reported to be “based on data from Henny et al. (2010).” The original paper, however, does not draw a best-fit line through these values, but rather, the data are presented in a bar graph with no statistically significant differences for the time periods: 1981–82, 1993, 1998, and 2000–01. In other words, there is no real trend for these four periods; only the values in 2006 and 2008 show a statistically significant decrease in DDE concentrations in osprey eggs. These are just a few examples of the discrepancies found in the WHO-UNEP 2012 report; they do not reflect all of the citations that have been mischaracterized or mistakenly referenced. In some cases, the discrepancies may only be mistakes in referencing the correct citation, but in others, there appears to be a tendency to exaggerate the findings or conclusions of the original authors.

Finally, a process for assessing the quality or reliability of the studies considered for review in the WHO-UNEP 2012 report is not described. In fact, the quality of the underlying studies does not appear to have been evaluated at all. This is of particular concern because not all studies should be given the same weight. For example, epidemiology studies that employ weak research designs (e.g., ecological and cross-sectional study designs) or include small sample sizes should not be given as much weight as studies with stronger designs (e.g., case-control and cohort studies) and larger sample sizes. Also, comparing in vitro exposures to relevant in vivo exposures is fraught with difficulties. In vitro studies can be relevant for investigating MOA and potential for endocrine activity, but cannot provide useful information on dose–response, do not take into account the disposition of a chemical in the body (its absorption, distribution, metabolism and elimination), and fail to account for homeostasis or other pathways and processes that respond to certain MOAs. Various methods exist to evaluate and weigh the quality and reliability of studies included in a review, such as the systematic approach for evaluating toxicology and ecotoxicology data described by Klimisch et al. (1997), which is also employed by the Organization for Economic Cooperation and Development (OECD) and relied upon by the OECD’s 34 member countries in the investigation of high production volume chemicals (OECD, 2005). Yet inexplicably, neither this approach nor an alternative approach for evaluating the quality of the data reviewed was applied in the WHO-UNEP 2012 review.

Most importantly, the WHO-UNEP 2012 report did not adopt a weight-of-evidence approach for the evaluation of data on endocrine disruption. In the 2002 review, an objective and transparent framework was developed for assessing the relationship between potential EDCs and health outcomes, which the 2012 report does not apply in its evaluation and interpretation of data. Further, no alternative approach was proposed or applied to evaluate the weight of evidence. However, an overall evaluation of the data must have been conducted because conclusions at the end of each section discussing health effects are titled: Strength of Evidence. No description is provided regarding how this “strength of evidence” was determined, although “best professional judgment” (WHO-UNEP, 2012a, p. 19) was said to be applied. Specific concerns about the lack of a structured approach for assessing causation in the 2012 report are further discussed in Section 2.4 of this paper.

In the WHO-UNEP 2012 report, evidence for endocrine disruption in humans and wildlife is presented as narrative reviews of the data. These assessments were described as being founded on an aggregation of the information related to biological plausibility, relevant exposures, consistency of the data across species, and dose–response and temporality (WHO-UNEP, 2012a, p. 19).
Despite this description, there appears to be little, if any, integration of the data. For example, biological plausibility is often cited in the 2012 report as the basis of concern for causality, but the evidence presented generally is limited to data on the role of endogenous hormones and not based on mechanistic data, potency or actual exposures for any of the chemicals of potential concern. Exposure data are clearly not integrated with the rest of the data, as exposure information is provided separately in Chapter 3 of the WHO-UNEP 2012 report. Species concordance is frequently used to try and bridge experimental animal or wildlife data with human data on particular observed effects, but there is a lack of integration of the data with exposure information or mechanisms. Potential species differences do not seem to have been considered, and in some cases, these differences may be critical to the interpretation of the data (an example is provided regarding thyroid effects in Section 2.6.2). Although the report purports to incorporate dose–response in its evaluation of the evidence, it tends to be ignored when specific chemicals or adverse outcomes are discussed; this issue is discussed in greater detail in Section 2.6 below.

Other factors not specifically mentioned in the WHO-UNEP 2012 report that should have been considered in the integration and interpretation of data include the reproducibility of the data and consistency of data across different lines of evidence (epidemiology, in vivo and in vitro data), data gaps, and the existence of controversy or differences in interpretation of study findings. As mentioned previously, the reproducibility of a finding in different studies, different research labs, or in different study populations was not addressed in the report. Although it is noted in a number of places where significant gaps still exist in the data, these data gaps and their implications are not always carried forward into the report conclusions. Further, no specific research recommendations are provided based on such data gaps in the 2012 report, which is an important and useful element for developing data in the future.

Other recently published reports have done a better job of providing state-of-the-science reviews on controversial issues in science. For example, the U.S. Environmental Protection Agency (USEPA) recently released a draft review on the scientific issues surrounding the phenomenon of a non-monotonic dose–response (NMDDR) (USEPA, 2013). USEPA took a methodical and even-handed approach to its review. The scope of the report was provided, clearly stating the scientific questions that were to be addressed. USEPA also clearly described and respected the limits of the report; for example, they acknowledged that it was not a comprehensive treatment. In a succinct manner, USEPA described both the supporting and conflicting evidence for NMDDRs. As part of this assessment, the uncertainties associated with the interpretation of the data were considered. Overall, the USEPA draft report on NMDDRs is a good example of a well-conducted, state-of-the-science review and illustrates how controversial issues, such as NMDDRs, can be approached in an objective manner. Similarly, the European Food Safety Authority (EFSA) reviewed a developmental neurotoxicity (DNT) study in rats and other recent scientific literature in context of the risk assessment for BPA (EFSA, 2010), a compound mentioned in the WHO-UNEP 2012 report as being an EDC. Although EFSA’s review had a narrow scope, specific parameters were provided for the inclusion of studies in the assessment. EFSA also described its reliance on clear “quality criteria” to assess the strengths and weaknesses of the studies it reviewed. Those criteria addressed issues of study design, conduct, recordkeeping, and interpretation. The EFSA review provides an example in which a comprehensive evaluation of the literature considers explanations for toxicity beyond those related to endocrine disruption; a feature which is not generally considered

---

1 When referring to adverse outcomes in this paper, this includes adverse health effects in humans, experimental animals, and wildlife.
weight-of-evidence framework developed in the 2002 report, it also does not provide an appropriate alternative framework by which the data could be assessed to determine causal relationships. Instead of using clear and objective principles as the foundation for the report, “best professional judgment” was used (WHO-UNEP, 2012a, p. 19) without fully describing any criteria by which such judgment was to be applied. Consequently, the WHO-UNEP 2012 report is not an update but, rather, a selective re-evaluation of information largely included in the 2002 report.

In order to better understand why and how the WHO-UNEP 2012 report is not an update of the WHO-IPCS 2002 report, selected examples are provided below of adverse outcomes reviewed in both documents. These examples compare the data relied upon, how the available evidence was characterized, the state of the science, and conclusions reached about the weight of evidence in the respective reports.

### 2.3.1. Sperm/semen quality

Sperm or semen quality was evaluated in both state-of-the-science reviews. The WHO-IPCS 2002 report concluded that a global trend for declining semen quality was not supported by the existing data. This conclusion was based on a broad review of studies investigating sperm counts, from the first study suggesting a decline (Nelson and Bunge, 1974), and included the first meta-analysis (Carlsen et al., 1992), longitudinal retrospective studies in single centers (e.g., Auger et al., 1995), and broader investigations around the world (e.g., Auger and Jouannet, 1997; Jørgensen et al., 2001; Swan et al., 1997; Younglai et al., 1998). The review in the 2002 report described the limitations and biases of the various studies. Concerns included the use of retrospective study designs, evaluation of semen or sperm samples from men that may not have been representative of the general population (e.g., patients at infertility clinics), differences in methods for recruiting study subjects, variability in analytical methods, and lack of control or consideration of the other factors that are known to impact sperm quality (e.g., age, sexual abstinence). The lack of data on specific chemical exposures raised questions about assessing the strength of the association. The WHO-IPCS 2002 report acknowledged that, while it was biologically plausible and some experimental evidence was available to support EDCs affecting sperm quality, the “lack of any demonstration to date of an endocrine-disrupting mechanism for other chemical exposures indicates the need for more studies before firm conclusions can be drawn” (WHO-IPCS, 2002, p. 56). In its evaluation of the strength of evidence for: Semen Quality and Testis Function in Humans (WHO-IPCS, 2002, p. 124), the 2002 report concluded that the overall strength of evidence was weak based on an assessment of temporality, strength of association, consistency, and biological plausibility.

The WHO-UNEP 2012 report described various studies of declining semen quality, also starting with the first meta-analysis by Carlsen et al. (1992) and cites many of the same references relied on in the WHO-IPCS 2002 report. The discussion is not limited to or focused on the newest studies on sperm or semen quality, nor is it a comprehensive review of all of the data. Although several prospective studies of the general population were mentioned as being conducted by Nordic, Baltic, German, Spanish and Japanese researchers, the only citations provided are two studies of Finnish and Danish men (Jørgensen et al., 2011, 2012). A retrospective analysis of French men with total infertile partners that were subjects in assisted reproductive technology was also cited (Rolland et al., 2013). However, several other studies published since the WHO-IPCS 2002 report that do not show a decrease or an increase in sperm counts are not cited (e.g., Axelson et al., 2011; Costello et al., 2002; Elia et al., 2012; Marimuthu et al., 2003; Pal et al., 2006). In a recent commentary from Bonde et al. (2011), it is shown that sperm cell counts in Danish military draftees have remained stable and, in fact, suggest higher counts in the last four years of the study between 2007 and 2010. This contribution to the extensive and ongoing discussion about sperm cell count trends is not cited at all. The selective citation of literature and the failure to include many studies that do not support a decline in sperm counts indicates that the 2012 report provided an unbalanced review of the literature.

Unlike the WHO-IPCS, 2002 report, there is no mention of the limitations and potential biases of any of the studies reviewed, despite the fact that all of the limitations noted in the 2002 report continue to apply. The new, prospective studies highlighted in the 2012 report also have low study participation rates: 13.4% in the Finnish study (Jørgensen et al., 2011) and 24% in the Danish study (Jørgensen et al., 2012). Such low participation rates raise concerns about whether the men in these studies were truly representative of the general population. Furthermore, there is only a brief mention that the issue of declining sperm counts remains controversial. Questions about the implications of conflicting findings related to sperm quality are not carried over into the conclusions on male reproductive health. Recent reviews of this issue continue to characterize the reports of declining sperm counts as controversial because of the differences seen geographically and temporally (Fisch and Braun, 2013; Sharpe, 2010). These reviews identify a number of factors that may account for these differences that are not related to endocrine disruption, such as lab techniques, sexual behavior resulting in differences in abstinence, lifestyle factors (e.g., obesity, drug use), and genetic variations.

The WHO-UNEP 2012 report presents limited data regarding specific EDCs and the potential for affecting sperm quality. Epidemiology studies are characterized as showing weak associations with EDCs, which is based on single citations for most of the chemicals of possible concern identified in the report. Single publications are not sufficient to assess the weight of evidence (or to provide one’s “professional judgment”) of a potential association between an exposure and an adverse effect. Furthermore, the 2012 report does not evaluate the limitations of these studies; nor does it consider experimental animal studies in conjunction with epidemiology to assess the hypothesis that EDCs could cause male reproductive disorders, including effects on sperm or semen quality. The 2012 report relies only on “suboptimal or poor semen quality in large proportions (20–40%) of men in countries in which this has been studied” and that there is “some evidence for a declining semen quality,” thereby suggesting that these perceived trends are the consequence of exposure to endocrine disruptors with no strong evidence to support this claim.

Despite acknowledging that the epidemiology data only show weak associations for a decline in sperm quality related to specific EDCs, the conclusions in the 2012 report for male reproductive health focus on the observation of decreased sperm counts, ignoring the variability in sperm quality reported around the world and the well-founded scientific questions that have been raised about this issue. Based on the evidence presented in the WHO-UNEP 2012 report, it does not appear that the evidence for changes in sperm quality differ from that reported in the WHO-IPCS 2002 report. Therefore, the reason for the discrepancy between the conclusions of the two reports is difficult to explain. 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.

### 2.3.2. Adrenal disorders

As noted in the WHO-UNEP 2012 report, adrenal dysfunction was not discussed in detail in the WHO-IPCS 2002 report because the “available research to date [was] very limited” (WHO-IPCS, 2002, p. 86). The observation of severe adrenocortical hyperplasia in the Baltic ringed and gray seal were mentioned and it was stated in the 2002 report that although adrenal effects in wildlife were...
associated with dichlorodiphenyldichloroethane (DDD), DDT, and PCBs, the involvement of these compounds in the cause of these disorders was uncertain.

The WHO-UNEP 2012 report contains a specific section discussing adrenal disorders, which seems appropriate given the lack of systematic review in 2002. The information presented in the WHO-UNEP 2012 report focuses heavily on data and literature from 2002 or earlier, particularly regarding effects in wildlife. This suggests that the state of knowledge regarding endocrine disruption related to perturbations of the adrenal gland has not changed significantly since 2002. No new data are provided on the observation of adrenocortical hyperplasia in Baltic seals – the one new article, Lind et al. (2003), addresses bone mineral density in Baltic grey seals, not adrenocortical hyperplasia. Significantly, there is a change in the conclusions about the weight of evidence regarding adrenocortical hyperplasia in wildlife and various compounds, despite the lack of new information. In 2002, this was considered an uncertain association, but in 2012, it was concluded that this was a causal relationship. This change in the interpretation of the data should have been highlighted and the evidence to support this change should have been stated explicitly. Additional concerns about the weight of evidence regarding the causal relationship between exposure to EDCs and the observation of adrenocortical hyperplasia in Baltic seals are discussed later in this paper (see Section 2.4.1). It is unclear how a more definitive conclusion could be reached regarding adrenocortical hyperplasia based on essentially the same data reviewed in 2002, which suggests that the WHO-UNEP 2012 report is not an update, but a re-interpretation, of the same data. Alternatively, the authors of the WHO-UNEP 2012 report may view the 2002 report as flawed. If this is the case, then the reasons behind the divergent opinions should have been explicitly stated.

2.3.3. Endometriosis

In the WHO-IPCS 2002 report, endometriosis was reviewed as a reproductive outcome possibly associated with EDCs. Some epidemiology studies of PCBs and dioxins indicating associations with endometriosis were contrasted with other studies that failed to observe an association. Experimental animal evidence was presented for 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) and the data were judged to be conflicting – some studies indicated that TCDD may have a role in the development of endometriosis, but other studies did not support such an association. The 2002 report mentioned the criticisms of the studies of TCDD in rhesus monkeys, including factors potentially confounding the results. The question of relevance of the rodent studies for humans was also raised given the high doses required to induce endometriosis in rodents. Regarding possible MOAs, the roles of estrogen and progesterone were discussed in the context of known disease development and the aryl hydrocarbon receptor (AhR) was reviewed in relation to TCDD exposure and endometriosis. The 2002 report concluded:

“Relative to the hypothesis of an association between a stressor and an outcome, evidence is judged to be weak because of conflicting data from humans and animals, lack of association in women exposed to high amounts of TCDD, and antiestrogenic effects of TCDD. In humans, occurrence of endometriosis shows dependency on estrogen–progesterone balance, suggesting that an EDC-related mechanism may be possible.” [WHO-IPCS, 2002, p. 125]

The WHO-UNEP 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. A majority of the literature cited in the WHO-UNEP 2012 report linking TCDD to endometriosis was cited in the WHO-IPCS 2002 report, but the 2012 report failed to note that many of these studies (e.g., Bruner-Tran et al., 1999; Rier et al., 2001) employed high doses of TCDD, which were considered of questionable relevance to endometriosis in humans in the 2002 report. Much of the PCB literature was also reviewed and cited in the earlier WHO-IPCS 2002 report. Although the WHO-UNEP 2012 report cites nine publications as finding a “relationship between circulating phthalate (and phthalate esters) and endometriosis” (p. 44), none of these publications mention phthalates. Four other publications are mentioned regarding studies on phthalate ester metabolites, but the findings are inconsistent for various phthalate esters, which is acknowledged in the WHO-UNEP 2012 report. It is interesting to note that, in the discussion of phthalate esters and endometriosis, there is no mention of a study by Itoh et al. (2009) that did not observe an association between measures of phthalate metabolites in urine and endometriosis. Regarding a potential MOA for endometriosis, general hormonal influences on the disease are discussed, but little information is reviewed on specific EDCs. The WHO-UNEP 2012 report stated that epigenetic changes have been reported to be involved in endometriosis – particularly those induced by in utero exposure. However, only two review articles are cited (Cakmak and Taylor, 2010; Guo, 2009), neither of which address causal relationships for specific chemicals. 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 2002 report.

The examples above illustrate how the WHO-UNEP 2012 report is not an update of the WHO-IPCS 2002 report. In some cases, the 2012 report reaches conclusions that conflict with those reached in the earlier report, despite the lack of new information to support a change in the weight of evidence. In other cases, the 2012 report simply presents a re-evaluation of the studies reviewed in 2002. The fact that the 2012 report often reaches more definitive conclusions based on the same data emphasizes the use of non-transparent, subjective decision making for evaluating potential causal relationships compared to the earlier 2002 report. In light of this, although the 2012 report is stated to be an update, in actuality, the 2012 report is a revised review of the state of the science presented in 2002 and does not build upon what was previously done.

2.4. Inference, not causation

Causation is a critical element in the definition of an endocrine disruptor and consideration of causal relationships should have been given more attention in the WHO-UNEP 2012 report. Instead, the 2012 report tended to focus on only part of the definition of an endocrine disruptor – the potential to alter some aspect of endocrine function. Consequently, information is presented or discussed on possible or potential endocrine disruption without considering whether alterations to the endocrine system caused any adverse effects. The failure to differentiate between the potential for endocrine system interaction of some sort and actual disruption of control of physiology or development as a result of such interactions misleads the reader about the weight of evidence for particular disruptive effects. This could result in inappropriate regulatory actions and research priorities when additional research still might be needed to establish these causal relationships. Any state-of-the-science review should be a balanced and objective review of all of the available literature with identification of data gaps along with clear statements of the conclusions supported by the data and the inferences that cannot be supported. Only through such an approach can all parties, in particular, stakeholders without a strong scientific background, be confident in the evidence-based decisions that arise from such a report.

There have been many definitions for an endocrine disruptor over the years and the WHO-UNEP 2012 report stated it relies on the definition in general use today; this same definition was used in the WHO-IPCS 2002 report. Integral to this definition are three important components:

- The substance must act through an endocrine MOA that alters function of the endocrine system;
- The substance must cause an adverse health effect; and
• That adverse effect must be causally related to and occur as a consequence of the altered endocrine function.

All three of these components are necessary to demonstrate that a chemical is an endocrine disruptor. This requires the differentiation between endocrine-mediated effects from other known MOAs and the linking of an observation of an adverse effect to that endocrine MOA. Various chemicals or classes of chemicals were presented in the WHO-UNEP 2012 report as though they were EDCs, but the report does not provide any information to show that the effects of these chemicals are, in fact, the result of alterations in endocrine function. In the 2012 report, the interference with endocrine function by a chemical is often considered sufficient evidence for endocrine disruption, when in fact this only demonstrates that a chemical has the ability to interact with the endocrine system (Tinwell et al., 2013) and the potential for endocrine disruption.

The WHO-UNEP 2012 report frequently relies solely on the observation of adverse effects in endocrine organs or the existence of possible endocrine MOAs. The report often reaches conclusions based on only one or two elements of the definition for an endocrine disruptor. It is the third and essential component of the definition that leads to the determination of endocrine disruption. In order to reach the conclusion that a chemical is an endocrine disruptor, one needs to have a systematic method for assessing causation. Various methods have been developed to assess causal relationships; the most commonly referenced approach relates to the criteria outlined by Sir Austin Bradford Hill (Hill 1965). The WHO-IPCS 2002 framework was proposed based on Bradford Hill’s criteria to assess relationships between exposures to potential EDCs and altered health outcomes (see WHO-IPCS, 2002, Chapter 7). Several factors were specifically considered important in establishing the overall weight of evidence for a causal relationship: temporality, strength of the association, biological gradient (dose response), consistency of the observations, biological plausibility, and evidence of recovery. A number of illustrative examples were presented to demonstrate how the framework would work to evaluate hypotheses that particular EDCs cause specific adverse outcomes.

The WHO-UNEP 2012 report did not adopt the WHO-IPCS 2002 framework for assessing causation. In fact, the authors criticize the earlier approach because it failed to distinguish between the quality of evidence and strength of the recommendations, as recommended in the Grades of Recommendation Assessment, Development and Evaluation (GRADE) scheme used in clinical medicine (GRADE, 2011). Despite the reference to this alternative approach, the 2012 report did not apply the GRADE scheme (which, because it was derived for clinical medicine, would not be appropriate in any case for the evaluation of toxicology data and studies in wildlife). Although there is substantial discussion of weight of evidence, no systematic approach is described or adopted to assess the weight of evidence for causation; instead, “best professional judgment” is the basis for making expert assessments of the data and aggregated data are purported to be presented on trends, biological plausibility, relevant exposures, consistency across species, dose–response, and temporality (WHO-UNEP, 2012a, p. 19). However, not all of these categories are captured in the discussion of all endpoints of concern; in particular, few data are presented on dose–response. Although a section, titled Strength of evidence, is included in the discussion of health effects and conclusion-like statements are provided for some adverse health outcomes, in many cases, these conclusions do not reflect the uncertainties or limitations that were described in the main text. More importantly, these statements do not represent the totality of the evidence, often ignoring contrary study results that were inappropriately excluded. Furthermore, little effort was made to synthesize the findings across the categories in the report – from temporal trends to exposure and biological plausibility – in order to provide a complete picture of the state of science for MOA, dose response, and adverse health effects. The use of the narrative approach in the WHO-UNEP 2012 report allowed for a selective presentation of information without a critical review of the data. In the absence of a formal assessment of causation, subjective inference is relied onto suggest causation.

For each of the specific health or environmental adverse outcomes discussed, information is presented in the 2012 report for a sequence of topics in such a way to suggest that they were related when, often, there is no connection at all. Each subchapter discusses the trends in the subject disease or health outcome first, followed by a suggestion that exposure to environmental chemicals contributes to these trends. In some cases, limited or no data are provided to support environmental chemical exposures as contributing to the trends and often, other causes of the trends are ignored or dismissed. For example, the report stated that there is a rising trend for breast cancer and notes that this trend cannot be explained by improved diagnosis or changes in risk factors, including genetic factors (WHO-UNEP, 2012a, p. 126). This statement about the trends is followed by the comment that twin studies have highlighted the importance of environmental factors. The juxtaposition of these two sentences gives the reader the impression that the rising trend in breast cancer must be a consequence of these environmental chemical exposures. However, the term “environment” as used in twin studies encompasses all modifiable (i.e., non-genetic) factors, not just environmental chemical exposures. Several publications have concluded that the observed increase in breast cancer incidence in some countries can be explained by the introduction and promotion of mammography and breast cancer screening (Glass et al., 2007; Séradour et al., 2009; Weeden-Fekjær et al., 2012). In addition, in several countries around the world, the trend for breast cancer has been in decline since 2002, which has been attributed to the dramatic reduction in hormone replacement therapy in post-menopausal women (Glass et al., 2007; IOM, 2012; LeClère et al., 2013; Séradour et al., 2009; Weeden-Fekjær et al., 2012). Therefore, the implication that the trends are due to environmental chemical exposures is misleading when evidence for some of the changes in breast cancer trends point to improved diagnostic tests or to changes in recommended post-menopausal therapies. In the IOM (2012) review of the state of the science regarding environmental risk factors for breast cancer, the factors with the clearest evidence included: hormone therapy products, oral contraceptives, being overweight or obese, alcohol consumption and ionizing radiation; the evidence for exposure to industrial chemicals was considered to be limited and in some cases only suggest a possible association for an increased risk of breast cancer. Further discussion of the issues associated with disease trends as presented in the WHO-UNEP 2012 report is described in more detail in Section 2.5.

Another part of the discussion in the 2012 report for each adverse outcome included a description of the normal role of endogenous hormones. This information is useful in understanding normal physiology, basic mechanisms and showing the potential biological plausibility for endocrine disruption; however, it does not demonstrate that environmental chemicals are acting in the same manner as endogenous hormones. Evidence is presented from epidemiology and experimental animal studies for particular chemicals associated with various health outcomes, but rarely is there any discussion of specific mechanisms for the highlighted chemicals. For example, androgens and estrogens are mentioned as playing a role in normal prostate development; however, an androgen- or estrogen-mediated MOA cannot be described for any chemical mentioned in this section. In fact, the report notes that “the precise mechanisms by which the chemicals related to prostate cancer induce the carcinogenic process remain to be
resolved” (p. 131). Overall, the reader is left with the impression that environmental chemicals could cause disease via the same mechanisms as endogenous hormones without any understanding of the normal feedback mechanisms that exist for homeostasis as well as differences in potency and dose–response (discussed in more detail in Section 2.6). Moreover, a failure to recognize the complexity of hormone-receptor interaction and activation is lacking. Specifically, despite a short discussion at the beginning of the WHO-UNEP 2012 report, the roles of co-activators, repressors, and transcription factor interactions, and receptor cross-talk are completely ignored throughout the rest of the report. Recent studies also point to competition for transcription factors (Kollara and Brown, 2006); yet these issues are completely overlooked.

Another shortcoming of the WHO-UNEP 2012 report that affects the cohesive evaluation of causation is the fact that little, if any, discussion of exposure occurred in the report in the context of specific effects or related to specific hormonal MOAs. Generally, when exposures were mentioned, a reference is made to Chapter 3 of the 2012 report. In Chapter 3, only general information is presented on major classes of chemicals that the WHO-UNEP 2012 report described as known or potential endocrine disruptors. No quantitative information on exposure is presented for individual chemicals, and potential human exposures are not considered in context of dose–response data. The segregation of exposure data from the information on chemicals of concern, potential adverse effects associated with that chemical, and biological plausibility or possible MOAs for endocrine disruption make the assessment of the causal relationship between exposure and effects impossible.

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. The selective citation of literature without discussion of contradictory studies and the failure to consider alternative causes of reported effects gives the reader the impression that the weight of evidence is stronger than is justified by the available scientific data. This calls into question the integrity of decisions at all levels of the report. Specific examples are provided below that further demonstrate the issues with the evaluation of causation in the WHO-UNEP 2012 report, with an emphasis on highlighting the key factors that are typically used in a causation assessment.

2.4.1. Adrenocortical hyperplasia in seals

The WHO-UNEP 2012 report concluded that there was sufficient evidence to demonstrate that a mixture of PCBs and DDT caused adrenocortical hyperplasia and a Cushing-like condition in Baltic seals. The report attributes recoveries in seal populations to the “drastic reduction of DDT and PCBs in Baltic biota” (p. 149). This conclusion is not supported by the discussion in the main text of the report; at best, it is based on limited data. Additionally, the report ignores conflicting data and fails to consider alternative causes for the adrenocortical hyperplasia.

The data on adrenocortical hyperplasia in seals or other aquatic mammals are limited and inconsistent. Although the WHO-UNEP 2012 report mentions that adrenocortical hyperplasia has not been reported in seal populations outside the Baltic Sea, it does not discuss evidence from Great Britain that reported contrasting results for other marine mammals (Kuiken et al., 1993); nor does the report address alternative causes for these observations. For example, Kuiken et al. (1993) measured the concentrations of chlorinated hydrocarbons in the carcasses of harbor porpoises and found that adrenocortical hyperplasia was not associated with increased levels of these chemicals, but rather associated with chronic stressors causing their death (e.g., malnutrition, prolonged illness). Another study that is not mentioned, Clark et al. (2006), also found a significantly higher adrenal gland mass in Atlantic bottlenose dolphins that were chronically stressed compared to those that were acutely stressed, suggesting that other factors may be involved in the observation of adrenocortical hyperplasia in seals. Lair et al. (1997) suggested that the adrenal hyperplasia seen in beluga whales may be part of the normal aging process. Although these latter studies are on cetaceans, rather than pinnipeds, they point to alternative causes of adrenocortical hyperplasia in marine mammals and should have been considered in the data review and assessment of causation.

The report also fails to address data inconsistencies for the specific persistent organic pollutants (POPs) mentioned, including the differences observed between experimental animal studies and reports in wildlife. For example, in two-year chronic toxicity studies of PCBs conducted in female Sprague–Dawley rats (NTP, 2006, 2010), increased adrenocortical atrophy was reported, which is in contrast to the findings of adrenocortical hyperplasia found in Baltic seals. Cancer bioassays for several commercial PCB mixtures (i.e., Aroclors) do not report adrenal effects in rats exposed to these PCBs for two years in the diet (Mayes et al., 1998), which further calls into question the identification of PCBs as the cause of adrenocortical hyperplasia in seals. While the degree of similarity between adrenal glands of seals and rats is unknown, it is expected that similar, rather than opposite, effects from the same chemical are more likely in mammals. The WHO-UNEP 2012 report acknowledges in the introduction to the section on adrenocortical hyperplasia that a DDT metabolite showed degeneration and necrosis in the adrenal cortex of laboratory mice—a finding inconsistent with hyperplasia—but this is not mentioned further.

No information on plasma cortisol levels is available for the Baltic Sea seals, as the WHO-UNEP 2012 report notes, which are critical for differentiating a stress response from a direct toxic effect on the hypothalamus–pituitary–adrenal axis as the cause of the observed hyperplasia (Harvey and Sutcliffe, 2010). Although the WHO-UNEP 2012 report recognizes the possible role of stress and aging of wildlife in the development of adrenal hyperplasia in the main text, when the strength of evidence is described, these factors are ignored. The 2012 report does not consider alternative causes for the observed species recovery, and stress can plausibly explain these observations given other changes occurring during the same time period in the Baltic Sea (e.g., reductions in nutrient inputs, eutrophication, oxygen deficiency, and oil discharges) (HELCOM, 2012).

It is presumed that because the effect occurred in an endocrine organ, this must be the result of endocrine disruption, but no data are provided to show that these effects are the result of an endocrine MOA. Given the limited data available on the observation of adrenocortical hyperplasia, inconsistent findings in experimental animal studies, conflicting data in other wildlife species and other regions, and stress as a plausible alternative cause for these observations, it is questionable that there is sufficient evidence to demonstrate that these compounds caused the adrenocortical hyperplasia observed in the Baltic seals.

2.4.2. Prostate cancer

Prostate cancer is included in the discussion of various hormonal cancers in the WHO-UNEP 2012 report and it is stated that there is sufficient evidence for a link between pesticide exposures and prostate cancer (WHO-UNEP, 2012a, p. 130). However, the report does not reach any conclusions or make any statements about the evidence for an endocrine-mediated MOA. The mere association between pesticides and prostate cancer is insufficient to demonstrate causation. More importantly, this link has not been shown to be attributable to an alteration in endocrine function. Moreover, based on the data presented in main text of the report, from an objective view, there is not sufficient evidence to conclude that a link between pesticide exposures in general and prostate cancer
even exists. The report mentions that individual pesticides have been reported to be associated with prostate cancer; however, these data are not consistent. For example, the report mentions that oxychlordane was linked with an increased risk for prostate cancer based on Ritchie et al. (2003). Another study cited in the report (Hardell et al., 2006) that did not observe an association between oxychlordane and prostate cancer is ignored. Other biomonitoring studies that examined oxychlordane levels are not cited in the 2012 report, including two studies that failed to observe an association with prostate cancer (Aronson et al., 2010; Sawada et al., 2010) and one that did (Xu et al., 2010). Thus, the evidence from biomonitoring studies for an association between oxychlordane and prostate cancer is inconsistent. In these same studies, similar results were seen for other organochlorine pesticides where only one or two statistically significant associations were reported for any individual pesticide. Although other epidemiology studies are cited in the WHO-UNEP 2012 report, these studies did not directly measure exposure through analysis of blood, fat, or urine, and therefore, are considered to be more susceptible to bias and should be given less weight in an overall assessment of the evidence. For example, the Agricultural Health Study (AHS) (Alavanja et al., 2003; Koutros et al., 2010) obtained exposure information based on a questionnaire that collected data on duration and frequency of pesticide use, which is a less reliable measure of exposure compared to biomonitoring studies.

It is interesting to note that the two meta-analyses of pesticide applicators (van Maele-Fabry and Willems, 2004) and pesticide manufacturers (van Maele-Fabry et al., 2006) characterized the weight of evidence for pesticide exposure for these workers as weak (rate ratios of less than two). A general limitation with the epidemiology studies on pesticide exposure and cancer is the use of multiple comparisons to assess risks for many different types of pesticides and various cancer endpoints, which increases the likelihood that a statistically significant association will be detected based on chance alone. In addition, the long latency between initiation and detection of the cancer make it very difficult to identify relevant exposures. Overall, the data for pesticides as a broad category are weak and those for individual pesticides are limited with regard to an association with prostate cancer, let alone to causation.

While it has been speculated that hormones play a role in the development of prostate cancer given the involvement of sex steroids in the development of the prostate, it is recognized that many other factors may be involved in the etiology of prostate cancer. These factors include age, family history (genetics), race, dietary fat, and other dietary factors (NIH, 2013). As the WHO-UNEP 2012 report notes, the mechanism by which pesticides could induce prostate cancer is currently unknown. Therefore, while it may be biologically plausible for endocrine disruption to be contributing to prostate cancer, insufficient data are available to show that pesticides are involved in the induction of prostate cancer by an endocrine-mediated MOA.

Based on the issues discussed above regarding the strength of the association, lack of consistency among studies of pesticide workers and manufacturers and lack of evidence for an endocrine MOA, the overall weight of evidence for pesticides causing prostate cancer through endocrine disruption should have been determined to be either inconclusive or weak, not “sufficient” as determined in the WHO-UNEP 2012 report.

As the above examples illustrate, the WHO-UNEP 2012 report presented information on chemicals and various adverse outcomes, but whether these exposures actually cause these effects was not established scientifically. Several factors for establishing causation, such as demonstrating exposure to the chemical, dose–response, and consistency in the data, were frequently ignored. 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 WHO-UNEP 2012 review.

2.5. What do temporal trends in diseases show?

The WHO-UNEP 2012 report indicates that the high incidence and increasing trends of many endocrine-related disorders in humans is one of “three strands of evidence [that] fuel concerns over endocrine disruptors” (p. vii). The report indicates that “worldwide, there has been a failure to adequately address the underlying environmental causes of trends in endocrine diseases and disorders” (WHO-UNEP, 2012a, p. ix).

It should be noted that, in some cases, the WHO-UNEP 2012 report does put trends in perspective. For example, the report indicates that elevated BPA levels in the body could be a result of polycystic ovarian syndrome and not the other way around. In other cases, however, alternative explanations are either not discussed, or when they are discussed, they are dismissed in favor of endocrine disruption as an explanation without a sufficient evaluation of the science. For example, as noted earlier, the report briefly acknowledges, but appears to dismiss, the role of alternative factors in reported trends for breast cancer or sperm/semen quality.

The 2012 report concluded that because the increase in disease trends has occurred primarily over the last few decades, these trends cannot be entirely attributable to genetic causes. It is implied that if these trends are not the result of genetic heritability, then the only other explanation is environmental exposure to chemicals. However, environmental factors go well beyond chemicals, and cover a multitude of characteristics in human populations including: diet, exercise, lifestyle factors, infectious agents, and even drug use; for wildlife, these include factors related to habitat, food supply, disease, predation, and competition – factors which can be completely unrelated to environmental chemical exposures. The WHO-UNEP 2012 report does not acknowledge that, much of the time, the environmental causes of the diseases being discussed are not chemical exposures.

There are many factors that can influence the appearance of an increasing trend (either temporally or geographically) in disease incidence or prevalence. For human health considerations, these include changes in diagnostic criteria, screening, medical interventions, and treatment. Other life style trends are important; for example, giving birth at an older age can have a substantial impact on the incidence of birth defects and congenital abnormalities. Another significant trend is the obesity epidemic and being overweight has been found to increase the risk of male infertility (Hammoud et al., 2008). In addition, the report sometimes described trends based on a compilation of data from different sources such that, what appears to be a trend actually may be a reflection of different data collection methods.

Through selective citation in the WHO-UNEP 2012 report, an impression is created that certain diseases have an increasing incidence or prevalence. However, publications that are not cited in the WHO-UNEP 2012 report often provide opposing evidence that the occurrence of the disease of interest is not on the rise. For example, in the 2012 report, several papers are cited to indicate that the prevalence of hypospadias is increasing. Fisch et al. (2010) was not cited, who stated that “[a] review of the epidemiologic data on this issue amassed to date clearly demonstrates that the bulk of evidence refutes claims for an increase in hypospadias rates.” The postulated decreasing trends in semen quality have also been highly contested by other scientists, as already discussed above.
In the end, even if an environmental exposure and a health outcome trend are related spatially or temporally in an ecological epidemiology study, all one can be sure of is that a statistical correlation exists. Such study designs cannot determine whether people with health effects are the same people with a particular exposure in these studies. In addition, it generally cannot be known if the exposure occurred before the health outcome (or vice versa), whether they each have a common cause, or whether they are completely independent. As a result, these analyses are the weakest form of scientific evidence for evaluating causation.

The reasoning presented in the WHO-UNEP report that the rising disease trends must be associated with exposure to EDCs becomes even more questionable given that exposure to most of the compounds named in the report have not increased over the last two or thirty years, but rather, have decreased. Concentrations of DDE in human milk in Germany have been reported to be reduced by approximately 90% from 1984 to 2001 (Wilhelm et al., 2007). Another biomonitoring study in Germany clearly shows that human exposure to phthalates has declined over time, with the exception of some new compounds that were recently introduced to the market (Wittassek et al., 2007). Historical biomonitoring data from a number of countries including the Czech Republic, Norway, and the U.S. indicates that human exposure to compounds like persistent chlorinated pollutants has decreased in the last two decades (Cerná et al., 2012; Ferribi et al., 2007; Nöst et al., 2013). These trends in declining human exposures over the last two decades in the Western world contradict the hypothesis presented in the WHO-UNEP 2012 report that a continuing rise in disease trends, if any, can be related to EDC exposure in humans.

Overall, any claims of trends indicating an endocrine cause must be supported by a systematic review of all relevant data regarding disease trends, exposure trends, and alternative explanations for observed statistical correlations. Below, two examples are discussed in which the WHO-UNEP 2012 report described trends of increasing endocrine-related disorders and concluded they are due to environmental EDCs, without considering whether the weight of evidence supports such conclusions or alternative explanations are more likely. Note that these examples are not weight-of-evidence analyses, but the identification of factors and limitations related to the discussion of health trends in the WHO-UNEP 2012 report and how these factors may lead to erroneous conclusions.

2.5.1. Autism spectrum disorders (ASDs)

Autism spectrum disorders (ASDs) – which include autistic disorder, Asperger’s syndrome, and pervasive developmental disorders not otherwise specified – are developmental disabilities that are diagnosed based on behavioral symptoms and failure to reach certain developmental milestones (CDC, 2012). ASDs are characterized by communication and socialization problems, as well as atypical behaviors and interests. ASD symptoms are usually apparent before the age of three and can vary in severity and presentation. While IQ decrement can co-occur with ASDs, they are not associated with ASDs per se (CDC, 2012).

The WHO-UNEP 2012 report claims that “the increase in autism spectrum disorders is indisputable” (p. 109) and there is “sufficient evidence to conclude that a number of factors, including environmental, contribute to the increases in autism spectrum disorders” (p. 119). The first claim is based on two studies, one published in 1976 and the other in 2007 (Rice, 2007; Wing et al., 1976). More current analyses were not cited; there is a considerable body of other literature evaluating ASD prevalence that the WHO-UNEP 2012 report did not consider.

At least some of the increase in ASD prevalence is due to changes in diagnostic criteria, better diagnostic techniques and increased case ascertainment (CDC, 2012). For example, when the Centers for Disease Control and Prevention (CDC) created new diagnostic standards for its Autism and Developmental Disabilities Monitoring Network (ADDM), there was a significant increase in ASD prevalence compared to the prevalence estimated using older standards (Rice et al., 2012). Also, a recent reanalysis of older studies found ASD prevalence to be consistent with current reports when the data were analyzed according to contemporary diagnostic criteria, indicating that ASD was likely underestimated in earlier studies (Duchan and Patel, 2012). Increases in ASD prevalence may also be partly attributed to diagnostic substitution, as children who would have been diagnosed with learning disabilities or mental retardation in the past are currently diagnosed with ASD (reviewed by Fombonne et al., 2009). ASD diagnosis relies on behavioral identification, which leaves room for wide variation in clinical judgment and is influenced by differing cultural and social norms worldwide (Elshabagh et al., 2012). Even within a culture, there is evidence of low inter-evaluator agreement about diagnoses, although it is not clear in which direction it would influence prevalence measures (Duchan and Patel, 2012). Finally, the success of national awareness efforts and the growth of alternative treatments may also contribute to a perceived increase in ASD prevalence, with more children being tested, diagnosed, and treated (Duchan and Patel, 2012). Therefore, it is not clear that there is a true increase in ASDs.

Regardless of temporal trends, the WHO-UNEP 2012 report presents no evidence that environmental factors, much less EDCs, contribute to ASDs. The report summary stated that “insufficiency of thyroxine during pregnancy is also associated with reduced intelligence quotient, ADHD and even autism in children” (p. xii). Yet, the WHO-UNEP 2012 report does not provide a reference to support this claim, nor discuss any other factor(s) that may contribute to ASDs. There is no clear etiology for autism and no known MOA for it being induced, thus the current science cannot substantiate endocrine disruption as causal. Many potential non-EDC risk factors have been studied, including genetics, older paternal age, sex, prenatal nutrition, and in utero exposure to antidepressants and pain-killers (e.g., Duchan and Patel, 2012; Gentile et al., 2013; Guinchat et al., 2012; Kinast et al., 2013; Schmidt et al., 2011). Although the weight of the evidence supporting these associations varies, the available data indicate that a good deal of research on possible causes of ASDs is not considered in the WHO-UNEP 2012 report.

Overall, the two references (Rice, 2007; Wing et al., 1976) regarding endocrine disruption and trends reported for ASDs on which the WHO-UNEP 2012 report relies are not representative of the literature as a whole. Thus, the report’s conclusions that there are actual increases in the spectrum of autism-related disorders and that these increases are due to endocrine disruption are not supported based on the current state of the science.

2.5.2. Wildlife population declines

The WHO-UNEP 2012 report stated that the evidence for “endocrine disrupting POPs such as PCBs and organochlorines” (p. 186) as causes of wildlife population declines has increased since 2002 due to observed increases in the populations since restrictions on the use of these chemicals. The logic presented – that as chemical exposures increased, populations declined and, conversely, as chemicals were removed from the market and as exposures declined, populations recovered, would be reasonable if: (1) the chemical exposures are documented; (2) the levels of exposure occurring are sufficient to impact the organisms; (3) the organism-level impacts are manifested in population-level impacts; and (4) other possible causes for population changes are adequately considered. The WHO-UNEP 2012 report falls short in demonstrating the linkages that would be required to make a case based on all of these points.
The two most prominent examples cited in the WHO-UNEP 2012 report are links between DDT and bird populations and between tributyl tin (TBT) and snail populations. For the latter example, the report cites publications by Jörundsdóttir et al. (2005) and Morton (2009). Jörundsdóttir et al. (2005) observed reductions in the levels of imposex in the dogwhelk (Nucilla lapillus) in Iceland, mainly near small harbors with no change seen in larger harbors. As no measurements were made of TBT concentrations, Jörundsdóttir et al. (2005) stated that the continued impacts in the large harbors are “presumably associated” with continued use of TBT paints on larger vessels. Morton (2009) documented a 20-fold increase in the population of N. lapillus on the southeastern coast of England during the period May 2004–August 2008, which coincided with the period over which TBT was banned as an antifouling paint globally. Morton (2009) stated that, “due to the 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 ambient TBT levels.” The WHO-UNEP 2012 report discusses recovery in the abundance of North Sea brown shrimp, although there is no known mechanism of endocrine disruption by TBT in crustaceans. Verhaegen et al. (2012), as cited in the WHO-UNEP 2012 report, state that the inability to demonstrate an “unarguable causative link” between decreased organotin concentrations and recovery of the shrimp stock is due to the lack of data on both exposure and effects in these organisms. None of these weaknesses in the conclusions of the cited studies are mentioned in the WHO-UNEP 2012 report, and therefore, the observed trends of imposex and the recovery of snail and shrimp populations cannot be definitively attributed to TBT, much less endocrine disruption.

Although the ability of organotins to cause masculinization of female gastropods (including the development of imposex) is probably the most-recognized EDC effect in wildlife over the past 30+ years, the vast majority of field studies on this phenomenon do not include chemical analyses of body burdens (Titley-O’Neal et al., 2011). This review article by Titley-O’Neal et al. (2011) is cited in the WHO-UNEP 2012 report, but a number of interesting points from the review were not mentioned. For example, the WHO-UNEP 2012 report does not note the lack of agreement among researchers on the mechanism for induction of effects, the observation of imposex prior to the use of TBT, the natural occurrence of imposex in some species, the lack of sensitivity of a number of species to TBT-induced imposex, and the fact that female masculinization by TBT or triphenyl tin (TPT) has been confirmed in the laboratory in only a small fraction of species affected (7.5% or 20 species confirmed out of 268 total species examined). Thus, the statement in the WHO-UNEP 2012 report that the “temporal relationship between a measure of exposure and population parameters” for TBT as an example of the “best evidence of a relationship between EDCs and wildlife populations,” does not reflect the uncertainties in the available information, including the studies cited in the report.

The WHO-UNEP 2012 report acknowledges the difficulty in making the link between declines/recoveries in wildlife populations and EDCs, stating that many factors may be responsible. These factors may include food, habitat, competition, predation, overall environmental quality, climate change, and human activities (e.g., harvesting, traffic, noise). Regardless, the report emphasizes chemicals, specifically EDCs, as the main causative factor. Even in the case of TBT and N. lapillus, which is arguably the best known example of EDC effects on wildlife, there are other factors that impact the distribution and abundance of this gastropod. This species is sensitive to changes in nutrient levels, substrate loss, toxic algal blooms, and oil spills (Bryan, 1968; Gibbs et al., 1999; Robertson, 1991). For the DDT example cited in the WHO-UNEP report, a number of confounding factors are likely to have affected the recovery of osprey populations, as discussed by Henny et al. (2010), who stated that “expansion of suitable habitat (reservoirs) and enhanced use of artificial nest sites confounds a simple conclusion that recent population increases were solely a recovery from earlier contaminant exposure,” especially in the western United States. The WHO-UNEP 2012 report concluded that the “strength of the evidence linking EDC exposure to most wildlife population declines is insufficient,” then goes onto make the statement that “an endocrine mechanism for wildlife declines is probable but not conclusive” (p. 186). It would be more appropriate to conclude that the evidence for an endocrine mechanism is hypothetical, particularly given the fact that for the two best known examples for wildlife declines, DDT and TBT, an endocrine mechanism, while possible, is only one of many potential factors that may be contributing to the observed population dynamics.

2.6. Importance of dose–response and potency

In Chapter 1 of the WHO-UNEP 2012 report, an effort is made to describe the endocrine system in general – the glands involved, hormones produced, molecular mechanisms involved in mediating responses, and the physiological processes that are regulated by this system. A few of the feedback mechanisms that are an integral part of this system (e.g., how insulin secretion is affected by changes in blood glucose levels) are also mentioned. These various negative feedback loops are important in regulating the production and release of hormones, the expression of various hormone receptors, and generally maintaining homeostasis (i.e., a stable internal environment). What the WHO-UNEP 2012 report fails to fully discuss, however, is the fact 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 (Goodman et al., 2010; Rhomberg et al., 2012). In fact, the responsive nature of the endocrine system is essential to health as seen in the hormonal changes that occur when a woman becomes pregnant. In other words, not all modulations of endocrine function are necessarily adverse. Based on this fact, it can be generally accepted that endocrine activity observed through in vitro testing (or even some in vivo assays) is not sufficient to classify a substance as an endocrine disruptor if these tests do not address whether the alterations cause actual harm in a whole organism or its offspring. Rather, such a substance may be considered endocrine-active only (EFSAs, 2013a); without a clear indication of consequent adversity in a living organism, the substance does not reach the level of an endocrine disruptor. Nevertheless, the WHO-UNEP 2012 report often presented evidence of in vitro or in vivo endocrine modulation (rather than adversity) as support for certain substances being classified as EDCs. For example, in the section on adrenal disorders, only in vitro data are presented as evidence of potential endocrine disruption in humans; no epidemiologic evidence of adrenal effects in people is available. Given the fact that a large proportion of the cited in vitro studies relate to alterations in gene expression, the data presented fall short of demonstrating an adverse effect and are not sufficient for establishing endocrine disruption. Despite the lack of robust evidence for adrenal disorders in humans as a result of exposure to environmental chemicals, the WHO-UNEP 2012 report identified the adrenal cortex as “the most commonly affected and vulnerable endocrine organ in toxicity” (p. 148).

The WHO-UNEP 2012 report contends that hormonal dose–response curves are non-monotonic and states that NMDRs for EDCs are to be expected. As support for NMDRs, the WHO-UNEP 2012 report cited a recent review by Vandenbarg et al. (2012) in which numerous examples were presented of EDCs that exhibit these types of behaviors. This review has been duly criticized, however,
for its selective dismissal of studies that do not show these effects and the general acceptance of those studies that do show these types of responses without any type of critical evaluation of study quality; the inclusion of studies that do not address adverse effects, but rather, transient, adaptive responses; and a failure to consider whether the doses examined in these studies are of any relevance to human exposure levels (Rhomberg and Goodman, 2012).

The Danish Centre on Endocrine Disrupters, in its examination of the evidence presented by Vandenberg et al. (2012) for NMDRs, noted that the majority of these data were from *in vitro* studies and inappropriately included findings for which the U-shaped or inverted U-shaped curves were the product of general toxicity (DTU Food, 2013). It was concluded that 45% of the *in vitro* examples cited by Vandenberg et al. (2012) were the result of cytotoxicity and thus were not examples of true NMDRs. Of the remaining examples cited, approximately one-third were judged to be false and another one-third were considered questionable. Further, only 5 of the 34 *in vivo* examples cited by Vandenberg et al. (2012) were considered to show “clear evidence” of NMDRs. In other words, while examples of NMDRs do exist, they are not as common as Vandenberg et al. (2012) suggests. Recently, USEPA also conducted an expert review of the experimental evidence for NMDRs (USEPA, 2013). In this draft report, EPA noted that such responses are not uncommon in *in vitro* studies and often relate to “lower-order biological endpoints” rather than apical endpoints. However, “[t]here is currently no reproducible evidence that the early key events involved in the expression of NMDRs that are identified at low doses are predictive of adverse outcomes that may be seen in humans or wildlife populations for estrogen, androgen or thyroid endpoints” (USEPA, 2013, p. 8). USEPA concluded that, while NMDRs for adverse effects have been occasionally seen in intact organisms, NMDRs are relatively uncommon. Further, such dose–response curves – when observed – typically occur at high doses, well above the NOAELs identified in standard testing paradigms. In summary, the limited available evidence for low dose effects and NMDRs does not preclude the need to consider dose in assessing the potential hazards of chemicals to the endocrine system.

The WHO-UNEP 2012 report also asserts the potential for EDCs to act at very low doses (i.e., doses below the no observed adverse effect level [NOAEL] or doses below a dose that is environmentally relevant to humans). The implication is that no threshold can be characterized for adverse effects; therefore, any exposure poses some kind of concern, though the magnitude of the effect, or even its direction, let alone its adversity, is not considered relevant from this perspective. However, there are many reasons to question the assertion of effects at very low doses. First, as noted by EFSA (2010), studies of low-dose effects often suffer from various methodological shortcomings (including the use of small numbers of animals and single doses). Additionally, their findings are of questionable toxicological relevance and frequently cannot be replicated in subsequent, more robust studies (EFSA, 2010).

Second, the WHO-UNEP 2012 report generally does not address the fact that many EDCs have much lower potency than endogenous hormones (Nohynek et al., 2013; Sharpe, 2003). In fact, the report claimed that in the diethylstilbestrol (DES) case study, while other chemicals may be less potent than DES, these effects are “equally undesirable when the exposure occurs in early development where potency seems less important” (WHO-UNEP, 2012a, p. 25). In other words, the report suggests that at vulnerable developmental stages, potency may not be very relevant. This statement confuses the issue of potency and sensitivity at different life stages. Although certain substances may be more or less potent depending on the particular life stage at which exposure occurs, potency is always important – no matter the developmental window. The WHO-UNEP 2012 report rightly stated that hormone potency and receptor affinity are not the same things (p. 12). It is further suggested that potency depends on many different factors, but it is vague as to what those factors might be (other than receptor abundance). At the receptor level, potency is determined by both the affinity of a substance to bind to a receptor site as well as the efficacy with which the affinity activates the receptor (Borgert et al., 2013). Endogenous hormones have both strong affinity for their receptor sites as well as high efficacy for activation of these receptors; thus, hormones generally are highly potent for modulating endocrine function. Exogenous chemicals, on the other hand, are rarely as potent as hormones, either due to reduced affinity, reduced efficacy, or both (e.g., Gaido et al., 1997; Nilsson, 2000). Although the WHO-UNEP 2012 report claims that “very low concentrations of environmental endocrine disruptors could add to the endogenous hormone effect to produce a response that is much greater than would be predicted based on the hormone alone” (p.8), this theory fails to consider the existence of biological thresholds. As described by Borgert et al. (2013), given the lower potency of most exogenous chemicals, the additional presence of these chemicals will not significantly alter hormone receptor occupancy; thus, a biological threshold for potency exists.

At the whole organism level, potency relates to the ability of a substance to produce a biological effect and may be substantially different from the potency measured with *in vitro* assays (EFSA, 2013a). In considering potential EDCs, therefore, potency should not be determined based on the results of *in vitro* studies. Rather, EFSA recommends that potency should be based on the ability of a substance to produce an adverse health effect in *vivo* (EFSA, 2013a). This ability will depend on not just a substance’s potency at its receptor site, but also on its disposition in the body, the timing of exposure (i.e., the particular life stage of development) and the dose and duration of exposure. Therefore, dose remains an important factor in assessing the potency of potential EDCs to cause adverse health effects.

Endocrine disruption is generally posited throughout the WHO-UNEP 2012 report in terms of “adverse” outcomes, yet the report fails to provide a concrete definition for what may be considered an adverse response. In particular, the 2012 report did not adopt the IPCS (2004) definition of an adverse health effect: “change in morphology, physiology, growth, development, reproduction or life span of an organism, system, or (sub) population that results in an impairment of functional capacity, an impairment of the capacity to compensate for additional stress, or an increase in susceptibility to other influence.” The need to clearly delineate adverse from adaptive responses, particularly when considering results from *in vitro* assays, was addressed and possible definitions for these terms proposed in a recent workshop (Keller et al., 2012). Certain endocrine-mediated adverse effects – such as cancer or reproductive disorders – clearly can be judged as detrimental. For other endpoints – such as alterations in hormone levels – it is more difficult to delineate an adaptive response that is within the limits of homeostasis from one that has gone beyond those limits for a sufficient period of time and, therefore, capable of causing an adverse effect. This is particularly true when hormone levels are only measured at a single time point shortly after exposure with no indications of whether the response is transient or more permanent. Thus, the mere presence of a change does not necessarily mean that the outcome is adverse. For example, in the discussion on thyroid-related disorders and diseases (WHO-UNEP, 2012a, p. 97), the report suggests that chemicals can interfere with thyroid hormone signaling without affecting serum hormone levels, but it is stated that methods to evaluate this are not yet available. Changes in thyroid hormone signaling alone cannot be characterized as adverse without evidence to show that these changes lead to impairment in function (Bianco and Kim, 2006). Furthermore, the failure to observe changes in serum thyroid hormone levels would indicate a lack of consequence from the change in signaling.
In its opinion on the scientific criteria for the identification of endocrine disruptors, EFSA specifically discussed a “threshold of adversity,” noting that toxicologically relevant responses occur only when the degree of endocrine modulation elicited is beyond that which could be counteracted through homeostatic mechanisms (EFSA, 2013a). These thresholds are similar to those that exist for responses measured from other physiological systems operating within the body. Further, their importance in characterizing endocrine disruption has been emphasized by toxicologists concerned about the European Commission’s recommendations for the regulation of EDCs (Dietrich et al., 2013). In contrast, the WHO-UNEP 2012 report proposes that thresholds for endocrine disruption “should not be assumed” (p. 19) and exposures to endocrine-active substances – no matter the level – will add to the already present hormone levels in the body and thus alter endocrine function in a threshold-independent manner. However, using mathematical calculations within a systems biology construct, Borgert et al. (2013) have posited that the endocrine system is able to discriminate potent hormonal signals from the “background noise” of other endogenous molecules, making the system relatively robust in its responses and resistant to spurious interferences by substances with lower potency.

Because thresholds exist, not only for inducing an endocrine response, but also for moving beyond adaptive modulation toward adversity, it is important to understand at what doses the observed responses occur and how these doses compare to the levels at which people or wildlife are typically exposed. In the discussions in Chapter 2 of the report regarding various adverse outcomes, however, the WHO-UNEP 2012 report often fails to mention the doses at which findings are observed. It is important to note that doses administered in experimental animal studies are often orders of magnitude above those to which people or wildlife are generally exposed. Frequently, if one delves deeper into the scientific literature as seen in the examples below, it is apparent that the doses associated with the reported findings are extremely high, well above those to which people or wildlife may be typically exposed.

It is unclear why the WHO-UNEP 2012 report fails to consider dose in the discussions of evidence for endocrine disruption in humans and wildlife. In the beginning of Chapter 2 of the report, it is noted that the focus was on the “identification of the characteristics of the hazards posed by endocrine disruptors rather than risk assessment,” (WHO-UNEP, 2012a, p. 23) because accurate risk assessments are difficult in light of limited human exposure data and the combined effects of mixtures. However, the report often draws conclusions that appear to go beyond a simple assessment of potential hazard. For example, the report concluded that environmental exposures play a role in the observed increased incidences of hormonal cancers — rather than saying that these exposures have been associated with the cancers (WHO-UNEP, 2012a,b, p. 137). Similarly, the report states that adrenal changes seen in Baltic seals were caused by exposure to DDT, PCBs, and their metabolites — instead of saying that they have been associated with these exposures (WHO-UNEP, 2012a, p. 147). Further, at the end of Chapter 1, the WHO-UNEP 2012 report stated that “best professional judgment was used to make expert assessments of the data linking exposure to chemicals with each disease/dysfunction,” (p. 19) and relevant exposures and dose–responses were considered. Therefore, although dose and exposure were specified as important factors in the evaluation of endocrine disruption, in reality, this does not appear to have been the case.

To illustrate, a few selected examples are discussed below. These are not isolated examples, but rather, representative of how the lack of the consideration of dose in the WHO-UNEP 2012 report leads to a false impression that humans are at risk of endocrine effects from their daily exposures to chemicals.

2.6.1. DES or genistein and endometrial cancer

In the discussion of animal studies of EDCs and endometrial cancer, the WHO-UNEP 2012 report cites the study of Kabbarah et al. (2005) as showing “[g]reater than 90% of CD-1 pups neonatally exposed to DES or the phytoestrogen genistein develop endometrial cancer by 18 months of age whilst C57Bl/6 mice are resistant” (WHO-UNEP, 2012a, p. 130). In this study, both DES and genistein were injected subcutaneously into the pups (a route of administration not relevant to environmental exposures for humans) at doses of 1 and 50 mg/kg/day, respectively, on postnatal days 1–5. The DES dose is over 1000-fold higher than the typical estrogen dose that women receive from low-dose daily oral contraceptive pills (Kripe, 2005). Further, the daily intake of genistein in Japanese subjects, a population known to have high intake of soy isoflavones, has been shown to be <1 mg/kg/day (Nakamura et al., 2000; Wakai et al., 1999), at least 50-fold lower than the dose administered to mice in Kabbarah et al. (2005). Thus, the doses of DES and genistein used in this study are well beyond those to which people would be typically exposed. Another issue is that this study was actually conducted in knockout mice with a genetic predisposition for DNA repair errors – a fact that is not mentioned in the WHO-UNEP 2012 report and that further brings the human relevance of the findings into question.

2.6.2. PCBs and neurodevelopmental effects

In another example, in the discussion on neurodevelopmental disorders, the WHO-UNEP 2012 report cites three studies as consistent evidence that “PCB exposures decrease serum thyroid hormone levels,” (p. 113) with no mention of study details. However, the doses at which effects were observed in those studies are extremely high. In Goldney et al. (1995) and Zoeller et al. (2000), rats were exposed to 1, 4, or 8 mg/kg/day of Arochlor 1254 on gestational days 6–21. Although circulating T4 levels were reduced on postnatal days 1–30, they recovered by postnatal day 45, indicating a transient effect. More importantly, Goldney et al. (1995) reported that pup mortality was 20% and 50% in the 4 and 8 mg/kg/day dose groups, respectively, indicating that these doses were extremely high. In the third study (Bastomsky, 1974), adult rats were injected with an even higher dose of 25 mg/kg/day of Arochlor 1254 for 4 days. In contrast, the mean intake of PCBs from consumption of the French diet was recently estimated at 2.71 ng/kg/day for adults and 3.77 ng/kg/day for children (Sirot et al., 2012), while that from consumption of the Japanese diet was estimated at 1.45–2.08 pg/kg/day (Nakatani et al., 2011). Thus, human exposures to PCBs in the diet are over 300,000 times lower than the doses used in the experimental studies (on a human equivalent dose basis) cited in the WHO-UNEP 2012 report as consistent evidence of effects of PCBs on thyroid function. This dramatic difference in dose calls into question the relevance of these studies to human environmental exposures.

It should be further noted that, although the thyroid develops and functions in a manner generally similar between rodents and humans, differences exist that make neonatal rats more susceptible and less capable of compensating for possible alterations in function than humans. For example, the human fetal pituitary can respond to thyrotropin–releasing hormone (TRH) as early as gestation week 25 and thyroid-stimulating hormone reaches peak serum levels somewhere around this same time, while the hypothalamic–pituitary–thyroid axis in rats does not respond to TRH signals until a couple of weeks after birth (Howdeshell, 2002). Further, free thyroid hormone levels can be maintained during pregnancy in humans via increased peripheral metabolism and enhanced thyroid hormone binding to serum proteins (Ahmed et al., 2008; Howdeshell, 2002). Consequently, children with congenital hypothyroidism may be born with low-normal concentrations of TRH.
thyroid hormone due to compensation by the maternal system (Ahmed et al., 2008).

2.6.3. BPA and adverse effects

Throughout the WHO-UNEP 2012 report, BPA is mentioned as being responsible for a variety of adverse findings in rodents, including fibroid development in mice and rats (p. 42); defeminization and other alterations in social behaviors in female rats (p. 115); altered mammary gland development leading to increased tumor induction (p. 128); endometriosis in offspring of exposed mice (p. 130); and modified immune responses in mice (p. 169), to name a few. In all of these cases, the doses of BPA associated with these findings are not reported. Further, in Chapter 3 of the WHO-UNEP 2012 report, the various ways in which people may be exposed to BPA are emphasized (e.g., in the call-out box on page 196 under the heading called “origin and use”). The report also notes that BPA is found in virtually all people (WHO-UNEP, 2012a, p. 225), but no information is provided on the magnitude of exposures, the biological concentrations that have been measured in people, or whether these would be sufficient to cause adverse effects. The implication is that people are at risk of adverse health effects because they are exposed to BPA. However, a number of recent weight-of-evidence evaluations have been conducted to assess the potential risks to humans from BPA exposure (EFSA, 2013b; Goodman et al., 2006, 2009; Hengstler et al., 2011; Teegarden and Hanson-Drury, 2013). These reviews document that some BPA results reported in investigatory experiments have not been replicated in subsequent studies and that many studies have used non-oral exposure routes that bypass first-pass liver metabolism and thus are not relevant to human oral exposures. The majority of BPA studies have been conducted at doses well above those to which humans are generally exposed and human exposures are generally well below the current BPA tolerable daily intake (TDI) of 0.05 mg/kg/day derived from two- and three-generation reproductive studies in rodents. More specifically, daily BPA exposures were recently estimated by EFSA (2013b) to be ≤857 ng/kg/day for toddlers and ≤495 ng/kg/day for infants 1–5 days of age; these values are 50–100-fold lower, respectively, than the BPA TDI value of 0.05 mg/kg/day. Daily BPA exposures were also estimated by the U.S. Food and Drug Administration (USFDA) to be 100–200 ng/kg/day for children and adults and 200–400 ng/kg/day for infants (USFDA, 2009); these values are even lower than those estimated by EFSA. In other words, the implication of human health risks from BPA exposure raised in the WHO-UNEP 2012 report is unfounded when the data are considered in the context of actual doses administered and concentrations to which people are typically exposed.

These examples demonstrate that dose was not considered in the discussion of experimental animal studies in the WHO-UNEP 2012 report. In all of the illustrated cases, potential human exposures are orders of magnitude lower than those administered in toxicology studies. In addition, some of these studies utilized routes of exposure (i.e., injection) that bypass normal metabolism and elimination of chemicals and therefore, are not relevant for assessing potential environmental exposure to humans. Further, the WHO-UNEP 2012 report did not address other dose-related issues such as thresholds for effects and potency, when discussing specific chemicals.

3. Conclusions

The WHO-UNEP 2012 state-of-the-science report on endocrine disruptors is purported to be an update of the WHO-IPCS 2002 state-of-the-science report – however, it is neither a state-of-the-science review, nor is it an update of the 2002 report. The 2012 report cannot be characterized as a state-of-the-science review because it lacks several key features for this type of assessment including: the lack of a defined scope for the review, the absence of a process for identification, integration, and interpretation of data, the lack of a structure for evaluating individual studies for relevance and reliability, and an objective method for evaluating the weight of the evidence. These deficiencies undermine the conclusions reached in the report. The WHO-UNEP 2012 report can be more appropriately characterized as a selected discussion of aspects of science that should be considered when discussing endocrine disruption, but it is not a summary of the current state of the science.

Neither can the WHO-UNEP 2012 report be considered an update to the WHO-IPCS 2002 report because it does not build on and modify the earlier analysis, giving reasons and support for the changes in the state of the science. A true update to the earlier report would cite the 2002 conclusions, articulate what data, findings, or new understanding since 2002 should be considered and evaluate how and whether the 2002 conclusions need to be modified in light of the newer information. In addition, the WHO-UNEP 2012 report does not address research recommendations from the earlier report. In some cases, the 2012 report reviews the same data from the 2002 report, but reaches conclusions that conflict with those of the earlier report, despite the lack of new information to support a change in the weight of evidence. 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. Although the WHO-UNEP 2012 report is stated to be an update, this report in actuality is a revised review of the state of the science that does not build upon what was previously done and disregards the WHO-IPCS 2002 proposed framework for causation in favor of “best professional judgment” on these matters.

A key concern with the WHO-UNEP 2012 report is the use of subjective inference instead of a formal framework to assess the potential role of causation for endocrine disruption. The report adopted a narrative approach for the data review that does not represent a weight-of-evidence assessment. Rather than demonstrate causation, the report relies on inference to suggest that exposures to chemicals and adverse outcomes are related. The WHO-UNEP 2012 report presented information on chemicals and various adverse outcomes, but whether the exposure causes these effects was not determined or adequately considered in an objective, transparent, and scientific manner. Several key factors for establishing causation, such as a demonstrated exposure to the chemical, dose–response, and consistency in the data were frequently ignored. For example, temporal trends in human diseases or wildlife populations are presented without consideration of alternative explanations for these trends (especially diagnostic criteria and reporting changes). Exposures to chemicals considered to have the potential for endocrine disruption exist and are suggested as contributing to the observed trends, but there is little consideration of whether these exposures are sufficient to explain the alleged effects and whether the patterns of exposure are congruent with the trends. Above all, the lack of a framework to collectively evaluate, in an objective and comprehensive manner, the data on specific chemicals and the alleged adverse outcomes is a significant shortcoming in the WHO-UNEP 2012 review.

The WHO-UNEP 2012 report fails to fully address a number of critical factors that must be considered when defining EDCs; specifically dose, dose–response, potency, and adversity. First, the substance must be shown to cause an adverse effect in an intact organism, their progeny or (sub)populations; therefore, in vitro data alone are insufficient for classifying a compound as an endocrine disruptor. Further, the observed effect must be shown to go
beyond adaptive modulation of endocrine function; that is, it must result in an adverse outcome. Second, thresholds exist for inducing such adverse effects. The 2012 report does not give appropriate consideration to thresholds and dose–response for adverse effects. Despite the chosen definition of endocrine disruption as producing adverse changes, the report treats all effects as evidence of disruption and makes an a priori rejection of thresholds. Where examples from animal testing are discussed there is little consideration of dose–response, when, in fact, only some doses of some compounds can cause endocrine disruption in the laboratory. The WHO-UNEP 2012 report also does not address the fact that most EDCs have much lower potency than endogenous hormones and potency is important regardless of the life stage at which exposure occurs. Finally, consideration of low dose effects and NMDRs do not exclude the need to consider potency and exposure.

It is also important to emphasize that the Summary for Decision-Makers, while implied by the title to be a synopsis of the main report, is not truly representative of the main report. In many cases, the Summary for Decision-Makers compounds the limitations of the main report by making statements without supporting references and providing more definitive conclusions. Consequently, this companion report cannot be considered a Summary nor should it be relied on to make decisions regarding the regulation of endocrine disruptors.

Overall, the WHO-UNEP 2012 report on endocrine disruptors fails to achieve its objectives as an updated state-of-the-science review on endocrine disrupting chemicals, and therefore, should not be used to support evidence-based decisions. The scientific literature on endocrine disruption is voluminous, complex, heavily nuanced and covers multiple disciplines, and has certainly expanded greatly since the WHO-IPCS 2002 report – thus, warranting a true update of the state of the science. Science and policy decisions should not rely on professional opinion alone; rather, efforts to advance our understanding of the potential impacts of endocrine disrupting chemicals on human health and wildlife need to be based on objective and systematic reviews that transparently capture the best available science and rely on explicit criteria for the evaluation of the evidence. Therefore, a balanced and objective review of all of the available literature with clearly stated objectives and limitations is essential. Moreover, any updated review of the literature must acknowledge controversies regarding the issues, identify data gaps, and provide clear statements of conclusions that are supported by the data. Only through such an approach can all parties, regardless of position on the issues, be confident in evidence-based decisions that arise from such a report.

Conflict of interest

The employment affiliations of the authors are as shown on the cover page. James Lamb, Karyn Hentsz, Jane Staveley, Gerard Swaen, and Amy Williams are employees of Exponent. Exponent is a publically traded engineering and scientific consulting firm that provides expertise in toxicology, ecological toxicology, and epidemiology. James Lamb has been a testifying expert in litigation involving several chemicals discussed in this review, including pesticides, PCBs and dioxins. Lorenz Rhomber and Julie Goodman are employees of Gradient, a private consulting firm that provides services to both private and public organizations on toxicological and human health risk assessment issues. Paolo Boffetta is on the faculty of the Icahn School of Medicine at Mount Sinai and, as an independent consultant, has provided technical advice to industry and government on cancer epidemiology and has been a testifying expert in litigation involving PCBs. Warren Foster is on the faculty of McMaster University and, as an independent consultant, has provided technical advice to industry and government organizations on the health impacts associated with exposure to environmental chemicals. Glen Van Der Kraak has served as a consultant to industry regarding endocrine-related effects on wildlife. This review has been conducted with funding support from several sponsors: American Chemistry Council (ACC), CropLife America (CLA), CropLife Canada (CLS), CropLife International (CLI), European Chemical Industry Council (Cefic), and European Crop Protection Association (ECPA). These sponsors were provided an opportunity to review a draft of the paper and offer comments for consideration by the authors. The authors have sole responsibility for the content and the writing of the paper. The interpretations and views expressed in the paper are not necessarily those of the sponsors, or the authors’ employers or clients. Both Glen Van Der Kraak and Warren Foster were members of the Steering committee for the WHO-IPCS 2002 Global Assessment of the State of the Science of Endocrine Disruptors; Glen Van Der Kraak also served as one of the four editors for this report.

References


Danish Centre on Endocrine Disrupters (DTU Food). 2013. Input for the REACH-review in 2013 on endocrine disruptors (tærskelværdi-projekt, j.nr. MST-621-0004). National Food Institute of Denmark (DTU), National Institute of Food, Division of Toxicology and Risk Assessment.


European Food Safety Authority (EFSA). 2013b. DRAFT scientific opinion on the risks to human health related to the presence of bisphenol A (BPA) in foodstuffs – part: exposure assessment. EFSA J.


