

Using systematic reviews for hazard and risk assessment of endocrine disrupting chemicals

Anna Beronius¹ · Laura N. Vandenberg²

Published online: 4 February 2016
© Springer Science+Business Media New York 2016

Abstract The possibility that endocrine disrupting chemicals (EDCs) in our environment contribute to hormonally related effects and diseases observed in human and wildlife populations has caused concern among decision makers and researchers alike. EDCs challenge principles traditionally applied in chemical risk assessment and the identification and assessment of these compounds has been a much debated topic during the last decade. State of the science reports and risk assessments of potential EDCs have been criticized for not using systematic and transparent approaches in the evaluation of evidence. In the fields of medicine and health care, systematic review methodologies have been developed and used to enable objectivity and transparency in the evaluation of scientific evidence for decision making. Lately, such approaches have also been promoted for use in the environmental health sciences and risk assessment of chemicals. Systematic review approaches could provide a tool for improving the evaluation of evidence for decision making regarding EDCs, e.g. by enabling systematic and transparent use of academic research data in this process. In this review we discuss the advantages and challenges of applying systematic review methodology in the identification and assessment of EDCs.

Keywords (4–6): endocrine disruptor · Guideline endpoint · Hazard assessment · Low dose · Study evaluation · Threshold

1 Introduction

The question of whether chemicals used in industrial processes, consumer products and food production have adverse effects on health is not a new one. Although Rachel Carson's book *Silent Spring* brought public awareness to this issue in the 1960s, since that time the number and volume of chemicals used in commerce has continued to increase [1]. A subset of chemicals has garnered significant attention over the last two decades, not only by public health officials, risk assessors and toxicologists, but also by endocrinologists. These chemicals, coined 'endocrine disrupting chemicals' (EDCs), are studied for their ability to interfere with normal hormone action. To date, databases from the US Food and Drug Administration and the Endocrine Disruption Exchange estimate that at least 1000 chemicals may have endocrine disrupting properties [2, 3]. Most EDCs identified to date interact with hormone receptors as agonists or antagonists [2, 4, 5], but others act via altering hormone "production, release, transport, metabolism, binding, action, or elimination" [4]. Importantly, these chemicals are used in a range of consumer products including personal care products, detergents, upholstery and fabrics, food contact materials, medical equipment, as well as other uses resulting in exposure to the general population, such as pesticides (herbicides, insecticides, fungicides, etc.) [6–8].

In 2013, an updated State-of-the-Science review of the EDC literature was published by the World Health Organization (WHO) and United Nations Environment Programme (UNEP) which concluded that there was sufficient evidence from controlled laboratory studies to indicate

✉ Laura N. Vandenberg
lvandenberg@schoolph.umass.edu

Anna Beronius
Anna.Beronius@aces.su.se

¹ Department of Environmental Science and Analytical Chemistry, Stockholm University, Stockholm, Sweden

² Department of Environmental Health Sciences, University of Massachusetts Amherst School of Public Health & Health Sciences, 149B Goessmann, 686 N. Pleasant Street, Amherst, MA 01003, USA

that EDCs can induce endocrine disorders, and thus EDC exposures may be implicated in many of these conditions in humans and wildlife [9]. The report also noted that human disease trends for many endocrine-related disorders are increasing and that growth and development of wildlife species have also been affected.

In fall 2015, the Endocrine Society released a Second Scientific Statement on EDCs [10], updating a prior statement that was published in 2009 [11]. The second report, termed EDC-2, reviewed 1300 studies of human populations and laboratory experiments, many of which showed relationships (associations or causal links) between EDC exposures and diseases. Focusing largely on data produced over the past five years, EDC-2 drew conclusions about the strength of evidence between EDC exposures and obesity, diabetes and cardiovascular disease; female reproductive disorders; male reproductive disorders; hormone-sensitive cancers; thyroid conditions; and neurodevelopmental and neuroendocrine effects [12].

The possibility that EDCs in our environment are a contributing factor to endocrine related effects in both human and wildlife populations implies that past risk assessments of these chemicals have failed to adequately protect human health and the environment. The publication of the 2013 WHO report [9] and the 2015 EDC-2 Statement [10], along with other reviews of large portions of the EDC scientific literature [11, 13–16], have heightened ongoing debates about the hazards and risks associated with these chemicals. In fact, a number of back-and-forth exchanges between different groups of scientists, including those associated with or funded by chemical industries, have highlighted what appears to be a deep divide in the scientific community about many issues associated with the identification, study and assessment of EDCs (see [17, 18], and [19–22], and [23–25]). Considering the strength of the evidence detailed in the UNEP/WHO report and EDC-2, some have begun to argue that this is a new example of ‘manufactured doubt’ [24]. This term indicates a financial incentive to block consensus building and public health protective decisions; methods to ‘manufacture doubt’ were originally developed by the tobacco industry to fight association of its products with lung cancer and other adverse health outcomes [26, 27].

One criticism of the WHO/UNEP report and other reviews that have been conducted on large portions of the EDC literature has focused on the methods that were used to select the studies to be assessed. A number of critics have noted that systematic review methods were not used, with the implication that the use of systematic review methods would have led to different conclusions about the strength of the evidence linking EDC exposures to health outcomes [17, 23, 28]. These critiques fail to note that agreed upon methods for systematic review for EDCs do not yet exist. Here, we will review the hazard and risk assessment process for EDCs and other environmental chemicals, and the role that systematic review

could play in improving the identification and assessment of EDCs for decision-making. We will describe the methods that are currently available to systematically review and integrate evidence from human epidemiology, in vitro mechanistic, laboratory animal, wildlife, and in silico studies of EDCs. We describe why EDCs may require systematic review methods that are distinct from the methods that are used for other environmental chemicals. Finally, we discuss some of the limitations to systematic reviews, especially as they relate to large analyses like the UNEP/WHO and EDC-2 reports.

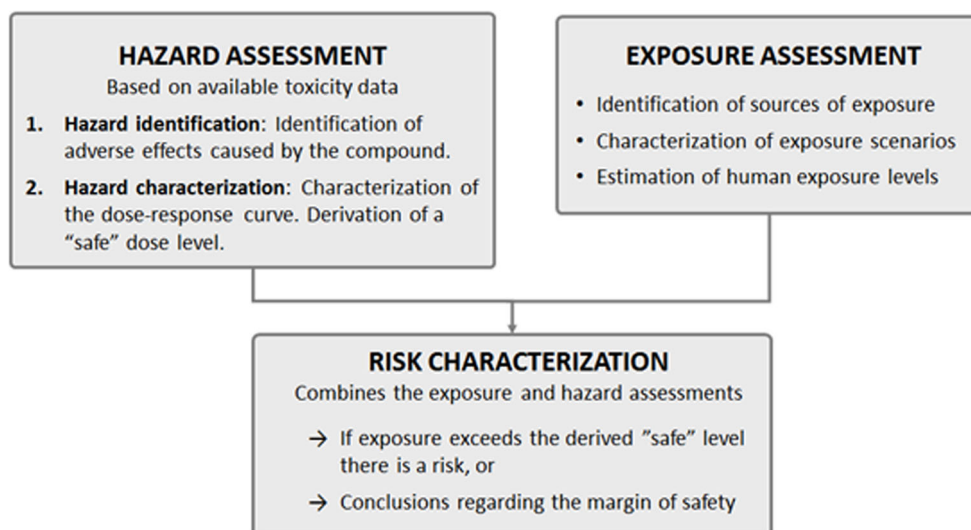
2 A brief overview of the risk assessment process for EDCs

Generally speaking, the purpose of risk assessment of chemicals is to evaluate scientific data in order to provide a scientific basis for decision makers to act (if needed) to reduce chemical risks to human health or the environment. For example, regulatory risk assessment is used as the basis for approving or restricting different uses of chemicals. The risk assessment process consists of three main parts: hazard assessment (including hazard identification and hazard characterization), exposure assessment, and risk characterization. Figure 1 summarizes these parts, as they relate to human health risk assessment.

In hazard assessment, toxicological data is reviewed to identify the adverse effects of the compound (hazard identification), as well as to characterize the dose-response and draw conclusions on a “safe” dose, e.g. calculating a reference dose for humans (hazard characterization). Hazard assessment is often based primarily on experimental animal data; if human data is available it is used as supportive evidence. The exposure assessment aims to identify all sources of exposure to a compound and estimate exposure levels in the population of interest, e.g. the general population, certain workers, or a certain age group. In the risk characterization step, the results from the hazard and exposure assessments are combined to draw conclusions concerning risk, e.g. if the estimated exposure exceeds calculated “safe” doses or if the margin between doses where adverse effects arise and exposure levels is/is not sufficient. This review will primarily focus on issues related to the hazard assessment step.

Risk assessment is based on a number of basic principles and assumptions; in some cases, these assumptions have been challenged by EDCs. These are summarized below in Sections 2.1–2.4. In addition, several aspects of the risk assessment process rely, more or less, on expert judgement, such as the identification and evaluation of relevant data, determining what effects are to be considered adverse, and the application of assessment (or uncertainty) factors to calculate reference doses. Expert judgement can be argued to play an important role in hazard and risk assessment by allowing

Fig. 1 Health risk assessment of chemicals is conducted in three parts; the results of the hazard and exposure assessments are combined to draw conclusions in the risk characterization



sufficient flexibility to account for all relevant aspects of the scientific question to be answered. However, it inevitably introduces value-based assumptions to the assessment that may influence conclusions. It is of key importance that these assumptions are transparently described and justified [29, 30].

2.1 All relevant data should be considered in risk assessment

Guidance for risk assessment of chemicals issued by different authorities and organizations generally requires or recommends that *all* relevant toxicity data should be considered in the assessment process (e.g. [31–34]). However, the process for identifying and selecting which data are considered as evidence in the assessment is rarely transparent [29, 35]. Risk assessors traditionally rely heavily on studies conducted in accordance with standardized test guidelines and guideline assays (such as OECD test guidelines) which examine ‘validated’ endpoints that have been shown to provide reproducible data in multiple laboratories with appropriate and predetermined levels of accuracy and precision [36, 37]. Such studies are often considered to be reliable and adequate for risk assessment by default, while non-guideline studies, e.g. many studies conducted in academic laboratories, typically are thoroughly evaluated for adequacy before they are included as evidence in health risk assessment [33, 34, 38]. Methods and practices for evaluating toxicity studies for use in risk assessment is further discussed in Section 3.1.

Standardized guideline assays examine a limited number of endpoints and commonly focus on endpoints that are considered apical (e.g. body weight, organ weight, histopathological lesions) and acknowledged to be toxicologically adverse. However, these assays are not comprehensive (i.e. they do not encompass all adverse outcomes) [5, 39, 40]. Numerous groups have concluded that standard toxicological studies in

many cases do not cover the most sensitive endpoints or sensitive windows of exposure and thus are not sufficient to identify EDCs as harmful or for calculating reference doses that are safe for the general public [5, 11, 14, 15, 20–22, 41–45]. For example, the uterotrophic assay, which measures the weight of a rodent uterus in response to a suspected estrogenic compound, was validated by the OECD in a process that took multiple years and participants from dozens of laboratories [46–48]. Although the uterotrophic response is considered ‘validated’, concerns have been raised about its precision; of the ten laboratories that tested the EDC nonylphenol, one identified the lowest observed effective concentration (LOEC) as 35 mg/kg/day, five as 80 mg/kg/day, and one as 100 mg/kg/day [48]. One laboratory could not identify any effective dose, and the final two laboratories did not perform the assay properly [48]. The use of traditional toxicology approaches including the heavy reliance on guideline studies has also been challenged because of the failure of many of the endpoints that are considered ‘adverse’ to anchor to a human disease. For example, many traditional toxicology assays use organ weight as a measure of an adverse effect whereas few human diseases are defined by changes in organ weight or size [14].

Studies conducted in academic laboratories commonly investigate endpoints such as assessments of gene expression, tissue morphology, neurobehaviors, responses to environmental stressors such as hormones and carcinogens, and others that are not covered in standardized test guidelines [39, 49–51]. Well conducted and reported academic research studies may thus provide information that can fill important information gaps in risk assessment and contribute to better targeted conclusions and recommendation for decision making. Unfortunately, many risk assessments dismiss academic studies because the endpoints examined are not considered toxicologically adverse or relevant to humans [5, 40]; this

has led to protracted arguments about what is meant by ‘adverse’ as well as the role of expert judgement and lack of transparency in its determination [41, 52–56]. The adequacy of non-standard studies for the purpose of chemical risk assessment is also often questioned for reasons such as methodological limitations and/or being poorly reported [57–60].

2.2 Thresholds for effect

One of the fundamental principles of toxicology, and consequently in chemical risk assessment, is that there is a threshold for a toxicological response below which no (adverse) effect is expected to occur [61]. In biological terms, a compound’s threshold for effect implies a dose-level below which there are no harmful effects in an organism [62]. However, in the experimental setting, a threshold is the dose-level below which no statistically significant adverse effects are *observed*. This is commonly termed the no observed (adverse) effect level (NO(A)EL). The NO(A)EL is thus by no means a no effect level. Authorities and other experts have lately started recommending a benchmark dose (BMD) approach for characterizing dose-response and deriving a benchmark dose lower bound (BMDL) as an alternative to the NOAEL [63–65]. Nevertheless, the presence or absence of a threshold can never be experimentally proven because all experiments have a limit of detection below which effects cannot be observed, i.e. no conclusion regarding the shape of the dose-response curve can be made below this detection limit [13, 66].

For the majority of chemicals, excluding genotoxic carcinogens, a threshold for effect is assumed when conducting risk assessment [61]. This means that a “safe” dose is derived for humans by dividing the NOAEL or BMDL identified in animal studies with assessment, or uncertainty, factors to account for differences in sensitivity between species as well as between individuals [41]. However, because endogenous hormones are already present, and very small changes in circulating hormone concentrations have been shown to induce biological responses, it can be argued that EDCs can ‘add’ to the actions of existing biological processes [67–69]; thus, a threshold cannot be assumed [22]. This may be specifically relevant during periods of development when hormones play important organisational roles [70–72]. The “additivity-to-background” argument has also been made to support a no-threshold-approach for genotoxic carcinogens [73, 74].

2.3 Epidemiology and other “low dose” studies challenge current risk assessment practices

Risk assessments combine information about hazard with exposure data during the risk characterization process (Fig. 1). Thus, risk assessment is conducted under the premise that reducing exposures to compounds can eliminate the risk,

particularly if a ‘threshold’ for adverse effects has been demonstrated.

Low level exposures to environmental chemicals including many EDCs have been well-documented by biomonitoring programs like those conducted by the US CDC [75–77]. Most of these exposures occur below the acceptable daily intake levels and thus are not expected to induce adverse outcomes. One major area of debate in the study of EDCs is whether “low doses” can induce adverse effects [78]. There are several definitions for a “low dose” that have been proposed including: 1) doses below the toxicological NOAEL; 2) doses in the range that humans (or wildlife) typically experience; or 3) doses that replicate in animals the concentrations that circulate in human bodies, taking into account differences in metabolism and toxicokinetics [15, 79, 80]. Human epidemiology studies are “low dose” studies, and hundreds of these studies have shown associations between EDC exposures and health outcomes; many have been published in just the last two years (for illustrative example, see [81–87]). Although these studies alone cannot indicate causal relationships between EDC exposures and health effects, they provide support for the hypothesis that low doses of EDCs, including those below regulatory ‘safe’ doses, can induce harm.

There are also now hundreds of low dose studies from controlled laboratory animal experiments [15, 50, 88, 89]. Although some low dose effects examined in non-guideline studies of EDCs are overt diseases or dysfunctions (see for example [90–93]), or increased susceptibility to diseases or dysfunctions (see for example [94–96]), most represent pre-disease endpoints [97–109], providing support for the links between EDC exposures and *disease pathways*, if not the apical endpoint itself [110, 111]. These low dose studies suggest that EDCs should not just be tested for their toxicity, but also for their contribution to diseases; the former may not accurately predict the latter. Furthermore, the presence of low dose effects in controlled animal studies suggests that current risk assessments for EDCs may not be sufficiently public health protective [112].

2.4 Non-monotonic dose response relationships

Another standard assumption in risk assessment is that the response above the NOAEL will increase with increasing dose, in a monotonic fashion, i.e. the dose-response curve will not change direction (or the slope will not change sign) anywhere along the dose span [113]. Similarly, monotonicity is used as the underlying principle to justify testing chemicals at high doses in experimental studies and extrapolating down to draw conclusions about toxicity at low doses [15]. A large number of studies have, however, reported non-monotonic dose response curves for both hormones and EDCs. This topic has recently become a highly debated issue for EDCs [16, 41, 113–117].

2.5 Risk assessments for EDCs operate at the interface of toxicology and endocrinology

Numerous groups including members of the Endocrine Society have argued that hazard and risk assessment of EDCs should take into account the principles of endocrinology [5, 15, 41, 45, 118]. Five general principles of endocrinology have been described as relevant to the study of EDCs [9, 41] (see Table 1). A number of different principles guide the study of EDCs from the perspective of toxicology, a field that was developed for the purpose of determining the hazards associated with chemical exposures (Table 2).

Clearly, because EDCs interact in some way with the endocrine system while also having toxic features, the study and assessment of EDCs must consider the features of both the endocrine system and toxic responses. Yet, some of the principles described in the fields of toxicology and endocrinology are at odds with each other, including issues related to “low dose” and the shape of the dose response curve, described above. Because of this, the interpretation of studies used for risk assessment purposes can lead to discrepancies that may be field-specific.

Other factors that are important to the assessment of EDCs may very well apply to environmental chemicals that do not have endocrine disrupting properties. For example, it is well understood that hormones have very different roles during

development compared to adulthood [72, 119], thus EDCs are expected to have more potent effects during critical periods of development [120, 121]. Other environmental chemicals may also have effects that are more profound when exposures occur during development (e.g. thalidomide induced adverse effects on adults including permanent nerve damage in some patients, but effects on developing fetuses were severe and profound, leading this compound to be classified as a teratogen [122, 123]). Similarly, other environmental compounds can act via binding to receptors that are not a part of the endocrine system, and thus alterations to receptor binding are not limited to EDCs (e.g. neurotoxicants, immunotoxicants, and others [124–126]).

3 Systematic review – what is it, and how can it help?

Systematic review is a method that aims to systematically identify, evaluate and synthesize evidence for a specific question with the goal to provide an objective and transparent scientific basis for decision making [127, 128]. Systematic reviews have been primarily used for decision making in the field of medicine and health care, where this approach is relatively well established [127–129]. In recent years, systematic review has been increasingly promoted in the field of environmental health sciences and for the risk assessment of

Table 1 Principles of endocrinology

Principle	Explanation
Hormones are responsible for coordinating the systems of the body, and allowing communication and integration of the various tissues of the body from conception until death.	Although hormones are often considered only for their role in the maintenance of ‘homeostasis’, they have much wider roles that include the development and differentiation of embryonic and fetal tissues, sexual reproductive functions, maintenance of bone density, body weight and cardiovascular health, and many other endpoints.
Hormones mediate effects via specific, context-dependent interactions with receptors.	The actions of hormones are dictated based on the specificity of their interactions with receptors, the concentration of the hormone in the blood (including the bioavailable fractions), the abundance and location of the receptor, and other factors.
Hormones act at low circulating concentrations.	Hormones induce significant biological changes at circulating concentrations in the part-per-billion or part-per-trillion range. This is because of the non-linear relationship between hormone concentration and the percentage of receptors that is bound, as well as the non-linear relationship between the percent of occupied receptors and induced biological effect.
Hormones often exhibit non-linear and non-monotonic dose responses.	Like vitamins, nutrients, and other pharmacological agents that interact with receptors, hormones can induce non-monotonic responses via mechanisms including receptor downregulation, receptor desensitization, receptor competition, receptor interactions, endocrine negative feedback loops, and even traditional toxicology mechanisms (i.e. the induction of cytochromes, or cytotoxicity).
The effects of hormones differ based on an individual’s stage of development.	Hormone exposures occurring during periods of development can disrupt the differentiation of cells, tissues, or organs; these so-called organizational effects can be observed long after exposures have ceased because the developmental trajectory of the cell, tissue, organ, or entire organism has been changed. In contrast, activational effects are those that are observed only during the period of exposure, thus when hormone treatment ceases these effects subside; these are typically associated with adult exposures.

Table 2 Principles of toxicology

Principle	Explanation
The relationship between dose and effect is an important feature, with an expectation of linearity or monotonicity. This is sometimes described using the maxim, “the dose makes the poison”.	The toxicological response that is observed is expected to be associated with the amount of exposure, with a monotonic relationship between these two factors. The effect that is examined can be quantal, such as life/death, or continuous, such as organ weight. Further, effects can be characterized as acute (rapidly developing), subacute (less severe than acute), or chronic (progressing at a slow and varying rate). Because a monotonic relationship is assumed across the entire dose range, extrapolation from high to low doses (or vice versa) is possible.
Exposures are characterized based on duration, exposure route, and developmental stage.	Exposures can be separated based on duration (acute, subacute, subchronic, and chronic). They also can be distinguished based on the route of exposure (oral, intravenous, intramuscular, dermal, inhalatory, subcutaneous). They are also sometimes considered separately if exposures occur during development versus in adulthood. These factors can influence the severity and type of the effects observed.
Compounds are studied and understood based on their absorption, distribution in the body, metabolism, and excretion from the body (ADME).	ADME can be influenced by the physiological state of the individual (i.e. young individuals with diminished metabolic capacities) and the physicochemical properties of the compound (i.e. lipophilic versus hydrophilic substances). Metabolism is a process to enhance the removal of compounds from the body (e.g. by changing its water solubility), not necessarily a process that makes chemicals less toxic.
Toxicants can induce general or tissue-specific effects.	Although some toxicants induce general toxic effects, many induce effects that are tissue- and organ-specific. Some of this specificity can be based on the route of exposure (i.e. the lungs are the entry-point for inhalation exposures), the chemical’s accumulation in a specific tissue, the presence of enzymes in specific organs that metabolize a compound to a more active form, the presence of specific receptors in a tissue or organ, and other factors including physiological sensitivities.

chemicals by regulatory agencies as well as expert groups [29, 130–132] and several specific methods have been developed [32, 133, 134].

As mentioned above, risk assessment of chemicals is traditionally founded on the principle of identifying one key toxicity study on which the hazard assessment is based; the key study that is chosen is often an *in vivo* toxicity study conducted according to standardized test guidelines. In contrast, systematic reviews promote a more integrated use of the entire body of evidence that is available and relevant for answering a specific question. A crucial and fundamental principle of systematic review is that it is a structured and clearly documented process that promotes objectivity and transparency. The key characteristics covered in the systematic review process can be summarized as follows [128]:

- a clearly stated set of objectives with pre-defined eligibility criteria for study inclusion
- an explicit, reproducible methodology
- a systematic search that attempts to identify all studies that would meet the eligibility criteria
- an assessment of the validity and/or quality of the findings of the included studies
- a systematic presentation, and synthesis, of the characteristics and findings of the included studies

There are several advantages to using a systematic review methodology in the risk assessment of chemicals in general, and for the identification and assessment of EDCs specifically, e.g.:

- It introduces structure and transparency in the process of identifying and selecting studies to be used as evidence, allowing this process to be reproducible [32, 133, 134]. It especially provides a useful tool for structured identification, evaluation and synthesis of the evidence in cases where there is a large amount of (heterogeneous) data and/or conflicting results between studies.
- It allows for applying expert judgment in a transparent manner, which is critical for promoting objectivity in assessment conclusions for environmental chemicals [29] and has been specifically debated for EDCs [24].
- It potentially allows for better use of academic research studies in regulatory risk assessment, for example by providing structured criteria to evaluate the validity and adequacy of studies for that purpose [133–135].
- The practice of identifying one (possibly limited) study as key evidence on which to base risk assessment conclusions, and subsequent decision making, can be avoided.
- Jointly considering available studies investigating the same or related endpoints may improve overall confidence

in the data, especially if individual studies have weaknesses [53].

- It provides a structured basis for integrating evidence from several streams of evidence (e.g. epidemiological, experimental in vivo and in vitro data), potentially improving confidence in the evidence if several streams support the same conclusions.

3.1 Assessment of study quality

A critical step in a systematic review is the evaluation of the quality or validity of individual studies that have been identified and included. Many different terms are used to describe study “quality” by different expert groups and authorities and in different contexts, rendering discussions sometimes confusing. For the purposes of regulatory risk assessment, the term study adequacy is commonly used, particularly for toxicity studies. Adequacy in this context refers to both the reliability and the relevance of a study [33, 34, 38].

Reliability is defined as the inherent quality of the study and is tightly linked to the reliability of the methods used, how the results have been interpreted, as well as how both methods and results have been reported. For systematic review methods developed in the clinical field, the term “risk of bias” has often been used to describe similar aspects when evaluating studies [128, 133, 134]. Risk of bias specifically relates to the evaluation of a study’s internal validity, reflecting characteristics in the study design that might introduce a systematic error and affect the magnitude and/or direction of study results. It is therefore a more specific measure than reliability. Relevance, in turn, refers to how appropriate the selected test methods are for evaluating a certain hazard, as well as how relevant the findings are to the risk assessment query in question, e.g. the characterization of the human health risks from a certain chemical. In other words, reliability (and risk of bias) is an intrinsic characteristic of a study, which is constant regardless of the context of study evaluation. In contrast, a study’s relevance depends on the scientific question to be answered and may vary on a case-by-case basis.

The evaluation of study adequacy also relies on how completely the study methods and results are reported. If the research aim, design, performance and results of a study are not well reported, it may not be possible to evaluate or ensure sufficient reliability and relevance for regulatory health risk assessment. The impact of reporting on study reproducibility and reliability, as well as consequences for decision making, have been extensively discussed in the scientific literature and several sets of guidelines for reporting research studies have been proposed recently (e.g. [135–137]).

A number of methods for evaluating primarily study reliability are available and have been discussed elsewhere [135, 138]. Some approaches for study evaluation are summarized below.

3.1.1 The Klimisch approach

Regulatory agencies and organizations, such as the European Chemicals Agency (ECHA), the US Environmental Protection Agency (EPA) and the Organisation for Economic Co-operation and Development (OECD), commonly promote the Klimisch approach for evaluating the “quality” of toxicity data [139]. This approach entails sorting available studies into four categories: 1) “Reliable without restrictions”, 2) “Reliable with restrictions”, 3) “Not reliable” and 4) “Not assignable”. However, a requirement for judging a study as category 1 is that it should comply with standardized test guidelines, such as OECD test guidelines or national standards, or be “comparable to a guideline study” [139]. Academic research studies are therefore likely to be categorized as “reliable with restrictions” at best, or “not reliable”. In practice, this means that if standardized studies are available they will be preferred over academic research studies in risk assessment, especially in cases where there are conflicting results. Given that there is evidence that standardized testing batteries may be inadequate to identify and assess EDCs [14, 40, 45, 140], this is a severe limitation rendering the Klimisch approach, as such, unsuitable for use when evaluating studies for systematic review of EDCs. An additional limitation to the Klimisch approach is that it provides no detailed criteria and very little guidance (other than calling for test guideline compliance) for evaluating study quality or for the categorization of studies. As a result, the evaluation of a study’s quality is likely to rely heavily on the application of expert judgment that is not transparent and will probably vary greatly between assessors [138].

3.1.2 Recently developed approaches for study evaluation

More recent approaches aim to make the study evaluation process more systematic and transparent, e.g. by providing detailed pre-defined criteria to evaluate the reliability and relevance of individual studies. New approaches are also moving away from awarding studies conducted according to standardized guidelines with higher scores for reliability by default, reflecting the attitude that this might not be appropriate and that academic research studies can be just as reliable as tests performed under strict implementation of standardized test guidelines.

Schneider et al. [141] developed the ToxRTool, a software-based tool intended to facilitate the reliability categorization of in vitro and in vivo studies according to the Klimisch approach. The ToxRTool assesses several criteria for reliability and then calculates an overall quantitative measure of reliability (i.e. qualitative assessments are converted to a quantitative score), which is then translated into Klimisch category 1–3. However, the assessor has the option to assign a different Klimisch category than the one provided by the tool based

on expert judgment. Although there is no specific reliability criteria for identifying standard guideline studies, this aspect is included in the items to be considered when reflecting on study relevance. The ToxRTool is available online at <https://eurl-ecvam.jrc.ec.europa.eu/about-ecvam/archive-publications/toxrtool>.

Science in Risk Assessment and Policy (SciRAP) is another initiative to develop a structured approach for the evaluation of study reliability and relevance [135, 142]. SciRAP provides criteria for the evaluation of *in vivo* toxicity and ecotoxicity studies; additional criteria for the evaluation of *in vitro* studies are under development. SciRAP also includes a web-based tool intended to help the assessor in the application of the criteria. These resources are freely available online at www.scirap.org. In contrast to the ToxRTool, the SciRAP method is based on a qualitative evaluation rather than a numerical score for reliability. Quantitative scores are avoided in the SciRAP method because numerical values can be argued to imply a level of scientific certainty in the evaluation that may be misleading. The SciRAP method also emphasizes the evaluation of study relevance in addition to reliability, illustrating that studies that are judged to be of low reliability may still be used as supporting evidence in risk assessment if the results are of very high relevance. In addition, guidelines for researchers on how to report studies to better meet the requirements of regulatory risk assessment have been developed and are available as checklists on the SciRAP website.

Recently there have been different initiatives to develop systematic review approaches for environmental chemicals based on existing methods. The National Toxicology Program (NTP) Office of Health Assessment and Translation (OHAT) approach was developed for the purpose of assessing the hazards and risks from environmental chemicals to produce NTP monographs and state of the science reports [133]. Another approach, the Navigation Guide, was initiated by the Program on Reproductive Health and the Environment at University of California San Francisco [134]. Primarily based on methods used for clinical medicine systematic reviews, the Navigation Guide was specifically developed as a systematic and transparent methodology to evaluate the quality of evidence and strength of recommendations about the relationship between environmental chemicals and reproductive health. In Europe, the European Food Safety Authority (EFSA) has also developed and published guidance for application of systematic review methodology to food and feed safety assessments to support decision making [32].

Study evaluation criteria have also been developed within the OHAT and Navigation Guide methods for systematic review [133, 134]. Both of these methods focus on the evaluation of internal validity of studies using risk of bias as a measure. Different “domains” of risk of bias are evaluated

including selection, performance, and detection bias, using specific questions, resulting in a qualitative assessment of the study. The OHAT Risk of Bias tool and handbook is available online at <http://ntp.niehs.nih.gov/pubhealth/hat/noms/index-2.html#Systematic-Review-Methods>. EFSA does not provide specific criteria for the evaluation of individual studies but refers to other available methods.

It should specifically be noted that the approaches developed within SciRAP, OHAT and the Navigation Guide do not, by default, attribute greater validity or adequacy to studies conducted in compliance with standardized test guidelines.

3.2 Systematic review methods for environmental chemicals depart from the field of medicine

Decision making in the fields of medicine and health care are chiefly based on human clinical trials. This is reflected in the approaches for systematic review in these fields, such as the GRADE and Cochrane approaches [127, 128], which were primarily developed for the evaluation and synthesis of data from randomized clinical trials. In contrast, drawing conclusions in the field of environmental health, including assessments that aim to characterize the potential health effects from EDC exposures, will be based on different types of studies such as observational epidemiology (human and wildlife) and experimental *in vivo* and *in vitro* studies. This heterogeneity in data provides an additional challenge, i.e. combining evidence from different “streams” (epidemiological, *in vivo*, *in vitro*, etc.) of evidence. Exposure scenarios will also be more complex for environmental chemicals considering that these exposures can be ubiquitous in populations and often involuntary. These issues have to be addressed in systematic review methods developed for application to environmental health questions.

4 Special features of systematic review for EDCs

As described above, systematic review methods were originally developed in the clinical field to improve assessment of pharmaceutical and other medical interventions. In contrast, exposures to environmental chemicals, including EDCs, are almost always involuntary, exposure scenarios are more complex, and there can be difficulty identifying and documenting unexposed populations. The question that has been raised is whether EDCs should be treated differently from other environmental chemicals, i.e. if specific considerations need to be included in risk assessment of EDCs [25].

4.1 Evaluating quality of EDC studies

A number of study design issues have been raised that should be considered when evaluating the quality of experimental EDC studies [41, 43, 118, 140, 143–148]. Some important features

include the use of sensitive species and strains, inclusion of sufficient numbers of doses over an appropriate dose span, appropriate routes of exposure, and inclusion of endocrine-sensitive endpoints [8, 12, 14, 80, 149, 150]. A number of groups have highlighted the importance of including appropriate positive and negative controls in EDC studies, and that the failure to include these controls should diminish their quality ratings during systematic review evaluations [40, 118, 146, 147, 149, 151]. In general, the requirements for experimental controls include: 1) To ensure that the experimental system was capable of responding to hormones, there must be a positive (significant) effect of either the test chemical or a known positive control; 2) Not only should the experiment include a positive control, it should include a positive control that shows significant effects at low doses. Hormones are active at very low doses – and the experimental system should be capable of responding at these low doses. If no effects are observed in the low dose range for the positive control, it suggests that the experimental system (or the animal species, or animal strain) is not appropriate for the assessment of hormones or EDCs; 3) Negative controls – treated in exactly the same way as the test group with the exception of the compound administered – are required and should be run concurrently to the test chemical. Negative control groups should account for the variability in the experiment including housing conditions, feed, water, environmental stresses, handling, etc. In a controlled laboratory experiment, the negative control should remain truly unexposed to the compound of interest. This can be challenging for some EDCs which can be found in environmental media, feed, water, etc. Yet, if no effects of the test compound are observed and the negative control group is contaminated (even with low levels of the compound of interest), this may confound data analysis.

While there is no method available that has been specifically developed for the evaluation of EDC studies, the SciRAP, OHAT and Navigation Guide methods include criteria, or can be adjusted, so as to accommodate these issues. SciRAP, for example, includes criteria for reliability evaluations that address a range of issues about the characterization of the test compound and vehicle; the animals, housing conditions and feed; the administration route and frequency of dosing; the number of animals and descriptions of evaluative endpoints; the statistical methods used including consideration of the litter as the statistical unit; and the discussion of the results within the context of the field [135]. Relevance evaluations within SciRAP address the species and strain used; the appropriateness of the timing of exposure; the relevance of the route of exposure to human exposures; and the dose levels selected.

4.2 Definitions of an EDC and how they might constrain systematic reviews in risk assessments

Risk assessors in different agencies use different definitions to delineate EDCs (reviewed in [22]). For example, the US EPA

defines an EDC as “An exogenous agent that interferes with the production, release, transport, metabolism, binding, action, or elimination of natural hormones in the body responsible for the maintenance of homeostasis and the regulation of developmental processes” [4]. The WHO defines an EDC as “An exogenous substance or mixture that alters function(s) of the endocrine system and consequently causes adverse effects in an intact organism, or its progeny, or (sub)populations” [152]. Finally, the Endocrine Society defines an EDC as “an exogenous chemical or mixture of chemicals that interferes with any aspect of hormone action” [14]. These different definitions thus establish different cut-offs for the evidence that is required to determine whether a compound is an EDC. The Endocrine Society definition requires only that a compound is shown to interfere with hormone action whereas the US EPA definition requires that a compound interferes with the action of a hormone that is responsible for homeostasis or regulation of development. The WHO definition sets the highest bar, requiring that a compound alters function of the endocrine system *and* induces an adverse effect.

Considering these different definitions, the systematic review methods that are used to assemble evidence for risk assessments would also likely differ. Using the US EPA’s definition, it is likely sufficient to ask only one question in a systematic review: Does Chemical X alter [specific hormone action]? In contrast, using the WHO definition, two questions are needed to conduct a systematic review: Does Chemical X alter [specific hormone action]? And, do Chemical X exposures induce [adverse health outcome]?

5 Limitations to systematic reviews

5.1 Using systematic review methods for broad questions

To date, a limited number of systematic reviews have been conducted for EDCs and published in the peer-reviewed literature (see for example [153–160]); these reviews have largely focused on chemical-disease dyads (i.e. Is chemical X associated with disease Y?). Yet, ‘state of science’ reviews are typically tasked with addressing much broader questions, such as, “Is there sufficient evidence linking low dose EDC exposures to (any) human diseases?” or “Are EDC exposures associated with adverse outcomes in wildlife populations?” Considering this need, the application of a systematic framework to address broad questions has not yet been attempted.

How would one go about conducting a broad systematic review like the kind required in a ‘state of science’ evaluation? If keywords like ‘endocrine disruptor’ were reliably used in studies of EDCs, it would be easier to identify and evaluate a large body of evidence. Unfortunately, terms like “endocrine disruptor” reveal only a small portion of the entire relevant literature. Further, there are political and economic interests in

the avoidance of a label like ‘endocrine disruptor’ [21, 161, 162]; acknowledgement that a compound is an EDC would likely have unwanted consequences for its producer. Instead, algorithms might search for relevant literature on a chemical-by-chemical basis (e.g. bisphenol A or BPA or bisphenol-A, or CAS 80-05-7), and these search results could be combed to identify any study showing effects on *any* biological endpoint. Once collected, the endpoints could be sorted and categorized for further evaluation, and a new search would commence on the next chemical, repeating until a long list of chemical searches has been completed. Alternatively, search algorithms could be used to identify all literature relevant to a single biological endpoint (e.g. anogenital distance or AGD or anogenital index), and these results could be evaluated individually to separate and assess all studies that included known or possible EDCs. This again would be repeated until all relevant diseases (and endpoints) were searched and the resulting literature was evaluated.

Clearly, conducting systematic reviews for very broad questions will require extensive resources and improvements in study indexing and evaluation. Perhaps the first step toward conducting a ‘systematic state of science’ evaluation will involve multiple complex but targeted questions and subsequent systematic reviews. The 2013 WHO/UNEP report and 2015 EDC-2 publication each broke their evaluations down by categories of health endpoints (i.e. female reproduction, endocrine cancers, metabolic diseases, etc.) [9, 10]. Systematic reviews that start with these types of questions (i.e. is a group of chemicals with similar properties associated with a group of diseases with a similar etiology?) will likely be a starting point for future systematic reviews.

5.2 Does this new focus on systematic reviews signal “the death” of narrative reviews?

Increasingly, journals are requiring authors to describe the methods they used to select and evaluate the literature included in review articles. Some journals require that authors “comprehensively review” all relevant studies including those that are not “consistent with expectations” or the authors’ *a priori* hypotheses [163]. Others require that authors follow a checklist and flow diagram to document how studies were selected for inclusion [164]. These policies suggest a shifting away from narrative reviews (where authors select papers for discussion without describing how specific studies were selected and others were ignored). Yet, the narrative review clearly has a role to play in providing historical views of a field, describing highlights from a broader scientific field of study, and providing experts’ “intuitive, experiential and explicit perspectives” [165]. Although some journals have clearly shifted toward favouring, or even requiring, systematic reviews, others continue to value the narrative review.

6 Conclusions

In the last few years, a number of large reviews and consensus statements including a large review of the EDC literature published in 2015 by The Endocrine Society have concluded that there is sufficient evidence that EDCs can affect the health of humans and wildlife [9–11, 13–15]. Many of these reviews have evaluated hundreds or thousands of publications to draw these conclusions. Yet, these reports have been criticized for using narrative rather than systematic review methods, with some criticisms suggesting that the use of systematic review methods would have decreased the strength of the conclusions reached. It is worth noting that one such critique [23] identified a small number of studies that were not included in the 2013 WHO/UNEP report, but the authors of the WHO/UNEP report subsequently determined that the inclusion of these studies would not have changed the evaluation considering they were relevant to a body of evidence that was already rated as relatively weak [24].

Methods for conducting systematic reviews for decision making in the field of environmental health sciences are just beginning to be developed. Most of the methods used to date focus on data derived from a single ‘stream’ of evidence (i.e. human epidemiology *or* controlled laboratory animal studies). Good systematic review methodologies will need to integrate the evidence from different evidence streams after their individual evaluation. Systematic reviews for EDCs also need to consider particular study design features that are essential for the assessment of chemicals that interfere with hormones including the sensitivity of the endocrine system to low doses, the possibility of non-monotonic responses, endpoints that are reflective of endocrine diseases rather than general toxicity, the importance of timing and dose duration, concerns about contamination, and strain/species differences in sensitivity. Many of these features are likely to apply to non-EDC environmental chemicals, so the methods developed for the systematic review of EDCs may be more broadly useful for the field of environmental toxicology. Finally, we have discussed the ways that defining an EDC could influence how a systematic review is conducted.

Another important part of this discussion that should be considered is whether the mechanism by which a compound acts must be understood before public health protective actions are taken. For an environmental chemical with strong evidence of an association between exposures and a disease of concern, or strong experimental studies showing causal relationships between exposures and disease, understanding the full mechanistic pathway from altered hormone action to disease onset should not be obligatory. Just as Bradford Hill did not intend for his ‘viewpoints’ on causal relationships to be *required* for action to be taken, evidence specific to

endocrine disrupting mechanisms should not be required for an appropriate response to be initiated by risk assessors and risk managers.

Acknowledgments LNV was supported by start-up funds from the University of Massachusetts and Award Number K22ES025811 from the National Institute of Environmental Health Sciences of the National Institutes of Health. The content of this manuscript is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The authors thank many colleagues for helpful discussions about systematic reviews and EDCs.

Compliance with ethical standards

Conflict of interest LNV has been reimbursed for travel expenses by numerous organizations including SweTox, Israel Environment Fund, the Mexican Endocrine Society, Advancing Green Chemistry, ShiftCon, US EPA, CropLife America, the World Federation of Scientists, BeautyCounter, and many universities, to speak about endocrine disrupting chemicals.

AB has no conflicts of interest to declare.

References

1. US CDC. National report on human exposure to environmental chemicals. 2015. <http://www.cdc.gov/exposurereport/faq.html>. Accessed 21 November 2015.
2. FDA US. Endocrine disruptor knowledge base. 2010. <http://www.fda.gov/ScienceResearch/BioinformaticsTools/EndocrineDisruptorKnowledgebase/default.htm>. Accessed August 20, 2012.
3. TEDX. TEDX list of potential endocrine disruptors. 2015. <http://endocrinedisruption.org/endocrine-disruption/tedx-list-of-potential-endocrine-disruptors/overview>. Accessed 21 November 2015.
4. Kavlock RJ, Daston GP, DeRosa C, Fenner-Crisp P, Gray LE, Kaattari S, et al. Research needs for the risk assessment of health and environmental effects of endocrine disruptors: a report of the U.S. EPA-sponsored workshop. *Environ Health Perspect*. 1996;104(Suppl 4):715–40.
5. Myers JP, Zoeller RT, FS v S. A clash of old and new scientific concepts in toxicity, with important implications for public health. *Environ Health Perspect*. 2009;117(11):1652–5.
6. Molander L, Ruden C. Narrow-and-sharp or broad-and-blunt—regulations of hazardous chemicals in consumer products in the European Union. *Regul Toxicol Pharmacol*. 2012;62(3):523–31.
7. Bergman A, Heindel JJ, Kasten T, Kidd KA, Jobling S, Neira M, et al. The impact of endocrine disruption: a consensus statement on the state of the science. *Environ Health Perspect*. 2013;121(4):A104–6.
8. Vandenberg LN, Luthi D, Quinerly D. Plastic bodies in a plastic world: multi-disciplinary approaches to study endocrine disrupting chemicals. *J Clean Prod*. 2015. doi:10.1016/j.jclepro.2015.01.071.
9. Bergman Å, Heindel J, Jobling S, Kidd K, Zoeller R, eds. The State-of-the-Science of Endocrine Disrupting Chemicals – 2012. WHO (World Health Organization)/UNEP (United Nations Environment Programme). available from: http://www.who.int/iris/bitstream/10665/78101/1/9789241505031_eng.pdf. 2013.
10. Gore AC, Chappell VA, Fenton SE, Flaws JA, Nadal A, Prins GS, et al. EDC-2: the endocrine society's second scientific statement on endocrine-disrupting chemicals. *Endocr Rev*. 2015;36(6):E1–150.
11. Diamanti-Kandarakis E, Bourguignon JP, Guidice LC, Hauser R, Prins GS, Soto AM, et al. Endocrine-disrupting chemical: an endocrine society scientific statement. *Endocr Rev*. 2009;30:293–342.
12. Gore AC, Chappell VA, Fenton SE, Flaws JA, Nadal A, Prins GS, et al. Executive summary to EDC-2: the endocrine society's second scientific statement on endocrine-disrupting chemicals. *Endocr Rev*. 2015;36(6):593–602.
13. Kortenkamp A, Martin O, Faust M, Evans R, McKinlay R, Orton F, et al. State of the Art Assessment of Endocrine Disruptors, Final Report. Brussels: European Commission; 2011. p. 442.
14. Zoeller RT, Brown TR, Doan LL, Gore AC, Skakkebaek NE, Soto AM, et al. Endocrine-disrupting chemicals and public health protection: a statement of principles from the endocrine society. *Endocrinology*. 2012;153(9):4097–110.
15. Vandenberg LN, Colborn T, Hayes TB, Heindel JJ, Jr, Jacobs DR, Lee DH, et al. Hormones and endocrine-disrupting chemicals: low-dose effects and nonmonotonic dose responses. *Endocr Rev*. 2012;33(3):378–455.
16. US EPA. State of the science evaluation: nonmonotonic dose responses as they apply to estrogen, androgen, and thyroid pathways and EPA testing and assessment procedures. 2013.
17. Rhomberg LR, Goodman JE, Foster WG, Borgert CJ, Van Der Kraak G. A critique of the European commission document, "state of the art assessment of endocrine disrupters". *Crit Rev Toxicol*. 2012;42(6):465–73.
18. Kortenkamp A, Martin O, Evans R, Orton F, McKinlay R, Rosivatz E, et al. Response to a critique of the European commission document, "state of the art assessment of endocrine disrupters" by rhomberg and colleagues—letter to the editor. *Crit Rev Toxicol*. 2012;42(9):787–9 author reply 90–1.
19. Dietrich DR, Aulock SV, Marquardt H, Blaauboer B, Dekant W, Kehrer J, et al. Scientifically unfounded precaution drives European commission's recommendations on EDC regulation, while defying common sense, well-established science and risk assessment principles. *Chem Biol Interact*. 2013;205(1):A1–5.
20. Gore AC, Balthazart J, Bikle D, Carpenter DO, Crews D, Czernichow P, et al. Policy decisions on endocrine disruptors should be based on science across disciplines: a response to Dietrich et al. *Endocrinology*. 2013;154(11):3957–60.
21. Bergman A, Andersson AM, Becher G, van den Berg M, Blumberg B, Bjerregaard P, et al. Science and policy on endocrine disruptors must not be mixed: a reply to a "common sense" intervention by toxicology journal editors. *Environ Heal*. 2013;12:69.
22. Zoeller RT, Bergman A, Becher G, Bjerregaard P, Bornman R, Brandt I, et al. A path forward in the debate over health impacts of endocrine disrupting chemicals. *Environ Heal*. 2014;13(1):118.
23. Lamb JC, Boffetta P, Foster WG, Goodman JE, Hentz KL, Rhomberg LR, et al. Critical comments on the WHO-UNEP state of the science of endocrine disrupting chemicals - 2012. *Regul Toxicol Pharmacol*. 2014;69(1):22–40.
24. Bergman A, Becher G, Blumberg B, Bjerregaard P, Bornman R, Brandt I, et al. Manufacturing doubt about endocrine disrupter science - a rebuttal of industry-sponsored critical comments on the UNEP/WHO report "state of the science of endocrine disrupting chemicals 2012". *Regul Toxicol Pharmacol*. 2015;73(3):1007–17.
25. Lamb JC, Boffetta P, Foster WG, Goodman JE, Hentz KL, Rhomberg LR, et al. Comments on the opinions published by bergman et al. (2015) on critical comments on the WHO-UNEP

- state of the science of endocrine disrupting chemicals (Iamb et al. 2014). *Regul Toxicol Pharmacol.* 2015;73(3):754–7.
26. Michaels D Doubt is their product. *Sci Am.* 2005;292(6):96–101.
 27. Michaels D Manufactured uncertainty: protecting public health in the age of contested science and product defense. *Ann N Y Acad Sci.* 2006;1076:149–62. doi:10.1196/annals.1371.058.
 28. Rhomberg LR, Goodman JE. Low-dose effects and nonmonotonic dose-responses of endocrine disrupting chemicals: has the case been made? *Regul Toxicol Pharmacol.* 2012;64(1):130–3.
 29. Scientific Committee on Emerging and Newly Identified Health Risks (SCENIHR). Memorandum on the use of the scientific literature for human health risk assessment purposes - weighing of evidence and expression of uncertainty. available from: http://ec.europa.eu/health/scientific_committees/emerging/docs/scenih_r_s_001.pdf 2012.
 30. Weed DL. Weight of evidence: a review of concepts and methods. *Risk Anal.* 2005;25(6):1545–57.
 31. (ECHA) ECA. Evaluation under REACH, Progress Report 2012. available from: www.echa.europa.eu. 2013.
 32. EFSA. Guidance of EFSA Application of systematic review methodology to food and feed safety assessments to support decision making. *EFSA J.* 2010;8:90.
 33. OECD. Manual for the assessment of chemicals. Chapter 3: Data evaluation. Available from: <http://www.oecd.org/2005>.
 34. US EPA. Determining the adequacy of existing data. Available from: <http://www.epa.gov/quality> 1999.
 35. Silbergeld E, Scherer RW. Evidence-based toxicology: strait is the gate, but the road is worth taking. *ALTEX.* 2013;30(1):67–73.
 36. Tyl RW. In honor of the teratology society's 50th anniversary: the role of teratology society members in the development and evolution of in vivo developmental toxicity test guidelines. *Birth Defects Res (Part C).* 2010;90:99–102.
 37. Tyl RW. Basic exploratory research versus guideline-compliant studies used for hazard evaluation and risk assessment: bisphenol a as a case study. *Environ Health Perspect.* 2009;117(11):1644–51.
 38. European Chemicals Agency (ECHA). Guidance on information requirements and chemical safety assessment. Part B: Hazard assessment. Available from: http://echa.europa.eu/documents/10162/13643/information_requirements_part_b_en.pdf. 2011.
 39. Schug TT, Abagyan R, Blumberg B, Collins TJ, Crews D, DeFur PL, et al. Designing endocrine disruption out of the next generation of chemicals. *Green Chem.* 2013;15:181–98.
 40. Myers JP, FS v S, BT A, Arizono K, Belcher S, Colborn T, et al. Why public health agencies cannot depend upon 'good laboratory practices' as a criterion for selecting data: the case of bisphenol-a. *Environ Health Perspect.* 2009;117(3):309–15.
 41. Vandenberg LN, Colborn T, Hayes TB, Heindel JJ, Jacobs DR, Lee DH, et al. Regulatory decisions on endocrine disrupting chemicals should be based on the principles of endocrinology. *Reprod Toxicol.* 2013;38C:1–15.
 42. FS v S, BT A, SM B, LS B, DA C, Eriksen M, et al. Chapel hill bisphenol A expert panel consensus statement: integration of mechanisms, effects in animals and potential to impact human health at current levels of exposure. *Reprod Toxicol.* 2007;24(2):131–8.
 43. FS v S, JP M. Good laboratory practices are not synonymous with good scientific practices, accurate reporting, or valid data. *Environ Health Perspect.* 2010;118(2):A60.
 44. Gore AC. Editorial: an international riposte to naysayers of endocrine-disrupting chemicals. *Endocrinology.* 2013;154(11):3955–6.
 45. Gore AC, Heindel JJ, Zoeller RT. Endocrine disruption for endocrinologists (and others). *Endocrinology.* 2006;147(Suppl 6):S1–3.
 46. Kanno J, Onyon L, Haseman J, Fenner-Crisp P, Ashby J, Owens W. The OECD program to validate the rat uterotrophic bioassay to screen compounds for in vivo estrogenic responses: phase 1. *Environ Health Perspect.* 2001;109:785–94.
 47. Kanno J, Onyon L, Peddada S, Ashby J, Jacob E, Owens W. The OECD Program to validate the rat uterotrophic bioassay. Phase 2: coded single-dose studies. *Environ Health Perspect.* 2003;111(12):1550–8.
 48. Kanno J, Onyon L, Peddada S, Ashby J, Jacob E, Owens W. The OECD Program to validate the rat uterotrophic bioassay. Phase 2: dose-response studies. *Environ Health Perspect.* 2003;111(12):1530–49.
 49. Birnbaum LS, Bucher JR, Collman GW, Zeldin DC, Johnson AF, Schug TT, et al. Consortium-based science: the NIEHS's multi-pronged, collaborative approach to assessing the health effects of bisphenol a. *Environ Health Perspect.* 2012;120(12):1640–4.
 50. Birnbaum LS. State of the science of endocrine disruptors. *Environ Health Perspect.* 2013;121(4):A107.
 51. Catanese MC, Suvorov A, Vandenberg LN. Beyond a means of exposure: a new view of the mother in toxicology research. *Toxicol Res.* 2015;4:592–612.
 52. Slovic P, Malmfors T, Mertz CK, Neil N, Purchase IFH. Evaluating chemical risks: results of a survey of the British toxicology society. *Hum Exp Toxicol.* 1997;16(6):289–304.
 53. Krimsky S The weight of scientific evidence in policy and law. *Am J Public Health.* 2005;95(Suppl 1):S129–36.
 54. Linkov I, Loney D, Cormier S, Satterstrom FK, Bridges T. Weight-of-evidence evaluation in environmental assessment: review of qualitative and quantitative approaches. *Sci Total Environ.* 2009;407:5199–205.
 55. Doull J, Rozman KK, Lowe MC. Hazard evaluation in risk assessment: whatever happened to sound scientific judgment and weight of evidence? *Drug Metab Rev.* 1996;28(1–2):285–99.
 56. Woodruff TJ, Zeise L, Axelrad DA, Guyton KZ, Janssen S, Miller M, et al. Meeting report: moving upstream-evaluating adverse upstream end points for improved risk assessment and decision-making. *Environ Health Perspect.* 2008;116(11):1568–75.
 57. Alcock RE, Macgillivray BH, Busby JS. Understanding the mismatch between the demands of risk assessment and practice of scientists—the case of deca-BDE. *Environ Int.* 2011;37(1):216–25.
 58. EFSA. Opinion of the scientific panel on food additives, flavourings, processing aids and materials in contact with food (AFC) related to 2,2-BIS(4-HYDROXYPHENYL)PROPANE. *The EFSA Journal.* 2006;428:1–75.
 59. Hengstler JG, Foth H, Gebel T, Kramer PJ, Lilienblum W, Schweinfurth H, et al. Critical evaluation of key evidence on the human health hazards of exposure to bisphenol a. *Crit Rev Toxicol.* 2011;41(4):263–91.
 60. NTP. NTP-CERHR monograph on the potential human reproductive and developmental effects of bisphenol A. In: NIH Publication No 08–5994. available from: <https://ntp.niehs.nih.gov/ntp/ohat/bisphenol/bisphenol.pdf>. 2008.
 61. Lewis RW. Risk assessment of 'endocrine substances': guidance on identifying endocrine disruptors. *Toxicol Lett.* 2013;223(3):287–90.
 62. Chen CW. Assessment of endocrine disruptors: approaches, issues, and uncertainties. *Folia Histochem Cytobiol.* 2001;39(Suppl 2):20–3.
 63. EFSA. Use of the benchmark dose approach in risk assessment. *EFSA J.* 2009;1150:1–72.
 64. US EPA. Benchmark dose technical guidance. available from: <http://www2.epa.gov/risk/benchmark-dose-technical-guidance> 2012.
 65. Hass U, Christiansen S, Axelstad M, Sorensen KD, Boberg J. Input for the REACH-review in 2013 on endocrine disruptors. available from: <http://www.mst.dk/NR/rdonlyres/54DB4583-B01D-45D6-AA99-28ED75A5C0E4/154979/ReachreviewrapportFINAL21March.pdf> 2013.

66. Slob W Thresholds in toxicology and risk assessment. *Int J Toxicol*. 1999;18:259–68.
67. Rajapakse N, Silva E, Kortenkamp A. Combining xenoestrogens at levels below individual no-observed-effect concentrations dramatically enhances steroid hormone activity. *Environ Health Perspect*. 2002;110:917–21.
68. Welshons WV, Thayer KA, Judy BM, Taylor JA, Curran EM, FS v S. Large effects from small exposures: I. Mechanisms for endocrine-disrupting chemicals with estrogenic activity. *Environ Health Perspect*. 2003;111:994–1006.
69. Ryan BC, Vandenberg JG. Intrauterine position effects. *Neurosci Biobehav Rev*. 2002;26:665–78.
70. Heindel JJ, Balbus J, Birnbaum L, Brune-Drisse MN, Grandjean P, Gray K, et al. Developmental origins of health and disease: integrating environmental influences. *Endocrinology*. 2015;156(10):3416–21.
71. Silbergeld EK, Flaws JA, Brown KM. Organizational and activational effects of estrogenic endocrine disrupting chemicals. *Cad Saude Publica*. 2002;18(2):495–504.
72. Wallen K The organizational hypothesis: reflections on the 50th anniversary of the publication of phoenix, Goy, gerall, and young (1959). *Horm Behav*. 2009;55(5):561–5.
73. Purchase IF, Auton TR. Thresholds in chemical carcinogenesis. *Regul Toxicol Pharmacol*. 1995;22(3):199–205.
74. Neumann HG. Risk assessment of chemical carcinogens and thresholds. *Crit Rev Toxicol*. 2009;39(6):449–61.
75. Calafat AM, Ye X, Silva MJ, Kuklenyik Z, Needham LL. Human exposure assessment to environmental chemicals using biomonitoring. *Int J Androl*. 2006;29(1):166–71 discussion 81-5.
76. Needham LL, Calafat AM, Barr DB. Assessing developmental toxicant exposures via biomonitoring. *Basic Clin Pharmacol Toxicol*. 2008;102:100–8.
77. Woodruff TJ, Zota AR, Schwartz JM. Environmental chemicals in pregnant women in the United States: NHANES 2003-2004. *Environ Health Perspect*. 2011;119(6):878–85.
78. Munn S, Heindel J. Assessing the risk of exposures to endocrine disrupting chemicals. *Chemosphere*. 2013;93(6):845–6.
79. Beausoleil C, Ormsby JN, Gies A, Hass U, Heindel JJ, Holmer ML, et al. Low dose effects and non-monotonic dose responses for endocrine active chemicals: science to practice workshop: workshop summary. *Chemosphere*. 2013;93(6):847–56.
80. Melnick R, Lucier G, Wolfe M, Hall R, Stancel G, Prins G, et al. Summary of the national toxicology program's report of the endocrine disruptors low-dose peer review. *Environ Health Perspect*. 2002;110(4):427–31.
81. Swan SH, Sathyanarayana S, Barrett ES, Janssen S, Liu F, Nguyen RH, et al. First trimester phthalate exposure and anogenital distance in newborns. *Hum Reprod*. 2015;30(4):963–72.
82. Barrett ES, Parlett LE, Wang C, Drobnis EZ, Redmon JB, Swan SH. Environmental exposure to di-2-ethylhexyl phthalate is associated with low interest in sexual activity in premenopausal women. *Horm Behav*. 2014;66(5):787–92.
83. Lam T, Williams PL, Lee MM, Korrick SA, Birnbaum LS, Burns JS, et al. Prepubertal serum concentrations of organochlorine pesticides and age at sexual maturity in Russian boys. *Environ Health Perspect*. 2015;123(11):1216–21.
84. Braun JM, Chen A, Romano ME, Calafat AM, Webster GM, Yolton K, et al. Prenatal perfluoroalkyl substance exposure and child adiposity at 8 years of age: the HOME study. *Obesity (Silver Spring)*. 2016;24(1):231–7.
85. Webster GM, Rauch SA, Ste Marie N, Mattman A, Lanphear BP, SA V. Cross-Sectional Associations of Serum Perfluoroalkyl Acids and Thyroid Hormones in U.S. Adults: Variation According to TPOAb and Iodine Status (NHANES 2007-2008). *Environ Health Perspect*. 2015. doi:10.1289/ehp.1409589.
86. Engel SM, Bradman A, Wolff MS, Rauh VA, Harley KG, Yang JH, et al. Prenatal organophosphorus pesticide exposure and child neurodevelopment at 24 months: an analysis of four birth cohorts. *Environ Health Perspect*. 2015. doi:10.1289/ehp.1409474.
87. Werner EF, Braun JM, Yolton K, Khoury JC, Lanphear BP. The association between maternal urinary phthalate concentrations and blood pressure in pregnancy: the HOME study. *Environ Health Perspect*. 2015;14:75.
88. Vandenberg LN, Ehrlich S, Belcher SM, Ben-Jonathan N, Dolinoy DC, Hugo ER, et al. Low dose effects of bisphenol a: an integrated review of in vitro, laboratory animal and epidemiology studies. *Endocrine Disruptors*. 2013;1(1):e25078.
89. Birnbaum LS. Environmental chemicals: evaluating low-dose effects. *Environ Health Perspect*. 2012;120(4):A143–4.
90. Acevedo N, Davis B, Schaeberle CM, Sonnenschein C, Soto AM. Perinatally administered bisphenol a as a potential mammary gland carcinogen in rats. *Environ Health Perspect*. 2013;121(9):1040–6.
91. Cabaton NJ, Wadia PR, Rubin BS, Zalko D, Schaeberle CM, Askenase MH, et al. Perinatal exposure to environmentally relevant levels of bisphenol a decreases fertility and fecundity in CD-1 mice. *Environ Health Perspect*. 2011;119(4):547–52.
92. McLachlan JA, Newbold RR, Shah HC, Hogan MD, Dixon RL. Reduced fertility in female mice exposed transplacentally to diethylstilbestrol (DES). *Fertil Steril*. 1982;38:364–71.
93. Hunt PA, Koehler KE, Susiarjo M, Hodges CA, Ilagan A, Voigt RC, et al. Bisphenol A exposure causes meiotic aneuploidy in the female mouse. *Curr Biol*. 2003;13(7):546–53.
94. Prins GS, Ye SH, Birch L, Ho SM, Kannan K. Serum bisphenol a pharmacokinetics and prostate neoplastic responses following oral and subcutaneous exposures in neonatal Sprague-Dawley rats. *Reprod Toxicol*. 2011;31(1):1–9.
95. Jenkins S, Rowell C, Wang J, Lamartiniere CA. Prenatal TCDD exposure predisposes for mammary cancer in rats. *Reprod Toxicol*. 2007;23(3):391–6.
96. Lamartiniere CA, Jenkins S, Betancourt AM, Wang J, Russo J. Exposure to the endocrine disruptor bisphenol a alters susceptibility for mammary cancer. *Horm Mol Biol Clin Investig*. 2011;5(2):45–52.
97. White SS, Stanko JP, Kato K, Calafat AM, Hines EP, Fenton SE. Gestational and chronic low-dose PFOA exposures and mammary gland growth and differentiation in three generations of CD-1 mice. *Environ Health Perspect*. 2011;119(8):1070–6.
98. Vandenberg LN, Maffini MV, Schaeberle CM, Ucci AA, Sonnenschein C, Rubin BS, et al. Perinatal exposure to the xenoestrogen bisphenol-a induces mammary intraductal hyperplasias in adult CD-1 mice. *Reprod Toxicol*. 2008;26:210–9.
99. Sharpe RM, Rivas A, Walker M, McKinnell C, Fisher JS. Effect of neonatal treatment of rats with potent or weak (environmental) oestrogens, or with a GnRH antagonist, on leydig cell development and function through puberty into adulthood. *Int J Androl*. 2003;26(1):26–36.
100. Do RP, Stahlhut RW, Ponzi D, Vom Saal FS, Taylor JA. Non-monotonic dose effects of in utero exposure to di(2-ethylhexyl) phthalate (DEHP) on testicular and serum testosterone and anogenital distance in male mouse fetuses. *Reprod Toxicol*. 2012;34(4):614–21.
101. Zoeller RT, Bansal R, Parris C. Bisphenol-a, an environmental contaminant that acts as a thyroid hormone receptor antagonist in vitro, increases serum thyroxine, and alters RC3/neurogranin expression in the developing rat brain. *Endocrinology*. 2005;146:607–12.
102. Richter CA, Taylor JA, Ruhlen RL, Welshons WV, FS v S. Estradiol and bisphenol a stimulate androgen receptor and estrogen receptor gene expression in fetal mouse prostate mesenchyme cells. *Environ Health Perspect*. 2007;115(6):902–8.

103. Gore AC, Walker DM, Zama AM, Armenti AE, Uzumcu M. Early life exposure to endocrine-disrupting chemicals causes lifelong molecular reprogramming of the hypothalamus and premature reproductive aging. *Mol Endocrinol*. 2011;25(12):2157–68.
104. Patisaul HB, Sullivan AW, Radford ME, Walker DM, Adewale HB, Winnik B, et al. Anxiogenic effects of developmental bisphenol a exposure are associated with gene expression changes in the juvenile rat amygdala and mitigated by soy. *PLoS One*. 2012;7(9):e43890.
105. Rubin BS, Lenkowski JR, Schaeberle CM, Vandenberg LN, Ronsheim PM, Soto AM. Evidence of altered brain sexual differentiation in mice exposed perinatally to low, environmentally relevant levels of bisphenol a. *Endocrinology*. 2006;147(8):3681–91.
106. Palanza P, Morellini F, Parmigiani S, FS v S. Ethological methods to study the effects of maternal exposure to estrogenic endocrine disrupters: a study with methoxychlor. *Neurotoxicol Teratol*. 2002;24(1):55–69.
107. Palanza P, Parmigiani S, Liu H, FS v S. Prenatal exposure to low doses of the estrogenic chemicals diethylstilbestrol and o,p'-DDT alters aggressive behavior of male and female house mice. *Pharmacol Biochem Behav*. 1999;64(4):665–72.
108. Palanza P, Parmigiani S, FS v S. Effects of prenatal exposure to low doses of diethylstilbestrol, o,p'-DDT, and methoxychlor on postnatal growth and neurobehavioral development in male and female mice. *Horm Behav*. 2001;40:252–65.
109. Steinberg RM, Juenger TE, Gore AC. The effects of prenatal PCBs on adult female paced mating reproductive behaviors in rats. *Horm Behav*. 2007;51(3):364–72.
110. NRC. Toxicity-pathway-based risk assessment: preparing for paradigm change. available from: http://www.nap.edu/download.php?record_id=12913# 2010.
111. Ankley GT, Bennett RS, Erickson RJ, Hoff DJ, Hornung MW, Johnson RD, et al. Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment. *Environ Toxicol Chem*. 2010;29(3):730–41.
112. Gee D. Late lessons from early warnings: toward realism and precaution with endocrine-disrupting substances. *Environ Health Perspect*. 2006;114(Suppl 1):152–60.
113. Vandenberg LN, Bowler AG. Non-monotonic dose responses in EDSP tier 1 guideline assays. *Endocrine Disruptors*. 2014;2(1):e964530.
114. Zoeller RT, Vandenberg LN. Assessing dose-response relationships for endocrine disrupting chemicals (EDCs): a focus on non-monotonicity. *Environ Heal*. 2015;14(1):42.
115. Vandenberg LN. Low-dose effects of hormones and endocrine disruptors. *Vitam Horm*. 2014;94:129–65.
116. Vandenberg LN. Non-monotonic dose responses in studies of endocrine disrupting chemicals: bisphenol a as a case study. *Dose-Response*. 2013;12(2):259–76.
117. EFSA. EFSA's 17th Scientific Colloquium on low dose response in toxicology and risk assessment. available from: <http://www.efsa.europa.eu/en/supporting/doc/353e.pdf> 2012.
118. FS v S, BT A, SM B, DA C, Crews D, LC G, et al. Flawed experimental design reveals the need for guidelines requiring appropriate positive controls in endocrine disruption research. *Toxicol Sci*. 2010;115(2):612–3.
119. Phoenix CH, Goy RW, Gerral AA, Young WC. Organizing action of prenatally administered testosterone propionate on the tissues mediating mating behavior in the female guinea pig. *Endocrinology*. 1959;65:369–82.
120. Heindel JJ, Vandenberg LN. Developmental origins of health and disease: a paradigm for understanding disease etiology and prevention. *Curr Opin Pediatr*. 2015;27(2):248–53.
121. Heindel JJ. Role of exposure to environmental chemicals in the developmental basis of reproductive disease and dysfunction. *Semin Reprod Med*. 2006;24(3):168–77.
122. Matthews SJ, McCoy C. Thalidomide: a review of approved and investigational uses. *Clin Ther*. 2003;25(2):342–95.
123. Vargesson N. Thalidomide-induced limb defects: resolving a 50-year-old puzzle. *BioEssays*. 2009;31(12):1327–36.
124. Ruegg J, Penttinen-Damdimopoulou P, Makela S, Pongratz I, Gustafsson JA. Receptors mediating toxicity and their involvement in endocrine disruption. *EXS*. 2009;99:289–323.
125. Ellis-Hutchings RG, Rasoulpour RJ, Terry C, Carney EW, Billington R. Human relevance framework evaluation of a novel rat developmental toxicity mode of action induced by sulfoxaflo. *Crit Rev Toxicol*. 2014;44(Suppl 2):45–62.
126. Tian J, Feng Y, Fu H, Xie HQ, Jiang JX, Zhao B. The aryl hydrocarbon receptor: a key bridging molecule of external and internal chemical signals. *Environ Sci Technol*. 2015;49(16):9518–31.
127. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*. 2008;336(7650):924–6.
128. Higgins J, Green S. *Cochrane Handbook for Systematic Reviews of Interventions*. Version 5.1.0 (updated March 2011). <http://handbook.cochrane.org/> (accessed 3 February 2013). 2011.
129. Eden J, Levit L, Berg A, Morton S. Finding out what works in health care. Institute of Medicine of the National Academies: Standards for Systematic Reviews; 2011.
130. Agerstrand M, Beronius A. Weight of Evidence evaluation and Systematic Review in EU chemical risk assessment: Foundation is laid but guidance is needed. *Environment International*. 2015.
131. US EPA. Framework for human health risk assessment to inform decision making, 2014.
132. Birnbaum LS, Thayer KA, Bucher JR, Wolfe MS. Implementing systematic review at the national toxicology program: status and next steps. *Environ Health Perspect*. 2013;121(4):A108–9.
133. Rooney AA, Boyles AL, Wolfe MS, Bucher JR, Thayer KA. Systematic review and evidence integration for literature-based environmental health science assessments. *Environ Health Perspect*. 2014;122(7):711–8.
134. Woodruff TJ, Sutton P. An evidence-based medicine methodology to bridge the gap between clinical and environmental health sciences. *Health Aff*. 2011;30(5):931–7.
135. Beronius A, Molander L, Ruden C, Hanberg A. Facilitating the use of non-standard in vivo studies in health risk assessment of chemicals: a proposal to improve evaluation criteria and reporting. *J Appl Toxicol*. 2014;34(6):607–17.
136. Kilkenny C, Browne WJ, Cuthill IC, Emerson M, Altman DG. Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research. *PLoS Biol*. 2010;8(6):e1000412.
137. Landis SC, Amara SG, Asadullah K, Austin CP, Blumenstein R, Bradley EW, et al. A call for transparent reporting to optimize the predictive value of preclinical research. *Nature*. 2012;490(7419):187–91.
138. Agerstrand M, Breitholtz M, Ruden C. Comparison of four different methods for reliability evaluation of ecotoxicity data: a case study of non-standard test data used in environmental risk assessment of pharmaceutical substances. *Environ Sci Eur*. 2011;23:17.
139. Klimisch HJ, Andreae M, Tillmann U. A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data. *Regul Toxicol Pharmacol*. 1997;25(1):1–5.
140. Hunt PA, Vandevort CA, Woodruff T, Gerona R. Invalid controls undermine conclusions of FDA studies. *Toxicol Sci*. 2014;141(1):1–2.
141. Schneider K, Schwarz M, Burkholder I, Kopp-Schneider A, Edler L, Kinsner-Ovaskainen A, et al. "ToxRTool", a new tool to assess the reliability of toxicological data. *Toxicol Lett*. 2009;189(2):138–44.
142. Molander L, Agerstrand M, Beronius A, Hanberg A, Ruden C. Science in risk assessment and policy (SciRAP) – an online

- resource for evaluating and reporting in vivo (Eco) toxicity studies. *Hum Ecol Risk Assess.* 2015;21:753–62.
143. Hayes TB. Atrazine has been used safely for 50 years? In: JE E, CA B, CA M, editors. *Wildlife Ecotoxicology: Forensic Approaches.* editors ed. New York, NY: Spring Science + Business Media, LLC; 2011. p. 301–24.
 144. Ruhlen RL, Taylor JA, Mao J, Kirkpatrick J, Welshons WV, FS v S. Choice of animal feed can alter fetal steroid levels and mask developmental effects of endocrine disrupting chemicals. *J Develop Origins Health Disease.* 2011;1–13.
 145. FS v S, Hughes C. An extensive new literature concerning low-dose effects of bisphenol A shows the need for a new risk assessment. *Environ Health Perspect.* 2005;113:926–33.
 146. FS v S, CA R, RR R, SC N, BG T, WV W. The importance of appropriate controls, animal feed, and animal models in interpreting results from low-dose studies of bisphenol. *A Birth Defects Res (Part A).* 2005;73:140–5.
 147. FS v S, WV W. Large effects from small exposures. II. The importance of positive controls in low-dose research on bisphenol. *A Environmental Research.* 2006;100:50–76.
 148. Mesnage R, Defarge N, Rocque LM, Spiroux de vendomois J, seralini GE. Laboratory rodent diets contain toxic levels of environmental contaminants: implications for regulatory tests. *PLoS One.* 2015;10(7):e0128429.
 149. Vandenberg LN, Welshons WV, Vom Saal FS, Toutain PL, Myers JP. Should oral gavage be abandoned in toxicity testing of endocrine disruptors? *Environ Heal.* 2014;13(1):46.
 150. Vandenberg LN, Catanese MC. Casting a wide net for endocrine disruptors. *Chem Biol.* 2014;21(6):705–6.
 151. FS v S, SC N, BG T, WV W. Implications for human health of the extensive bisphenol a literature showing adverse effects at low doses: a response to attempts to mislead the public. *Toxicology.* 2005;212(2–3):244–52 author reply 53-4.
 152. Damstra T, Barlow S, Bergman A. In: Kavlock RJ, van der Kraak G, editors. *Global assessment of the state-of-the-science of endocrine disruptors.* Geneva, Switzerland: World Health Organization; 2002.
 153. WHO. State of the science of endocrine disrupting chemicals - 2012. An assessment of the state of the science of endocrine disruptors prepared by a group of experts for the United Nations Environment Programme (UNEP) and WHO. 2013.
 154. Dechanet C, Anahory T, Mathieu Daude JC, Quantin X, Reyftmann L, Hamamah S, et al. Effects of cigarette smoking on reproduction. *Hum Reprod Update.* 2011;17(1):76–95.
 155. Oppeneer SJ, Robien K. Bisphenol A exposure and associations with obesity among adults: a critical review. *Public Health Nutr.* 2014;1-17.
 156. Kuo CC, Moon K, Thayer KA, Navas-Acien A. Environmental chemicals and type 2 diabetes: an updated systematic review of the epidemiologic evidence. *Current Diabetes rEports.* 2013;13(6):831–49.
 157. Goodman M, Mandel JS, DeSesso JM, Scialli AR. Atrazine and pregnancy outcomes: a systematic review of epidemiologic evidence. *Birth Defects Res B Dev Reprod Toxicol.* 2014;101(3): 215–36.
 158. Salay E, Garabrant D. Polychlorinated biphenyls and thyroid hormones in adults: a systematic review appraisal of epidemiological studies. *Chemosphere.* 2009;74(11):1413–9.
 159. Johnson PI, Sutton P, Atchley DS, Koustas E, Lam J, Sen S, et al. The navigation guide-evidence-based medicine meets environmental health: systematic review of human evidence for PFOA effects on fetal growth. *Environ Health Perspect.* 2014;122(10): 1028–39.
 160. Koustas E, Lam J, Sutton P, Johnson PI, Atchley DS, Sen S, et al. The navigation guide-evidence-based medicine meets environmental health: systematic review of nonhuman evidence for PFOA effects on fetal growth. *Environ Health Perspect.* 2014;122(10):1015–27.
 161. Trasande L, Zoeller RT, Hass U, Kortenkamp A, Grandjean P, Myers JP, et al. Estimating burden and disease costs of exposure to endocrine-disrupting chemicals in the European Union. *J Clin Endocrinol Metab.* 2015;100(4):1245–55.
 162. Mackay B Canada's new toxic hit list called "inadequate". *CMAJ.* 2007;176(4):431–2.
 163. Environmental Health Perspectives. Instructions to Authors. 2015. <http://ehp.niehs.nih.gov/instructions-to-authors/#what>. Accessed 22 November 2015.
 164. PLOS ONE. Best practices in research reporting. 2015. <http://journals.plos.org/plosone/s/best-practices-in-research-reporting>. Accessed 22 November 2015.
 165. Pae CU. Why systematic review rather than narrative review? *Psychiatry Investigation.* 2015;12(3):417–9.