



Evidence-Based Toxicology: Strait is the Gate, But the Road is Worth Taking

Ellen Silbergeld and Roberta W. Scherer

Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

Summary

The concept of evidence-based toxicology (EBT) was proposed in 2006, but progress since that time has been impeded by differing definitions and goals. This paper describes the parallels and discontinuities between the approach and methods of evidence-based medicine and health care and those proposed for toxicology. The critical element of an evidence-based approach for either discipline is the adoption of unbiased, transparent methodologies during the collection, appraisal, and pooling of evidence. This approach, implemented during the conduct of a systematic review, allows evaluation of the breadth and quality of available evidence. At present, systematic reviews are rarely done in toxicology by regulatory agencies, international organizations, or academic scientists. Adopting an EBT approach will necessitate significant changes in practice as well as attention to distinctive characteristics of toxicological studies, notably their emphasis on identifying harms and their reliance on experimental animal studies. An evidence-based approach does not obviate the role of judgment and values in decision making; its goal is to ensure provision of all available information in a transparent and unbiased manner.

Keywords: evidence-based toxicology, evidence-based medicine/health care, systematic reviews

1 Introduction

The concept of evidence-based toxicology (EBT) has been under discussion for several years (Hoffmann and Hartung, 2006). EBT is about assembling the evidence related to hazards and risks of exposure, or to the evaluation of methodologies for assessing toxicology for the purpose of using this systematically collected evidence during decision making. In this way it is similar to Evidence-based Medicine and Health Care (EBM/HC), which uses evidence derived from randomized controlled trials on which to base healthcare decisions. EBM/HC is defined as the application of *systematically acquired evidence* within the experience and expertise of the clinician, as well as patient values (Sackett et al., 1996). The essential premise is that decisions should be based on the evidence. It is important that the evidence be obtained in a transparent and systematic manner that is clearly described, enabling other investigators to obtain the same evidence. Like EBM, the impetus for EBT clearly is related to the increasingly important role of the discipline of toxicology in decision making related to public health as well as clinical and preclinical sciences. Progress in

EBT has been impeded by differing definitions (Guzelian et al., 2005; Griesinger et al., 2009), both of which advocate the use of methods developed for assessing and using evidence from randomized controlled trials for EBM, an approach that is not feasible for the study of agents suspected of toxicity, as we will discuss below. Efforts also were impeded by a relatively limited focus on the application of evidence-based approaches to the validation and acceptance of alternative methods in applied toxicology (Hartung, 2010).

Evidence-based decision making can be defined as the translation of information into accepted practice using methods that reduce bias and increase confidence (Grimshaw et al., 2006). As in the law, evidence-based methods involve the evaluation of information for its admission into consideration in decision making through the process of applying specified norms and methods. In order to avoid bias, these norms and methods must stand apart from the information under consideration, and their application must be undertaken with complete transparency.

These characteristics differentiate evidence-based approaches from current approaches used in the translation of toxicological studies into decision making by agencies concerned with

Received September 10, 2012; accepted in revised form October 31, 2012.

This paper is an elaboration of presentations by Roberta Scherer and Ellen Silbergeld at the workshop on "Evidence-based Toxicology for the 21st Century: Opportunities and Challenges," held on January 24-25, 2012 at the Environmental Protection Agency campus in Research Triangle Park, NC, USA (see <http://www.ebtox.com>). See Judson et al. (2013), for a written version of another major presentation from the workshop and Stephens et al. (2013), for an overall summary of the workshop.



occupational and environmental health and consumer protection, as we will demonstrate in this paper. In present practice, the identification of relevant primary studies and the norms by which these studies are evaluated in toxicology are largely implicit (the so-called Delphi method). As a result, the process clearly is not transparent and, because of this, it is difficult to avoid or reduce controversies over policy decisions incorporating toxicology. A previous paper commented on the opacity of the Delphi method often used in risk assessment (Silbergeld, 2009), in terms similar to critiques of medical decision making using these methods (Flower et al., 2007).

There is an understandable skepticism on the part of practitioners and experts in a field to the suggestion that the adoption of major changes in practice may be advantageous. This skepticism was expressed in the early days of EBM (Feinstein, 1995; Williams and Garner, 2002; Chalmers, 2005). We acknowledge and respect this natural skepticism in toxicology. This paper makes the case that adoption of evidence-based methods in toxicology may benefit from awareness of the history of evidence-based approaches in medicine and health care (EBM/HC). The goal of this paper is to introduce a consistent vocabulary for EBT and to examine the extent to which our experience in EBM/HC can inform the development of EBT.

At the outset, we recognize that it is reasonable to ask if adopting EBT will increase efficiency and quality of decision making. The history of EBM/HC demonstrates that the evidence-based approach has accomplished these goals in medicine and many health care-related fields (Dickersin and Manheimer, 1998). Moreover, this history shows that a commitment to an evidence-based approach in these fields has stimulated expansion and improvement in the field, specifically through the development of systematic reviews as the instrument for translating information into evidence. Systematic reviews often are considered the highest source of evidence in that primary studies are systematically identified and appraised and the totality of evidence is synthesized. This did not occur without considerable effort. When systematic reviews were initially conducted in medicine in the early '80s, many authors noted that methods associated with conducting systematic reviews were wanting in several areas, including reporting the primary studies, methods for identification and appraisal of the data, and methods for statistical pooling of the data (Mulrow, 1987; Oxman and Guyatt, 1988). The need to develop these approaches was not accepted readily by all practitioners (Chalmers, 2005). Nevertheless, over time, standards were developed through consensus for reporting primary studies (e.g., the CONSORT statement and its extensions¹), for reproducibly searching for these studies (Dickersin et al., 1994), and standardized methods to identify and account for biases in the primary studies (Moher et al., 1996). Also over time, further statistical methods and inferential models were put forward to synthesize similar research efforts. This focus on methods used during the conduct of a systematic review process, in turn, has

led both to greater transparency in reporting primary studies and to an increased focus on the quality of the studies comprising the evidence.

Also of interest to the field of toxicology, the focus on study quality in EBM/HC, in turn, has influenced researchers in relevant fields to improve the quality of their research designs and the rigor of their statistical analyses in order to meet the criteria for inclusion in systematic reviews as well as to support evidence-based strategies. From the perspective of the development of toxicological sciences, this may be one of the most important benefits to consider in adopting EBT.

There is concern that an evidence-based approach introduces rigidity into decision making (Gatchel and McGeary, 2002) and through this may exclude valuable information through the use of scoring systems and meta-analysis. In answer to these concerns, it should be noted that in EBM/HC the evidence provided by transparent systematic reviews provides only one stage of the evidence-based process of application of the evidence. This is not dissimilar to the role of toxicology in decision making as part of the overall process of risk management (NRC, 1994.) Any decisions made in EBM/HC or toxicology must include consideration of other factors, such as cost, feasibility, and the bounds of accepted practice. Thus, in medicine, application of systematically acquired evidence is done taking into account the needs and values of the individual seeking health care (Sackett et al., 1996). Moreover, there is no requirement for evidence-based decision making to employ formal meta-analysis or to use forest plots to express integrated findings.² The use of systematic tools, when appropriate, is an important means of ensuring reproducibility of analysis, as well as the quality of the review, by ensuring comparability in design and conduct across the individual data sources, and, above all, enhanced transparency of conclusions reached in the systematic review.

We argue that toxicologists should consider key lessons learned over the evolution of EBM/HC. First, such transitions are best managed by the community of researchers and practitioners, rather than by imposition from outsiders (such as regulators and other consumers of toxicological evidence). Second, as demonstrated in current practice in EBM/HC, evidence-based methods do not reduce or replace the importance of expert and experienced judgment. Rather, they simply provide the totality of evidence upon which to base those decisions. Third, the process in itself does not generate decisions. Simply put, an evidence-based approach assists the community by providing systematically collected information using clearly described methods that reliably represent the state of relevant knowledge. Thus, this approach assists decision makers in increasing the acceptability of their decisions by ensuring transparency during evidence collection. Fourth, a systematic and transparent approach to collecting and appraising the available evidence in EBM/HC has had a positive influence on researchers in terms of study design and data analysis.

¹ <http://www.consort-statement.org/>

² <http://www.cochrane.org/resources/handbook/index.htm>

2 Toxicology is not medicine or health care

Despite the relevance of understanding the history and experience in EBM/HC, there are characteristics of toxicology and its applications in public health that require more than simple adoption of EBM/HC methodologies. Some of these are related to differences in fundamental objectives. EBM/HC focuses primarily on developing evidence of the efficacy of therapy, together with an emerging focus on the accuracy of diagnostic tests, as well as some focus on etiology, prognosis, and screening. In contrast, the main focus of toxicology is on developing evidence for harms (hazard) and the magnitude or likelihood of harms (risk). Although questions of harm have occasionally been the subject of EBM systematic reviews, as discussed below, many study designs utilized in generating evidence in EBM are not specifically intended to detect or characterize harms. Second, EBM/HC draws almost exclusively upon studies conducted in humans and human populations; toxicology draws primarily upon studies conducted in nonhuman animals and nonanimal models in order to achieve its societal goals of preventing disease and disability. Thus it is important to recognize that adoption of evidence-based approaches for toxicology will require considerable work by the community, as discussed below.

3 Assessing current practice in toxicology

To date, there have been relatively few explorations of the application of evidence-based practices to resolving issues of importance in toxicology. Toxicology has matured in the context of increased demands for its information through the growth of public concerns and regulation in environmental and occupational health. The structure of information needs for decision making in these domains of public health is relatively well defined to include understanding the elements of relevant toxicological studies and the major decision rules into which these elements are to be incorporated. For the purposes of this paper, we focus on those toxicological studies related to defining hazard and quantifying risk; exposure assessment, which is the other element of risk-based decision making, involves other disciplines and methodologies. Hazard and risk are common to the practice of risk assessment and to application of the precautionary principle, which has been advanced as a partial alternative to risk assessment based methods related primarily to reducing the burden of information required for undertaking assessments (Silbergeld et al., 2004).

Current evaluations of toxicological information (from human and nonhuman subjects), almost without exception, have failed to utilize systematic or transparent methods. These limitations are exemplified by a review on lead and cancer by one author of this paper (Silbergeld, 2003) and a review of the carcinogenicity of lead compounds by the International Agency for Research on Cancer (IARC, 2006). Both of these examples are distinguished by lack of transparency such that it is not possible to determine or to replicate the process of identifying

studies or their selection for review. No information was provided on the search strategy or on screening criteria in terms of study quality. Without this information, it is not possible to ascertain the completeness of the review. There is no disclosure of which studies were discarded or why they were discarded. Further, there is no information on why certain studies were emphasized in the discussion. In the case of experimental studies, a similar lack of transparency informed the identification and selection of studies. A recent comment on the failure of IARC monographs to utilize systematic approaches or to cite systematic reviews echoed these same concerns with additional examples (Straif et al., 2012).

In these two examples, the review of epidemiological studies combined cross-sectional, longitudinal, cohort, or secondary analyses without acknowledgement or discussion of heterogeneity, even though it was unlikely that their results could be combined in any meaningful manner. Similarly, the *in vitro* studies were discussed without consideration of study design, dose or *in vitro* concentration, animal strain or cell line. Other sources of heterogeneity were obvious as well. Sometimes studies actually reported on different endpoints. These problems are increased when multiple experimental tests are used to define an endpoint, such as multiple *in vitro* systems and different animal strains (for example, in current US EPA guidelines for developmental neurotoxicity (Crofton et al., 2004) and endocrine disruption (Daston et al., 2003)). When the methods of such studies are so diverse, it may not be appropriate to combine results except in the most general way. Similarly, in EBM/HC studies are not combined if they show either clinical or statistical heterogeneity.

In place of a formal integration of results using clearly described methods (e.g., formal meta-analyses or focused narrative syntheses of the data), these reviews included only tables that summarize selected findings. The only qualified judgments relate to carcinogenicity using EPA or IARC criteria. Even more disturbing than these examples is the practice in some health assessments to base conclusions on only a few or even one study, judged to be the most appropriate or reliable (on nontransparent criteria). Facing two alternative conclusions, one must “choose” which one, if either, to believe. In contrast, a systematic approach uses all the accepted evidence on which to provide a basis for decision making. The concept of a “key study” is contrary to the notion of a systematic review because of its deliberate exclusion of the body of relevant information. This selective practice was followed in a recent NRC review of mercury, in which a nontransparent decision was made to reject one of two large prospective epidemiological studies on early exposures to methyl mercury and neurodevelopmental outcomes (NRC, 2001). Another approach on this same topic utilized a self-described Bayesian “integrative” approach to examine several studies, but no reason was provided for why only some pertinent studies were included (Axelrad et al., 2007). The recent NTP review of lead (2011) moves closer to the practice of systematic reviews as practiced using an evidence-based approach, but it is still a mixture of transparent and nontransparent methods. There are clear statements related to framing



specific questions and to some extent explicating the initial criteria for searching the literature for relevant primary studies, but it fails to present an explicit means by which these studies were identified or evaluated. In addition, as stated in the report, NTP explicitly relied upon other “authoritative sources” (from US government agencies) to identify citations for review, supplemented by some searches of the literature and consultation of experts rather than systematically reviewing all relevant citations. Thus, it is difficult overall to define the methods by which the primary studies were identified or selected, and it is likely to be difficult to replicate the process in an independent exercise. Most importantly, the document does not describe how these study results were integrated to support qualitative judgments based on IARC criteria. Tables in the document are rated as either “supporting” or “not supporting” these qualitative judgments without defining or describing the criteria used to classify a study as supporting or not. Furthermore, the authors appear to have selected which studies are cited in these tables rather than showing all data. Evaluation also involved nontransparent processes such as expert consultation and review by a selected panel. The conclusions were further influenced by the committee review, as well as by the conclusions of the “authoritative sources,” which, as noted above, did not adopt or implement transparency.

4 Why EBT and why now

The need for EBT is arguably driven by several forces: the increased demand for transparency and a stronger scientific basis for decision making in both public and private sectors, as well as longstanding dissatisfaction with the pace and contentious nature of current modes of decision making in public health (EEA, 2001). Examples such as the divergent risk assessments for methyl mercury and bisphenol A in public health policy in the US and the EU (Beronius et al., 2010) do not encourage confidence. Stakeholders with an interest in efficient government and public health should be greatly concerned by the fact that EPA’s evaluation of the human health effects of dioxins took 18 years. How the data used to make these decisions was obtained is neither clear nor replicable. EBT mandates the provision of methods used to develop a set of primary studies which are then used as the evidence for decision making. Clearly, the use of EBT can promote reduction of controversies, as all can obtain exactly the same data on which to base decisions; the methods used to obtain, assess, and integrate the data are described clearly enough to allow replication. In addition, through increasing the efficiency of decision making, EBT can respond to societal pressure to decrease the resources of time, money, and vertebrate animals utilized in reaching decisions related to hazard and risk (Rovida and Hartung, 2009). These pressures have increased interest in developing alternative methods that reduce the time required to obtain relevant information (NRC, 2007). For this reason, the need to validate these alternative methods adds further impetus to EBT.

5 Initial steps towards systematic reviews in toxicology

We have carried out some of the more detailed studies using principles of EBM/HC to evaluate the evidence for associations between environmental toxicants and human health risks, and this experience provides some perspective on the challenges in adopting and adapting these methods to EBT (Navas-Acien et al., 2005, 2006, 2007; Maull et al., 2012). These reviews follow the norms of transparency and methods that have been developed for systematic reviews of diagnostics and interventions in medicine and health care. They incorporate the following steps: development and explicit framing of research questions that can be answered by a systematic review plus explicit statement of a publicly available protocol for conducting the systematic review. This protocol includes a defined and annotated strategy for locating sources of evidence; *a priori* conditions for exclusion and inclusion; defined analytic procedures to evaluate study designs and statistical methods; criteria for evaluating selected studies; methods for integrating study results. These rules are based on the assumption that all studies are well intentioned but no study is perfect. The goal is to identify all relevant sources of information in an unbiased manner and then to screen this body of information by identifying aspects of each study that can increase bias or uncertainty and to consider the impact of these aspects on analytic confidence.

Our attempts to integrate toxicological studies into our reviews were limited in terms of availability of studies, due in most cases to the variability in study design or in the endpoints selected, as well as to differences or lack of precise information on dosing and dose duration, and uncertainty as to the relevance of measured outcomes to the inference of human health risk. Some of these issues relate to toxicology, in which a range of endpoints often are utilized as relevant indicators of human disease risk; this is related to the lack of accepted phenotypic animal (or *in vitro*) models for many human health endpoints and uncertainty as to mechanisms involved in human disease. Lacking a coherent nosology, toxicological studies are likely to be more varied in design and endpoint than epidemiological or clinical studies. Integration of different endpoints may be possible using a systems biology approach to group endpoints in terms of common pathways, but this has not been tested in practice. These concerns also were cited by Maull et al. (2012).

A similar experience is presented in an excellent recent systematic review of formaldehyde and reproductive and development endpoints (Duong et al., 2011). The review of epidemiological studies is a model in transparency and rigor. In contrast (and similar to our reviews on lead and arsenic mentioned above), the review of experimental animal studies was less transparent. No clear information is presented on search terms and criteria for inclusion or exclusion of studies. Large differences were noted among studies in terms of species, routes of exposure and dose, as well as endpoints, which probably impeded any attempt at integration such that only a summary of “key findings” was presented. A thorough narrative discussion of mechanisms and modes of action also was included.

6 Challenges for EBT

The results of our analyses, along with more recent experience from an expert working group convened by the National Institute of Environmental Health Sciences (NIEHS) to evaluate associations between environmental chemicals and diabetes, indicate that toxicologists have considerable work to do to implement an evidence-based approach (Silbergeld, 2009). Innovations and modifications are especially needed to develop evidence-based methods tailored for toxicology and experimental nonhuman studies. Some of the major limitations noted in our reviews are discussed here for human studies and experimental studies. First, the amount of primary information available from independently conducted epidemiological studies in the published literature is relatively sparse for many exposures of interest. Second, many of the available epidemiological studies have significant problems in terms of study design or data reporting such that it is difficult to identify biases in them. For example, in many studies of arsenic, there are limited or no data on individual exposures and many studies failed to collect or report information on important covariates and confounders or information sufficient to determine heterogeneity. Many studies are relatively small and likely underpowered; many of the studies of larger cohorts (such as NHANES) are not actually independent of each other, and none are longitudinal, and so causality cannot be inferred in terms of exposure preceding outcome. In addition, there are broad differences in definition and measurement of outcomes of interest. This is understandable for toxicological studies, but is also characteristic of many epidemiological studies on, for example, lead and arsenic. For the toxicological studies, there is enormous heterogeneity in all aspects of study design and interpretation, as discussed above and in Duong et al. (2011). These criticisms were similar to the evaluations of the medical literature in the early '80s when systematic reviews in EBM/HC were first widely applied and just beginning to be appreciated (Dickersin and Manheimer, 1998).

Nevertheless, our reviews demonstrated that important elements of the methodology of systematic reviews can be adopted by EBT with little change, notably an allegiance to transparency in methods for searching the available literature for potential evidence, in selecting studies for review, and application of *a priori* criteria for assessing each selected study. Toxicologists can examine existing criteria for systematic reviews of observational epidemiology (Blair et al., 1995; AMS, 2007; Longnecker et al., 1988). When appropriate, some of the methods for integrating results across studies also may be adopted. From our analyses, we also observed that the greatest challenges for developing EBT are related to handling information from experimental nonhuman studies, where there is no consensus on analytic procedures and where even the construction of research questions may be more complex owing to the many test systems and endpoints used in studies on the same topic. In addition, there is no consensus on methods for screening primary studies, for evaluating the selected studies,

and on appropriate statistical methods to integrate study results from the range of experimental designs. This challenge will not be met by selecting information only from standard toxicology test guidelines or Good Laboratory Practice requirements as the definition of acceptability for evidence-based decisions. Many of these designs are extremely limited and, while they may produce data of use in standard risk assessment methods, they are underpowered and not robust (Reuter et al., 2003). As has been noted in endocrine disruptor research, these types of studies may be less informative than research studies that are more specifically designed to investigate defined hypotheses rather than to generate minimal information on hazard (Myers et al., 2009). Rather, all relevant studies should be sought and then evaluated using methods for appraising sources of biases identified through a consensus process in order to determine the strength of the evidence provided by each. Achieving this goal will foster a closer relationship between environmental epidemiology and experimental research, going beyond the invocation of experimental research merely to satisfy one of Bradford Hill's recommendations.

Achieving the goals of evidence-based and systematic analysis, as argued by practitioners in EBM/HC, has involved two strategies implemented at the beginning: involvement of a broadly based community for achieving consensus in methods and evaluations and a commitment to complete transparency. These commitments are exemplified within the Cochrane Collaboration. At its inception, the Collaboration included only a few dedicated investigators with a shared vision to help people make good health care decisions. This goal drove the development of systematic reviews and the dissemination of these reviews, which now cover a broad range of topics related to health care interventions. Key principles of transparency and continuous improvement in methods based on empirical evidence underlie the growth of the Cochrane Collaboration and its influence in the field of EBM/HC.³ This paper argues that these strategies, as well as a commitment to continuous growth and improvement in methods, are equally critical for the successful development and adoption of EBT.

The decision for EBT involves a commitment by the field of toxicology, not only to science but to community. As noted above, practitioners in EBM/HC stress that its success has involved the engagement of a broadly based community for consensus evaluations and a commitment to complete transparency. These steps cannot be rushed by establishing structural frameworks and empty institutions but must be grown from an organic discussion among the community of stakeholders, including scientists, technicians, governments, private sector, and the public (Chalmers, 2005).

Our success may transform the field of toxicology, as well as the practice of decision making in regulation. EBT can contribute to the efficient adoption of alternative methods through consensus agreement on identifying the evidence and on criteria for evaluation, drawing on experience from diagnostic evaluations in EBM/HC. However, there must be a commitment to

³ <http://www.cochrane.org/resources/handbook/index.htm>



empirically testing the methodology for systematic reviews of toxicological data; without such methodological studies, the field cannot move forward. This will not be a simple task. Since toxicology is fundamentally a science of prevention (Silbergeld et al., 2004), its aim is to detect likely harms prior to human exposure. For this purpose, experimental studies are the only source of truly preventive information, and thus the focus of EBT should be on experimental toxicology and test methods in the broadest sense.

Adoption of an evidence-based approach does not mean the adoption of the clinical trial design as the “highest” or only form of reliable information (Silbergeld, 2009). Evidence may come from any type of study, and although many reviews focus on randomized clinical trials, the type of evidence (i.e., study design) required depends on the type of research question (e.g., the use of randomized controlled clinical trials to answer questions of efficacy and cohort or case-control studies to answer questions related to etiology). This has facilitated the development of both “rules of practice” and the *post hoc* evaluation of research results (Dickersin and Manheimer, 1998). EBM/HC also provides a rich source of valuable guidance to EBT in its methods for evaluating observational epidemiology (Blair et al., 1995; Longnecker et al., 1988; AMS, 2007). While we can learn from EBM/HC, as noted at the outset of this paper, the issues of concern to toxicology, for the most part, are not the same as those in medicine and health care. In EBM/HC, the evidence-based approach has been developed most fully for answering questions related to therapy and diagnosis. The evaluation of novel test methods (such as alternative systems) may draw usefully upon methods used in evaluating diagnostics. Systematic reviews using only evidence from randomized controlled trials (RCTs) are not well suited to identifying harms, primarily due to study designs focused on identifying benefit, often with insufficient power to detect adverse effects because of the relatively low number of individuals exposed and the short time frame of many RCTs (Chou and Helfand, 2005).

The investment of our community in developing EBT will be worthwhile. In the absence of an evidence-based process, decision making is dependent upon a pseudo-Delphi process, in which experts are convened to undertake a qualitative process of integrating and weighing information (e.g., the NTP and IARC). This is less and less satisfactory to the public and other stakeholders; it is also highly resource-intensive in terms of repetitious studies and expert consultation (Rovida and Hartung, 2009). EBT will lead us into new domains of science and assessment, but we should remember that, in identifying harms and assessing risks, as in the law, an evidence-based approach does not remove the need for the application of judgment (Sackett et al., 1996; NRC, 1994). The premise and promise of EBT is the reduction of uncertainties by assuring a consistent body of information and enhancing confidence in the selection and evaluation of this information through a fully transparent process dedicated to continuous improvement through experience. These were the goals that inspired Archie Cochrane and the early community of analysts; by adopting them, the com-

munity of toxicologists can enhance the development of science and better serve the social goals of health protection and safety assurance.

References

- AMS (2007). *Identifying the Environmental Causes of Disease: How Should We Decide What to Believe and When to Take Action*. London, UK: The Academy of Medical Sciences.
- Axelrad, D. A., Bellinger, D. C., Ryan, L. M., et al. (2007). Dose-response relationship of prenatal mercury exposure and IQ: an integrative analysis of epidemiologic data. *Environ Health Perspect* 115, 609-615.
- Beronius, A., Rudén, C., Håkansson, H., and Hanberg, A. (2010). Risk to all or none? A comparative analysis of controversies in the health risk assessment of Bisphenol A. *Reprod Toxicol* 29, 132-46.
- Blair, A., Burg, J., Foran, J., et al. (1995). Guidelines for application of meta-analysis in environmental epidemiology. ISLI Risk Science Institute. *Regul Toxicol Pharmacol* 22, 189-197.
- Chalmers, I. (2005). Academia's failure to support systematic reviews. *Lancet* 365, 469.
- Chou, R. and Helfand, M. (2005). Challenges in systematic reviews that assess treatment harms. *Ann Intern Med* 142, 1090-1099.
- Crofton, K. M., Makris, S. L., Sette, W. F., et al. (2004). A qualitative retrospective analysis of positive control data in developmental neurotoxicity studies. *Neurotoxicol Teratol* 26, 345-352.
- Daston, G. P., Cook, J. C., and Kavlock, R. J. (2003). Uncertainties for endocrine disruptors: our view on progress. *Toxicol Sci* 74, 245-252.
- Dickersin, K., Scherer, R., and Lefebvre, C. (1994). Identifying relevant studies for systematic reviews. *BMJ* 309, 1286-1291.
- Dickersin, K. and Manheimer, E. (1998). The Cochrane Collaboration: evaluation of health care and services using systematic reviews of the results of randomized controlled trials. *Clin Obstet Gynecol* 41, 315-331.
- Duong, A., Steinmaus, C., McHale, C. M., et al. (2011). Reproductive and developmental toxicity of formaldehyde: a systematic review. *Mutat Res* 728, 118-138.
- EEA (2001). *Late Lessons from Early Warnings: The Precautionary Principle 1896-2000*. Copenhagen, Denmark: European Environment Agency.
- Feinstein, A. R. (1995). Meta-analysis: statistical alchemy for the 21st century. *J Clin Epidemiol* 48, 71-79.
- Flower, A., Lewith, G. T., and Little, P. (2007). Seeking an oracle: using the Delphi process to develop practice guidelines for the treatment of endometriosis with Chinese herbal medicine. *J Altern Complement Med* 13, 969-976.
- Gatchel, R. J. and McGeary, D. (2002). Cochrane collaboration-based reviews of health-care interventions: are they unequivocal and valid scientifically, or simply nihilistic? *Spine J* 2, 315-319.

- Griesinger, C., Hoffmann, S., Kinsner, A., et al. (2009). Special issue: Evidence-based toxicology (EBT). Preface. *Hum Exp Toxicol* 28, 83-86.
- Grimshaw, J. M., Santesso, N., Cumpston, M., et al. (2006). Knowledge for knowledge translation: the role of the Cochrane Collaboration. *J Contin Educ Health Prof* 26, 55-62.
- Guzelian, P. S., Victoroff, M. S., Halmes, N. C., et al. (2005). Evidence-based toxicology: a comprehensive framework for causation. *Hum Exp Toxicol* 24, 161-201.
- Hartung, T. (2010). Evidence-based toxicology – the toolbox of validation for the 21st century? *ALTEX* 27, 253-263.
- Hoffmann, S. and Hartung, T. (2006). Toward an evidence-based toxicology. *Hum Exp Toxicol* 25, 497-513.
- IARC (2006). *Inorganic and Organic Lead Compounds, IARC Monographs on the Evaluation of Carcinogenic Risks to Humans*. Lyon, France: International Agency for Research on Cancer.
- Judson, R., Kavlock, R., Martin, M., et al. (2013). Perspectives on validation of high-throughput assays supporting 21st century toxicity testing. *ALTEX* 30, 51-66.
- Longnecker, M. P., Berlin, J. A., Orza, M. J., et al. (1988). A meta-analysis of alcohol consumption in relation to risk of breast cancer. *JAMA* 260, 652-656.
- Mauill, E. A., Ahsan, H., Edwards, J., et al. (2012). Evaluation of the association between arsenic and diabetes: A National Toxicology Program Workshop Review. *Environ Health Perspect*. <http://dx.doi.org/10.1289/ehp.1104579>
- Moher, D., Jadad, A. R., and Tugwell, P. (1996). Assessing the quality of randomized controlled trials. Current issues and future directions. *Int J Technol Assess Health Care* 12, 195-208.
- Mulrow, C. D. (1987). The medical review article: state of the science. *Ann Intern Med* 106, 485-488.
- Myers, J. P., vom Saal, F. S., Akingbemi, B. T., et al. (2009). Why public health agencies cannot depend on good laboratory practices as a criterion for selecting data: the case of bisphenol A. *Environ Health Perspect* 117, 309-315.
- Navas-Acien, A., Sharrett, A. R., Silbergeld, E. K., et al. (2005). Arsenic exposure and cardiovascular disease: a systematic review of the epidemiologic evidence. *Am J Epidemiol* 162, 1037-1049.
- Navas-Acien, A., Silbergeld, E. K., Streeter, R. A., et al. (2006). Arsenic exposure and type 2 diabetes: a systematic review of the experimental and epidemiological evidence. *Environ Health Perspect* 114, 641-648.
- Navas-Acien, A., Guallar, E., Silbergeld, E. K., et al. (2007). Lead exposure and cardiovascular disease – a systematic review. *Environ Health Perspect* 115, 472-482.
- NRC – National Research Council (1994). *Science and Judgment in Risk Assessment*. Washington, DC: National Academies Press.
- NRC (2001). *Arsenic in Drinking Water*. Washington: NAS Press.
- NRC (2007). *Toxicity Testing in the 21st Century*. Washington, DC: National Academies Press.
- NTP – National Toxicology Program (2011). *NTP Monograph on Health Effects of Low Level Lead*, NTP, NIEHS, USDHHS, draft October 14, 2011. <http://ntp.niehs.nih.gov/NTP/ohat/Lead/DraftNTPMonographonHealthEffectsofLowLevelLead.pdf> (downloaded April 4, 2012).
- Oxman, A. D. and Guyatt, G. H. (1988). Guidelines for reading literature reviews. *CMAJ* 138, 697-703.
- Reuter, U., Heinrich-Hirsch, B., Hellwig, J., et al. (2003). Evaluation of OECD screening tests 421 (reproduction/developmental toxicity screening test) and 422 (combined repeated dose toxicity study with the reproduction/developmental toxicity screening test). *Regul Toxicol Pharmacol* 38, 17-26.
- Rovida, C. and Hartung, T. (2009). Re-evaluation of animal numbers and costs for in vivo tests to accomplish REACH legislation requirements for chemicals – a report by the transatlantic think tank for toxicology (t⁴). *ALTEX* 26, 187-208.
- Sackett, D. L., Rosenberg, W. M., Gray, J. A., et al. (1996). Evidence based medicine: what it is and what it isn't. *BMJ* 312, 71-72.
- Silbergeld, E. K. (2003). Facilitative mechanisms of lead as a carcinogen. *Mutat Res* 533, 121-133.
- Silbergeld, E., Lerman, S., and Hushka, L. (2004). Ethics. Human health research ethics. *Science* 305, 949.
- Silbergeld, E. (2009). 4.2 Applying an evidence-based approach: arsenic as a health risk. *Hum Exp Toxicol* 28, 127-129.
- Stephens, M., Andersen, M., Becker, R. A., et al. (2013). Evidence-based toxicology for the 21st century: Opportunities and challenges. *ALTEX* 30, 74-103.
- Straif, K., Stayner, L., Demers, P. A., et al. (2012). Use of meta-analyses by IARC working groups. *Environ Health Perspect* 120. doi:10.1289/ehp.1205397
- Williams, D. D. and Garner, J. (2002). The case against 'the evidence': a different perspective on evidence-based medicine. *Br J Psychiatry* 180, 8-12.

Correspondence to

Ellen Silbergeld, PhD
Department of Environmental Health Sciences
Johns Hopkins School of Public Health
615 N. Wolfe Street, E6644
Baltimore, MD, 21205
USA
e-mail: esilberg@jhsph.edu