Regulatory Toxicology and Pharmacology 64 (2012) 130–133

Contents lists available at SciVerse ScienceDirect

Regulatory Toxicology and Pharmacology

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

Commentary

Low-dose effects and nonmonotonic dose–responses of endocrine disrupting chemicals: Has the case been made?

Lorenz R. Rhomberg *, Julie E. Goodman

Gradient, Cambridge, MA, USA

ARTICLE INFO

Article history:
Available online 27 June 2012

Keywords:
Endocrine-disrupting chemicals (EDCs)
Low-dose effects
Nonmonotonic dose–response
Weight of evidence

ABSTRACT

Vandenberg et al. (2012) claim that "most if not all [endocrine-disrupting chemicals (EDCs)] are likely to have low-dose effects" and "nonmonotonicity is a common occurrence after exposures to hormones and EDCs in cell culture and animals and across human populations." They present examples as anecdotes without attempting to review all available pertinent data, selectively citing studies without evaluating most of them or examining whether their putative examples are consistent and coherent with other relevant information. They assume that any statistically significant association indicates causation of an adverse effect, and their limited evaluation of specific studies is not done uniformly (*i.e.*, studies with positive results are evaluated differently than those with null results). They also do not evaluate whether exposures in studies are truly "low-dose" and relevant to humans. They propose a number of different nonmonotonic dose–response curves, but do not consider reasons for why they should be expected to apply generally across species. Many of their examples would be – and indeed have been – questioned by many scientists. Overall, Vandenberg et al. put forth many asserted illustrations of their two conclusions without providing sufficient evidence to make the case for either and while overlooking evidence that suggests the contrary.

1. Introduction

In their recent publication, Vandenberg et al. (2012) undertook the formidable task of assessing whether endocrine-disrupting chemicals (EDCs) as a whole challenge the traditional toxicology concept that "the dose makes the poison." Based on their assembly of a number of examples, they conclude that "most if not all EDCs are likely to have low-dose effects" and that "nonmonotonicity is a common occurrence after exposures to hormones and EDCs in cell culture and animals and across human populations." According to the authors, biological effects of concern are to be expected even well below exposure levels traditionally defined as no observed adverse effect levels (NOAELs) with respect to the frank toxicity observed at still higher exposures. They assert that, by failing to explore dose levels below those found to avoid high-dose toxicity, conventional toxicity testing of chemicals misleads us to believe that still lower doses should be regarded as without adverse consequences.

Our comments below are a direct response to this particular review, but it should be noted that they also pertain more broadly to several earlier publications that have argued for the theory of low-dose nonmonotonicity of endocrine-mediated effects (*e.g.*, vom Saal et al., 2007). These earlier arguments have drawn their own responses (*e.g.*, Sharpe, 2010; Goodman et al., 2009).

If the conclusions of Vandenberg et al. were to be accepted as a general proposition, there would be profound consequences for toxicity testing and its interpretation in risk analysis and safety assessment. It is therefore important to evaluate the basis for these claims. Vandenberg et al. base their argument largely on assembling a large set of putative examples of observed low-dose impacts and nonmonotonic dose–response curves (NMDRCs) in endocrine-mediated endpoints. In our view, the validity and interpretation of many of these examples would be – and indeed has been – questioned by many scientists familiar with the data and details of each case. In the present brief critique, we cannot discuss each of these examples, but it will be important to subject the interpretations of Vandenberg et al. to wider scientific analysis before the commonness of such effects at low doses is accepted. Below, we focus our comments on the methods for assembling and interpreting the examples, as well as on the bearing of some generalized mechanisms that Vandenberg et al. put forth in support of their hypothesis that nonmonotonicity is not only plausible, but to be expected for endocrine-mediated effects.

At the outset, it is important to be clear that Vandenberg et al. are not simply making the "no-threshold" argument (that frankly adverse effects seen with high frequency at high doses also appear

* Corresponding author. Address: Gradient, 20 University Road, Cambridge, MA 02138, USA. Fax: +1 617 395 5001.
E-mail address: lrhomberg@gradientcorp.com (L.R. Rhomberg).

0273-2300
http://dx.doi.org/10.1016/j.yrtph.2012.06.015

at smaller frequencies at much lower doses, even when experiments of practical sample sizes cannot detect the low rates of occurrence). It also seems that they are not simply arguing that high-dose testing discovers only "major" and obvious toxicities while missing subtler adverse effects that might have lower thresholds. For chemicals interacting with endocrine-mediated processes, they argue that it should be common for perturbations (of sufficient magnitude to prompt health concern) to occur at substantial frequencies through a range of low exposure levels, even when these same effects have not been observed at higher dose levels.[1]

Although the authors discuss a large list of EDCs and endpoints as examples of observed causal effects at low doses, they do not actually evaluate the case for causality of any particular EDC and endpoint with rigor. They present their examples essentially as anecdotes without an attempt to review all of the available pertinent data, selectively citing studies without evaluating most of them or examining whether the putative examples are consistent and coherent with other relevant information. They assume that any statistically significant association is indicative of causation, and their limited evaluation of specific studies is not done uniformly across studies. As noted above, many of their chosen examples would be – and indeed have been – questioned by many in the scientific community. Vandenberg et al. put forth many asserted illustrations of their two main conclusions (that "most if not all EDCs are likely to have low-dose effects" and that "nonmonotonicity is a common occurrence after exposures to hormones and EDCs in cell culture and animals and across human populations.") without providing sufficient evidence to make the case for either and while overlooking evidence that suggests the contrary.

2. Weight of evidence

As discussed by Vandenberg et al., there are many ways to evaluate the weight of evidence (WoE). Several key aspects are central, including a systematic review of relevant individual studies that accounts for data quality and study reliability; a systematic evaluation of consistency, specificity, and reproducibility of specific outcomes; an articulation and evaluation of hypotheses that bear on available data; and a comparison of how well each hypothesis describes the available data (*e.g.*, Kamrin, 2007; Rhomberg et al., 2010; US EPA, 2011). While Vandenberg et al. list several purported examples of low-dose effects, the majority consist of one study per endpoint. There is no evaluation of the study noted or whether the findings are consistent and coherent with other scientific evidence. This is also true for their case studies. For example, for bisphenol A (BPA)/prostate effects, Vandenberg et al. name one reason why each study with a null result should be ignored, give no reasons why those with positive results should be considered indicative of causation, and do not evaluate whether positive results are consistent with one another and BPA's purported mode of action.

Kamrin (2007) suggested that reproducibility, consistency, and proper conduct of studies are required to support the low-dose hypothesis. Vandenberg et al., citing Kamrin, state they and others agree with these criteria, with caveats. In practice, however, they do not evaluate reproducibility or consistency with any specificity (by comparing dose–response patterns or dose ranges over which nonmonotonic effects are asserted across studies). They argue that a single negative result (or even several studies showing negative results) cannot negate other studies that show adverse effects because "it is more difficult to actually find effects, particularly when using highly sophisticated techniques." In a WoE analysis, one evaluates all evidence – regardless of whether it is positive or negative – using the same criteria and determines whether it indicates causation. One also evaluates the alternative hypothesis that apparent positive results are statistical fluctuations or the result of confounding factors. Vandenberg et al. have not done this for any EDC; further, by excluding negative data for not following some uncommonly used experimental design but retaining positive findings for any experimental design, they create an inherent bias toward positive studies. Using this logic, most potentially contradictory results are excluded from consideration, yet positive findings in almost all studies occur among null results for many endpoints. How can null results be unquestionably ascribable to study design issues but positive ones are unaffected?

Vandenberg et al. place great emphasis on their standards for the proper conduct of studies, but they focus mainly on the use of positive and negative controls and sensitive animal strains. In practice, these criteria are used by Vandenberg et al. largely to dismiss studies containing negative findings that contradict the putative examples of low-dose effects. There is no discussion of other aspects of study conduct, such as how endpoints are measured or the statistical analyses used, that might affect results and their interpretation. When evaluating consistency among studies, Vandenberg et al. argue that every study may not show the same effect if some studies use insensitive species. Although this is true, ascribing a lack of effects to an inherent lack of sensitivity is itself a scientific conclusion that should be subject to a WoE evaluation that considers, among other factors, the historical record of the use of the experimental system to show causation of the effects in question. It is also important to consider whether available historical control data indicate that a strain or species is so prone to developing a particular effect that its response to certain chemicals overestimates human risks. That is, if effects of an agent are so tenuous as to appear only in particularly sensitive strains, it is not clear that such effects should be deemed relevant for humans (*i.e.*, the reasoning for why humans should also be regarded as especially sensitive needs to be evaluated). Similarly, automatically ascribing an observed lack of effects in studies without positive controls to the lack of ability to show such effects (an argument that is frequently raised by low-dose proponents, *e.g.*, see vom Saal and Welshons, 2006) is presumptuous unless couched in a WoE evaluation that considers the historical record of the assay to show effects.

Other experimental design issues that would seem equally critical are not considered. For instance, Vandenberg et al. include a section on the impact of intrauterine position – the placement of individual fetuses in the uterine horns between neighbors of different sexes – on subtle changes in endocrine-modulated effects. Their point is to show that even such small changes in the hormonal environment as wrought by the sex of one's uterine neighbors can produce changes of the kind and magnitude that they also ascribe to chemical exposures. But if this is so, then any chemical study that does not control for intrauterine position is likely to be susceptible to false positives arising from chance differences among dose groups in the distribution of intrauterine positions. Controlling for intrauterine position is rarely done, yet the impact of this on potential false positive results is not considered in the Vandenberg et al. analysis.

3. Exposure

To assess whether effects occur at low doses, one must know what a low dose is. Vandenberg et al. discuss many ways to define

[1] We speak of "perturbation" instead of toxicity because Vandenberg et al. explicitly decline to distinguish between adverse and non-adverse effects; rather, they implicitly presume that all perturbations, being impositions on normal control processes, should be avoided. This approach fails to consider how dose–response patterns for apical effects might differ from those for precursor perturbations and components of the overall mode of action.

this, but the definition that is probably most important is the dose to which humans are exposed. To determine whether an EDC has a low-dose effect, it is critical that it is studied at doses relevant to assessing human risk. For the vast majority of EDCs listed by Vandenberg et al., the relationship between exposures in the experimental studies is not related to human exposures. In other cases, estimates of human exposure are not based on a balanced review of the literature.

For example, Vandenberg et al. claim that human exposure to BPA is 400 μg/kg day based on one citation that did not adhere to the basic principles of pharmacokinetics. Based on all of the available pharmacokinetic data, the US Food and Drug Administration (FDA) recently reported mean dietary intake of BPA from its presence in food-contact articles to be 0.1–0.2 μg/kg day for children and adults and 0.2–0.4 μg/kg day for infants (US FDA, 2009).

Vandenberg et al. acknowledge that exposure route can affect the dose that reaches the target site, but they argue that the route of exposure is not an issue as long as the dose that reaches the target site is applicable to humans. But EDCs given by alternative routes of exposure can lead to concentrations at target sites that are much higher than human exposures; in those instances, investigators are no longer assessing low-dose, but high-dose, effects. As argued at length by Vandenberg et al., one cannot establish whether low-dose effects are likely or even plausible if one only assesses effects at high doses. To date, none of the non-oral rodent studies of BPA have approached human exposures, and, while these can be informative for assessing the overall WoE, they are not informative regarding low-dose exposures.

4. Biological effects

Vandenberg et al. state "any biological effect, whether it is observed to follow linear relationships with administered dose or not, provides conclusive evidence that an EDC has biological activity." This simply presumes that the apparent effects are indeed causal; as we have argued, such a conclusion is not self-evident. It requires evaluation of study-by-study consistency and weighing of contradictory data before being acceptable. Moreover, as discussed by Goodman et al. (2010), adaptive effects are non-adverse biological effects, or perturbations, that reflect homeostasis, or the body's response to environmental stressors to maintain normal biological function. Mere perturbations, such as compensatory effects that occur over a short period of time, transient effects that are not sustained during exposure, reversible effects that cease once exposure stops, and early precursor effects that are distant from the apical effect, may not be adverse. Also, effects that are sustained during exposure but are of low severity and do not lead to any biological functional impairment may not be adverse. Vandenberg et al. have made no efforts to distinguish these types of effects, calling *any* effect indicative of activity, insinuating adversity. While it is true that a non-adverse effect could be a biomarker for an apical effect, in any toxicity study, this must be evaluated. In particular, biological effects that have a homeostatic response should not be considered conclusive evidence of an EDC's activity. This holds for *in utero* exposures as well. Vandenberg et al. argue that the fetus does not have fully developed systems for maintaining homeostasis and will necessarily be particularly sensitive to EDCs, yet there are many instances where this has been shown not to be the case (*e.g.*, BPA administered orally to rats is not found in the fetus, and BPA in the fetus can be metabolized to its inactive form; Doerge et al., 2011). As toxicology becomes more mechanistic, all toxicity endpoints – not just those associated with endocrine perturbation – encounter this issue of how to use observations of early events at lower levels of organization as indicators of potential apical toxicity. To simply decline to distinguish perturbation and modulation from dysfunction dodges this key question and is retrogressive, stepping away from the insights contributed to modern toxicology by considering mode of action.

In addition, the fact that a biological effect is observed, be it adverse or not, does not mean it was caused by an EDC. An exposure is not likely to be causal if a biological effect is not statistically significantly different in exposed and non-exposed study subjects; isolated (occurring in very few study subjects) or independent (occurring in the absence of other effects that are expected *via* the same mode of action); secondary to primary exposure-related effects; observed because of study limitations; or unrelated to the apical effect and not associated with functional impairment (Goodman et al., 2010). Vandenberg et al. assert effects from dozens of EDCs without an evaluation of any of these factors, which often impact the interpretation of results.

5. Nonmonotonic dose–response

Vandenberg et al. state that NMDRCs occur in cell culture, animal, and epidemiology literature and "it is plausible that...NMDRCs are not the exception but should be expected and perhaps even common." Vandenberg et al. have shown that this phenomenon is in the realm of possibility, but have not shown that nonmonotonicity is actually demonstrated in any particular case. If enough parameters are used to model a dose–response function, it will always be possible to fit curves to the variable response patterns observed for some biological endpoints. It is critical, however, to distinguish between reflecting a true biological response and simply over-fitting patterns of study-specific dose-by-dose fluctuations that occur by chance alone. In the case of traditional monotonic dose–response relationships, it is the very constraint of assumed monotonicity that provides the statistical basis for sorting out trends from random statistical fluctuations. If this presumption is abandoned, it places especial importance on being able to repeat the findings, with similar nonmonotonic patterns of dose-dependence, if one is to claim the patterns as real phenomena and not just study-specific patterns of error. Without *a priori* expectations of dose–response curve shape, this is the only basis for establishing generalizable dose–response phenomena empirically. Vandenberg et al. have not shown that statistically significant results (or even trends that are not statistically significant) are observed with any consistent patterns across studies for any particular EDC and endpoint. Although they have put forth possible mechanisms that could generate nonmonotonicity, they have not shown that these mechanisms actually operate for particular EDCs. For instance, studies using estrogen and androgen antagonists can help to demonstrate the role of particular mechanisms and putative outcomes, but these studies have not been brought to bear to establish how specific interactions with the endocrine system bring about certain dose–response phenomena.

A contrasting perspective has recently been articulated in a commentary by Wolff (2011) on epidemiology studies of BPA, for which, like many other EDCs, most exposures are low and ranges small. This leads to low statistical power and, sometimes, specimen contamination can account for some measured biomarker levels. Wolff notes that this can lead to a high noise-to-signal ratio, causing "unusual statistical associations, including inverse relationships." Some studies use a one-time measurement that may be poorly correlated with exposure at some earlier critical time at which the putative causal processes would operate. For these reasons, single instances of a positive association in a study, even if statistically significant, should not be taken as sufficient evidence of causation, as Vandenberg et al. have done uncritically in their evaluation of not only human studies, but experimental studies as well (*e.g.*, Elswick et al., 2000, 2001).

Vandenberg et al. propose a number of different NMDRCs, and the doses at which they purport effects occur vary across EDCs. They do not evaluate whether these dose levels are applicable to human exposures, nor do they consider reasons for why the patterns should be expected to apply generally across species.

6. Mechanisms for nonmonotonicity

Vandenberg et al. review a number of potential mechanisms by which complex biological interactions could generate nonmonotonic relationships. It is noteworthy that the examples they show are all of phenomena that happen at substantial dose levels, much higher than typical human exposures. Moreover, they are largely examples of effects on quantitative continuous endpoints (for which the magnitude of a physiological effect is the outcome), rather than changes in the incidence of all-or-none effects. Such relationships between dose and the changing magnitude of continuous endpoints are sometimes termed dose–effect curves (to distinguish them from dose–response curves that trace the probability of a quantal response as a function of dose). Tying changes in continuous variables to changes in incidence rates of distinct disease states is a challenge for mechanistic toxicology, but the issue is elided in the Vandenberg et al. treatment.

The various mechanisms have in common that the changes in direction of the measured endpoint with increasing dose arise because some countervailing effect (which also increases with dose) acts to extinguish or quash it (by killing the target cells, down-regulating the receptor, or saturating a competing process, *etc.*). These countervailing mechanisms themselves are major effects that would require substantial exposures; moreover, they should cause other observable consequences and physiological disruptions. That is, even though these mechanisms may prevent one key endpoint at a high dose, they should still affect physiology in some observable way. Indeed, they might well be recognized as toxicity in their own right. It is not clear how such mechanisms would be able to operate in a way that produces apparently unaffected individuals at an intermediate dose and yet prompt dysfunction at lower doses.

7. Conclusions

The Vandenberg et al. analysis framework lacks the scientific rigor necessary for an objective evaluation of the extent to which a body of scientific evidence does, or does not, support the NMDR hypothesis. In our view, the case for widespread nonmonotonicity leading to undetected toxicity at low doses has not been made, and indeed cannot be made, simply through assembling selected cases that are presumed to represent causal effects while ignoring contradictory findings.

Conflict of interest statement

This paper was prepared with funding from the American Chemistry Council (ACC). The authors retain sole responsibility for its content. Both authors have done other work funded by ACC (including WoE analyses and testimony to state legislatures and city councils regarding BPA) as part of their work at Gradient, and Lorenz Rhomberg has been a named expert in litigation regarding BPA.

References

Doerge, D.R., Twaddle, N.C., Vanlandingham, M., Brown, R.P., Fisher, J.W., 2011. Distribution of bisphenol A into tissues of adult, neonatal, and fetal Sprague-Dawley rats. Toxicol. Appl. Pharmacol. 255, 261–270.

Elswick, B.A., Welsch, F., Janszen, D.B., 2000. Effect of different sampling designs on outcome of endocrine disruptor studies. Reprod. Toxicol. 14 (4), 359–367.

Elswick, B.A., Miller, F.J., Welsch, F., 2001. Comments to the editor concerning the paper entitled "Reproductive malformation of the male offspring following maternal exposure to estrogenic chemicals" by C. Gupta. Exp. Biol. Med. (Maywood) 226 (2), 74–75.

Goodman, J.E., Dodge, D.G., Bailey, L.A., 2010. A framework for assessing causality and adverse effects in humans with a case study of sulfur dioxide. Regul. Toxicol. Pharmacol. 58, 308–322.

Goodman, J.E., Witorsch, R.J., McConnell, E.E., Sipes, I.G., Slayton, T.M., Yu, C.J., Franz, A.M., Rhomberg, L.R., 2009. Weight-of-evidence evaluation of reproductive and developmental effects of low doses of bisphenol A. Crit. Rev. Toxicol. 39, 1–75.

Kamrin, M.A., 2007. The 'low dose' hypothesis: validity and implications for human risk. Int. J. Toxicol. 26 (1), 13–23.

Rhomberg, L.R., Bailey, L.A., Goodman, J.E., 2010. Hypothesis-based weight of evidence. A tool for evaluating and communicating uncertainties and inconsistencies in the large body of evidence in proposing a carcinogenic mode of action – naphthalene as an example. Crit. Rev. Toxicol. 40, 671–696.

Sharpe, R.M., 2010. Is it time to end concerns over the estrogenic effects of bisphenol A? Toxicol. Sci. 114 (1), 1–4.

US EPA, 2011. Weight of evidence. Evaluating results of EDSP Tier 1 screening to identify the need for Tier 2 testing. Office of Chemical Safety and Pollution Prevention, Endocrine Disruptor Screening Program, Washington, DC.

US Food and Drug Administration (US FDA), 2009. Memorandum to M. Hepp (Division of Food Contact Notifications, Regulatory Group 1) Re: Exposure to Bisphenol A (BPA) for Infants, Toddlers and Adults from the Consumption of Infant Formula, Toddler Food and Adult (Canned) Food, FDA-2010-N-0100-0001. Available at: <http://www.regulations.gov/#!documentDetail;D=FDA-2010-N-0100-0009>.

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

vom Saal, F.S., Welshons, W.V., 2006. Large effects from small exposures. II. The importance of positive controls in low-dose research on bisphenol A. Environ. Res. 100, 50–76.

vom Saal, F.S., Akingbemi, B.T., Belcher, S.M., Birnbaum, L.S., Crain, D.A., Eriksen, M., Farabollini, F., Guillette, L.J., Hauser, R., Heindel, J.J., Ho, S.M., Hunt, P.A., Iguchi, T., Jobling, S., Kanno, J., Keri, R.A., Knudsen, K.E., Laufer, H., LeBlanc, G.A., Marcus, M., McLachlan, J.A., Myers, J.P., Nadal, A., Newbold, R.R., Olea, N., Prins, G.S., Richter, C.A., Rubin, B.S., Sonnenschein, C., Soto, A.M., Talsness, C.E., Vandenbergh, J.G., Vandenberg, L.N., Walser-Kuntz, D.R., Watson, C.S., Welshons, W.V., Wetherill, Y., Zoeller, R.T., 2007. 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. 24, 131–138.

Wolff, M.S., 2011. Correspondence re: Bisphenol A. Chem. Eng. News 89 (29), 4–5.