

FORUM ARTICLE

The Emergence of Systematic Review in Toxicology

Martin L. Stephens,^{a,1} Kellyn Betts,^b Nancy B. Beck,^c Vincent Cogliano,^d Kay Dickersin,^e Suzanne Fitzpatrick,^f James Freeman,^g George Gray,^h Thomas Hartung,^{a,i} Jennifer McPartland,^j Andrew A. Rooney,^k Roberta W. Scherer,^e Didier Verloo,^l and Sebastian Hoffmann^m

^aJohns Hopkins Center for Alternatives to Animal Testing, Baltimore, Maryland, ^bFreelance Science and Technology Writer, Takoma Park, Maryland, ^cAmerican Chemistry Council, Washington, District of Columbia, ^dUS Environmental Protection Agency, Arlington, Virginia, ^eJohns Hopkins University Bloomberg School of Public Health, Baltimore, Maryland, ^fCenter for Food Safety and Applied Nutrition, Food and Drug Administration, College Park, Maryland, ^gExxonMobil Biomedical Sciences, Annandale, New Jersey, ^hGeorge Washington University Milken Institute School of Public Health, Washington, DC, ⁱUniversity of Konstanz, CAAT-Europe, Germany, ^jEnvironmental Defense Fund, Washington, District of Columbia, ^kOffice of Health Assessment and Translation, Division of the National Toxicology Program, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, North Carolina, ^lEuropean Food Safety Authority, Parma 43126, Italy and ^mseh consulting + services, Paderborn 33098, Germany

¹To whom correspondence should be addressed. Fax: (410) 614-2871. E-mail: msteph14@jhu.edu.

ABSTRACT

The Evidence-based Toxicology Collaboration hosted a workshop on “The Emergence of Systematic Review and Related Evidence-based Approaches in Toxicology,” on November 21, 2014 in Baltimore, Maryland. The workshop featured speakers from agencies and organizations applying systematic review approaches to questions in toxicology, speakers with experience in conducting systematic reviews in medicine and healthcare, and stakeholders in industry, government, academia, and non-governmental organizations. Based on the workshop presentations and discussion, here we address the state of systematic review methods in toxicology, historical antecedents in both medicine and toxicology, challenges to the translation of systematic review from medicine to toxicology, and thoughts on the way forward. We conclude with a recommendation that as various agencies and organizations adapt systematic review methods, they continue to work together to ensure that there is a harmonized process for how the basic elements of systematic review methods are applied in toxicology.

Key words: systematic review; risk of bias; data integration.

The Evidence-based Toxicology Collaboration (EBTC) is a group of North American and European researchers from government, academia, and industry who are interested in applying systematic review methods to toxicology (EBTC, 2015). The EBTC hosted a workshop on “The Emergence of Systematic Review and Related Evidence-based Approaches in Toxicology,” on

November 21, 2014 at the Johns Hopkins Bloomberg School of Public Health, in Baltimore, Maryland, USA. The workshop featured speakers from agencies and organizations developing and applying systematic review methods to questions in toxicology, speakers with experience in conducting systematic reviews in medicine and healthcare, and stakeholders from industry,

© The Author 2016. Published by Oxford University Press on behalf of the Society of Toxicology.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work properly cited. For commercial re-use, please contact journals.permissions@oup.com

government, academia, and non-governmental organizations. The full program and a brief summary are available online (EBTC, 2015). In this article, we expand on the workshop presentations to look at the use of systematic review approaches in toxicology, their roots in medicine and healthcare, the challenges facing practitioners, and some thoughts on the way forward, including implications for toxicologists in designing and reporting their studies.

Systematic review methods constitute a standardized approach for identifying and analyzing evidence related to clearly formulated questions (Higgins and Green, 2011; Institute of Medicine, 2011). Systematic reviews proceed through a sequence of steps, typically formulating a specific research question, developing a review protocol, performing a comprehensive literature search, selecting relevant studies, assessing the risk of bias of included studies, extracting and synthesizing the study data, rating the certainty in the findings, and interpreting and summarizing the findings (Table 1).

Compared to traditional narrative reviews, the systematic review framework is aimed at minimizing subjectivity and enhancing transparency, rigor, and consistency in the way reviews are conducted and reported (Silbergeld and Scherer, 2013). Transparency is enhanced through the drafting and posting of a protocol prior to commencing the review. The protocol specifies the research question; the literature search strategy; the inclusion/exclusion criteria for identifying relevant studies returned in the literature search; the framework for judging the quality of included studies; and the plan for data analysis, synthesis and presentation of findings. In this way, clear criteria for conducting the review are developed and specified in advance.

Although a systematic review approach increases the transparency and objectivity in the process, the conclusions of such reviews rely on the scientific judgment of the reviewers. Therefore, it is important to note that multiple reviews of a given topic will not necessarily address the same set of studies or come to the same conclusion. The basis of the scientific judgments and the conclusions should be made clear in a systematic review.

Some of the advantages and disadvantages of systematic review approaches are summarized in Table 2.

Systematic review methods, first developed for clinical medicine, have been adapted and are being used by toxicologists to synthesize available evidence, eg, on the potential association of exposure to a chemical with a particular health effect. The US National Toxicology Program, the US Environmental Protection Agency's Integrated Risk Information System program, the European Food Safety Authority, the Evidence-based Toxicology Collaboration, the UCSF Navigation Guide, and others are implementing systematic review methods in ways that meet their diverse programmatic needs. Despite their different applications

of systematic review methods, these organizations' approaches share commonalities including the fundamental steps of a systematic review.

SYSTEMATIC REVIEW METHODOLOGY AS APPLIED TO TOXICOLOGY

The initial steps of systematic review begin with scoping and problem formulation to identify the question of interest, gain a sense of the relevant literature, define and refine the question, define a PECO statement (see below), and develop the review protocol. Formulating the question to be answered is a simple-sounding process but it requires careful deliberation because the question guides the review. Then, a PECO statement is developed to identify the population (P), exposure (E), comparisons (C), and outcome (O) of interest to address the review question (European Food Safety Authority, 2010). This PECO statement is used to develop the literature search criteria and the inclusion/exclusion criteria for selecting the evidence relevant to answering the research question (Krauth et al., 2013).

A subsequent step is evaluating the risk of bias or methodological quality of the included studies, as pre-specified in the review protocol. Methodological quality refers to all aspects of a study's design, conduct, analysis, and outcome reporting that influence the study's ability to accurately answer the question posed. Risk of bias is a major component of methodological quality and refers to systematic errors that may lead to either an overestimation or an underestimation of the true effect (Higgins and Green, 2011). Shortcomings in the design, conduct, analysis, and outcome reporting of experiments add to the "risk of bias" or reduce methodological quality. The explicit evaluation of study bias is an important feature of systematic review historically and one not considered in traditional toxicology literature reviews. Examples of risks of bias include failure to (1) adequately randomize the administered dose or exposure level to each research subject in clinical trials or experimental studies, (2) account for important confounding or modifying variables, and (3) report all measured outcomes (Rooney et al., 2014). Dose selection per se, obviously an important issue in toxicology, is usually not considered a risk of bias issue. Rather, it is an element of external validity (or directness, applicability, and the extent to which a study's finding can be generalized to other circumstances). It remains to be determined to what extent issues of methodological quality beyond risk of bias should be incorporated in quality assessments in systematic reviews in toxicology, and the extent to which such issues are reported in research papers and thus amenable to assessment.

A growing number of tools have been developed to assess the risk of bias of environmental health studies (Krauth et al., 2013). There are published risk of bias frameworks for epidemiological studies and animal studies, but no such tools for *in vitro* studies or mechanistic data (Samuel et al., in press). However, there is a published approach in which mechanistic data are assessed for both methodological and reporting quality. The ToxRTool was created with funding from the European Commission and uses Klimisch codes (1 = reliable without restriction, 2 = reliable with restrictions, 3 = not reliable, and 4 = not assignable) (Klimisch et al., 1997) to evaluate and categorize the quality of toxicological data, including *in vitro* studies (Schneider et al., 2009).

Another major step in applying systematic review approaches to toxicology is integrating the evidence within and across diverse study types (eg, *in vivo*, *in vitro*, and human

TABLE 1. The basic steps of a typical systematic review.

1. Formulating a focused research question.
2. Preparing a protocol.
3. Applying the pre-defined literature search strategy.
4. Selecting the relevant papers by applying pre-defined inclusion and exclusion criteria.
5. Assessing the risk of bias of the included studies.
6. Extracting data on both the results relevant for addressing the research question and the study methods.
7. Synthesizing the data.
8. Rating the certainty in the findings.
9. Interpreting the results and presenting a summary of findings.

TABLE 2. Some advantages and disadvantages of systematic reviews.

Advantages.

A protocol for how the review will be conducted—written in advance—reduces the likelihood that ad hoc changes will be made that bias the outcomes. In cases where the protocol is published or otherwise shared with interested parties in advance of the actual review, stakeholders are thereby given the opportunity to recommend changes.

The incorporation of explicit criteria for including and excluding individual studies gives readers of the review a clear rationale for why some studies were included or excluded.

Assessing the risk of bias or broader methodological quality of the included studies gives reviewers and readers a sense of how much confidence to have in the review's conclusions.

Reviews that assess certain studies as having a high risk of bias are likely to encourage the authors of those studies to improve the quality of their future research.

The explicit and transparent nature of the review process and its published review give readers a clear sense of how the review was carried out. This also enables interested parties to replicate the review, with or without making any protocol amendments deemed desirable.

Under certain conditions, data synthesis lends itself to meta-analysis, which provides a quantitative summary of the data from individual studies and overall.

Disadvantages.

Even once familiar with the process and tools, conducting a systematic review is still likely to take considerable time and labor. Review teams are likely to include, at a minimum, an information specialist, a systematic review "methodologist," and subject-matter experts.

Although the basic framework for systematic reviews has remained the same across the fields to which it has been applied already, those seeking to apply this methodology to a new field will likely face some challenges not fully addressed by the experience gained in these other fields.

observational studies), in order to reach a conclusion. Data integration frameworks should be true to the evidence-based principles of transparency, objectivity, and consistency.

Groups of scientists in both the United States and European Union are collaborating to advance systematic review approaches in toxicology. Guidance for conducting systematic reviews in toxicology has been published (European Food Safety Authority, 2010; Rooney *et al.*, 2014; Woodruff and Sutton, 2014). There is substantial consistency across the approaches that have been developed in terms of how they frame the review (developing a specific research question and PECO statement), identify the evidence, evaluate its study quality (eg, by using risk of bias tools), and rate the certainty in the findings. The certainty rating in these approaches has been based on the GRADE approach, a system for evaluating the quality of the body of evidence, which is widely accepted in healthcare (Guyatt *et al.*, 2011).

AGENCY FRAMEWORKS AND OTHER EFFORTS

The Office of Health Assessment and Translation (OHAT) of the National Institute of Environmental Health Sciences has created a framework for applying systematic review methods to environmental health questions, including methods to develop conclusions from the full range of relevant data (human, animal and *in vitro* data) (Rooney *et al.*, 2014). The OHAT approach was developed in a process involving public comment and consultation with experts from toxicology and systematic review, building on and extending guidance from major systematic review groups (eg, the Cochrane Collaboration (Higgins and Green, 2011), Agency for Healthcare Research and Quality (Viswanathan *et al.*, 2013), GRADE Workgroup (Guyatt *et al.*, 2011), and the Navigation Guide Work Group (Woodruff and Sutton, 2014)). The evaluation process begins with a problem formulation step to form the specific research question and the PECO statement, and then involves the development of a protocol for conducting the review. The protocol outlines the methods for the evaluation tailored to the research question, including the literature search strategy, inclusion/exclusion criteria, risk of bias approach, establishing confidence in the evidence, and methods for evidence integration.

In the OHAT framework, evidence integration begins with the process for rating confidence in the findings for each body of evidence separately (eg, human and animal studies on a particular outcome) based on the GRADE approach (Guyatt *et al.*, 2008) with modifications on the initial starting point for observational human studies. It includes guidance for human and animal studies and a process for considering mechanistic studies. Rating confidence in the body of evidence is developed using the GRADE factors that reflect strengths and weaknesses of a body of evidence (eg, dose response, or indirectness) with an additional factor that may increase confidence in the association between exposure and health outcome when there is consistency of the response across species, study designs, or human populations. These ratings are translated into levels of evidence for each health effect based on whether the reviewed studies do or do not show an adverse effect. Finally, the degree of support from mechanistic studies is considered and the 3 evidence streams (human, animal, mechanistic) are integrated to reach a hazard conclusion of "known," "presumed," "suspected," or "not classifiable" as a hazard to humans that reflects the confidence and consistency across each body of evidence.

OHAT released a detailed methods guide of standard operating procedures for using systematic review in its evaluations in 2014 (National Toxicology Program, 2015). This guide reflected OHAT's then-current practices, with the expectation that the procedures will be updated and refined as best practices in the field of environmental health and systematic review continue to evolve.

The US Environmental Protection Agency's Integrated Risk Information System (IRIS) program is adapting and implementing systematic review methods (US Environmental Protection Agency, 2015a). These procedures, modeled after the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins and Green, 2011), also reflect recommendations from 2 US National Research Council (NRC) reviews of the IRIS process (National Research Council, 2011; National Research Council, 2014).

IRIS assessments involve multiple systematic reviews to determine whether a chemical causes specific adverse effects, eg, for endpoints such as carcinogenicity or neurotoxicity, in humans or in animals. The data include mainly animal and *in vitro* studies, and observational epidemiological studies. IRIS

assessments must address disparate data, such as different animal species and strains that may tolerate different doses, differing results (ie, an effect occurs in one species but not another), or occupational studies conducted while exposure levels change, eg, use of protective equipment, or changing industrial processes.

The IRIS program's emerging approach to systematic review is similar to OHAT's, and it includes a step for systematic integration of evidence for each health outcome. Both the NRC review (NRC 2014) and a subsequent workshop on the subject suggested that guided expert judgment, coupled with structured processes, are required for integrating IRIS evidence streams (US Environmental Protection Agency, 2015b).

After evidence integration, IRIS assessments characterize exposure-response relationships related to the EPA's need for toxicity values. The process for selecting studies to assess those relationships is similar to that of a systematic review. The agency is currently developing methods to combine results of the selected studies.

Systematic reviews typically include a literature-search cutoff date, after which "late-breaking" studies are not considered. Because IRIS evaluations are expected to consider late-breaking studies if they would change major conclusions, the EPA has developed a process for considering pivotal studies that are published after the literature search has closed (US Environmental Protection Agency, 2014). In general, new studies can be included until an assessment is readied for peer review. After peer review, the presumption shifts to not including new studies unless they have an impact on the credibility of an assessment's conclusions. Examples might be a strong new study that indicates a heretofore undiscovered health effect, or a strong new study that might change, in either direction, a major conclusion.

The European Food Safety Authority (EFSA) has been using systematic review approaches for a few years to fulfill its mandates. The Authority uses the reviews mainly for 2 different kinds of risk assessments: (1) for the evaluation of applications submitted with the goal of having a specific product, such as a pesticide, feed additive, or genetically modified organism, authorized for use in the EU, and (2) for generic assessments. Generic assessments review issues that arise within a wide range of areas where EFSA has jurisdiction, including animal health and welfare, plant health, feed additives, food actives, food contact materials, and health and nutrition claims. EFSA must also appraise systematic reviews conducted by applicants.

EFSA published its first guidance document on systematic reviews in 2010 with a team of authors that included experts from the Cochrane Collaboration and other groups performing systematic reviews in relevant fields (European Food Safety Authority, 2010). The organization has been conducting systematic review trainings.

EFSA has also produced reports on prioritizing questions for systematic review in risk analysis and on sources of evidence relevant for EFSA risk assessments (European Food Safety Authority, 2015a; O'Connor et al., 2012). In 2013 and 2014, EFSA authorized the creation of 23 systematic reviews on topics including pesticides, nutrition, feed, animal health, plant health, contaminants, biological hazards, genetically modified organisms, and methodologies. The Authority is committed to making the data from systematic reviews publicly available.

More recently, EFSA began what it calls the PROMoting METHods for Evidence Use in Science (Prometheus) project to further enhance the scientific rigor of the methodological approaches used in dealing with evidence. The project was

based on the recognition that evidence is needed in all assessments and the process for collecting, appraising, and analyzing it should be the same regardless of the objectives of the assessment or who conducts it. Assessments focused on efficacy, safety, and risk should all follow the same process. Another rationale for the Prometheus project is to address the issues presented when evidence is not available or there is insufficient time for applying extensive or complex approaches. EFSA recently published a report on the resulting methodological framework (European Food Safety Authority, 2015b). The Authority is also working on a report on how to analyze data gaps and the impacts thereof.

The EBTC is particularly interested in the Cochrane Collaboration's emerging methodology for systematic reviews of diagnostic test accuracy in medicine (Cochrane Collaboration, 2015) and its application to test method assessment in toxicology (Hoffmann and Hartung, 2005). The EBTC is using this approach to conduct a systematic review of zebrafish embryo testing as a predictor of developmental toxicity (de Vries et al., 2014). The aim is determine how well zebrafish embryo testing identifies teratogenesis, as compared to results from standard mammalian test protocols in rats and rabbits.

A primary driver for this review is to identify whether the zebrafish could serve as a partial replacement for the routine test for prenatal development, Test Guideline 414 of the Organisation for Economic Cooperation and Development. This test is costly in terms of money, time, and animals (primarily rats and rabbits) (Selderslaghs et al., 2009).

ANTECEDENTS

Systematic reviews are the hallmark of evidence-based medicine, which has been defined as the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients (Sackett et al., 1996). Evidence-based medicine involves integrating individual clinical expertise with the best available external evidence from systematic research (Sackett et al., 1996). The creation of organizations that shaped and promoted evidence-based medicine, such as the Cochrane Collaboration, facilitated the development of rigorous methods (Cochrane Collaboration, 2015). Systematic reviews have been defined as "an overview of primary studies which contains an explicit statement of objectives, materials and methods and has been conducted according to explicit and reproducible methodology" (Greenhalgh, 1997). In clinical medicine, such knowledge syntheses have proven "essential to advance practice and research through consolidation of evidence" (Colquhoun et al., 2014).

The U.S. government has launched numerous initiatives for systematic review, including the Evidence-based Practice Centers, which perform systematic reviews of treatment interventions across a wide spectrum of health conditions, and the U.S. Preventive Services Task Force, which performs reviews in the area of preventive medicine, including topics related to screening, counseling, and preventive medicines. In 2011, an Institute of Medicine panel published recommendations for the preparation of systematic reviews (Institute of Medicine, 2011).

Early promoters of the idea of translating the systematic review process from medicine to the field of toxicology included Philip Guzelian, who coined the term "evidence-based toxicology," and Thomas Hartung and Sebastian Hoffmann, who were interested in applying the process to assessments of test method performance/validation (Guzelian et al., 2005; Hoffmann and Hartung, 2006). Hartung founded the Evidence-based

Toxicology Collaboration in 2011 (Stephens *et al.*, 2013). As in medicine, evidence-based approaches in toxicology include not only evidence synthesis across studies (systematic review) but also the application of individual elements of systematic review methodology to other contexts, such as appraising the risk of bias of an individual article or appraising the quality of evidence in regulatory submissions.

Among the first actual systematic reviews in toxicology and environmental health were those conducted by Navas-Acien, Silbergeld, and colleagues, examining the association between exposure to environmental chemicals and human health effects (Navas-Acien *et al.*, 2005; Navas-Acien *et al.*, 2006; Navas-Acien *et al.*, 2008). These early reviews examined *in vivo*, *in vitro*, and epidemiological evidence to address specific questions. More recent reviews, including those by The Navigation Guide Work Group, have further explored how these diverse data streams could be integrated in a systematic review (Johnson *et al.*, 2014; Koustas *et al.*, 2014; Lam *et al.*, 2014; Woodruff *et al.*, 2011).

CHALLENGES

The challenges currently facing the more widespread application of systematic review approaches in toxicology are manifold.

- Data integration: Toxicology includes a diversity of study types providing relevant data. How and when evidence is integrated across study types is a subject that deserves careful thought to ensure that the process is transparent and replicable.
- Data accessibility:
 - Much of the data in toxicologically relevant databases is not publicly accessible. Some study data is available in databases that are not traditionally considered part of the scientific literature. How to include information from this format in systematic reviews is unclear, particularly if it is presented only in summary form. Other concerns relate to data that is proprietary or in formats that may not be exchangeable.
 - Retrieving toxicology data from PubMed is challenging because of the lack of MeSH (Medical Subject Headings) terms to subdivide toxicology. Therefore current best practices include MeSH terms and text word searches to identify the relevant literature.
 - In toxicology, no one information portal exists that is analogous to the one available for evidence-based medicine's online Cochrane Library, which provides up-to-date information independently generated by practitioners throughout the world about the effectiveness of health care interventions via 6 databases, including one focused on systematic reviews.
 - Efforts are underway to encourage industry stakeholders to share detailed data without putting competitive advantage at risk.
- Risk of bias: Application of risk of bias assessment methods to toxicology studies suggests that several possible sources of bias (randomization of treatment, lack of allocation concealment, and lack of blinding of outcome assessors) may be widespread among toxicology studies (Koustas *et al.*, 2014). This, in turn, suggests that the toxicological community should be better trained in using study design and conduct procedures to avoid risk of bias issues. Moreover, information from studies that generate negative data are not always published—a form of publication bias.

- Expert judgment: Exactly what constitutes the proper role for expert judgment in the context of a systemic review also merits some consideration and, potentially, guidance. The kind of expert judgment used in conducting a systematic review is and should be separate from the kind of expert judgment involved in making policy. A related challenge is the misperception that evidence-based approaches leave no room for professional judgment. Systematic reviews should strive to make expert judgments clear along with the scientific basis for those judgments in developing conclusions for a systematic review. Analyzing the approach that has been developed for involving expert judgment in risk analysis may prove helpful in efforts to determine how to best use expert judgment in systematic reviews for toxicology (Cooke and Goossens, 2008; Morgan, 1992).
- Workload manageability: For the EPA, which must sometimes contend with evaluating topics that have been the subject of thousands of studies, there is interest in finding ways to limit the literature search at the outset of a study while still including all truly informative studies. Some of the workload issues that the agency must contend with may have more to do with external validity than risk of bias. The EBTC is also seeking to develop streamlined approaches to data identification to enhance workload manageability without compromising evidence-based principles. There are also inherent challenges with the goals of each review, as an evaluation of all health effects potentially associated with a chemical will be necessarily broad compared with a focused review of a single health effect.

Other issues:

- Sufficient primary studies may not exist to adequately answer the review question. In these cases, the outcome of the systematic review would identify data gaps and research needs.
- Thought should be given as to who should be included on the work groups that conduct systematic reviews or subsequent peer reviews, including whether they should include regulators or other “customers” of systematic reviews.

WAY FORWARD

The challenges identified above to the advancement of systematic review approaches in toxicology, including issues of data accessibility, data integration, and workload manageability, are formidable. Recognition of these challenges is the starting point for further discussion and priority setting. Many of the workshop speakers—who are among the vanguard of those seeking to apply systematic review approaches in toxicology—expressed a willingness to continue to work together closely, where appropriate, to advance the field.

A strength of evidence-based medicine is that approaches such as systematic review and meta-analysis are quite uniformly applied. In translating these approaches, it is important that the safety sciences pursue a harmonized process and avoid, where possible, major discrepancies in terminology and approach driven by organizational preferences. A recent harmonization effort explored the similarities and differences in the use of risk of bias methods across organizations (Rooney *et al.*, 2016). The EBTC is committed to fostering the necessary international dialogue to facilitate this harmonization.

For its part, the EBTC is hoping, in time, to apply systematic review methodology to “qualify” biological pathways and pathway-based test methods for application to 21st century toxicology approaches. In this context, pertinent questions to pursue via literature review are whether proposed pathways reflect

actual pathways in the human body and whether proposed pathway-based tests do a good job of tracking perturbations to the pathways in question. These questions are related to the thorny issue of test method validation. More generally, evidence-based toxicology can aid in evaluating new mechanistic *in vitro* tools for assessing toxicity (Hartung, 2010). The new evaluation approaches compare well to traditional validation approaches in that they are more systematic and able to focus on mechanistic relevance, rather than on predicting animal data (Hartung et al., 2013). Approaches for how the methodology can be used to validate high-throughput assays in support of 21st century toxicity testing are being developed (Judson et al., 2013).

The issue of training came up repeatedly during the workshop. The Johns Hopkins Center for Alternatives to Animal Testing (CAAT) is developing a course, to commence in 2016, on systematic review and evidence-based toxicology.

Finally, the emergence of systematic review frameworks in toxicology has implications for practicing toxicologists, who, like any scientist, would want their data to be used in decision-making. For that to happen in the context of a systematic review, the relevance of published studies would be judged based on the PECO statement and the review's inclusion/exclusion criteria. These criteria depend heavily on the topic being reviewed but they generally favor studies that are well-designed with respect to issues such as choice of study subjects (eg, species, strain, age), dosages, and routes of administration. If included, the results would be assessed for study quality including risk of bias, which addresses issues that may be less familiar to toxicologists. To increase the likelihood that study data are used, toxicologists can minimize risk of bias through choices in study design and reporting, such as incorporation of techniques to ensure randomization and allocation concealment in assigning animals to treatment groups, and then blinding of the outcome assessors to the treatment groups. Interested toxicologists could begin to gain a familiarity with this topic by consulting, eg, the discussion of risks of bias in the NTP/OHAT systematic review framework (Rooney et al., 2014). At a more basic level, toxicologists should take care to draft the titles, abstracts, and key words of their published work to ensure ready retrieval in literature reviews of the subject, and to report study methods in sufficient detail as to allow an assessment of risk of bias/methodological quality by reviewers (Samuel et al., in press).

FUNDING

This work was supported by funding from an anonymous private charitable foundation and the ExxonMobil Foundation to the Johns Hopkins Center for Alternatives to Animal Testing (CAAT). CAAT serves as the secretariat for the Evidence-based Toxicology Collaboration, which hosted the workshop on which this article is based. Only those authors working at or for CAAT (K.B., T.H., S.H., and M.L.S.) received compensation from these sources via CAAT for their participation in the workshop or their preparation of this article.

Disclaimer: This paper does not necessarily represent the views or policies of the U.S. EPA.

REFERENCES

- Colquhoun, H. L., Levac, D., O'Brien, K. K., Straus, S., Tricco, A. C., Perrier, L., Kastner, M., and Moher, D. (2014). Scoping reviews: time for clarity in definition, methods, and reporting. *J. Clin. Epidemiol.* **67**, 1291–1294.
- Cooke, R. M., and Goossens, L. H. J. (2008). TU Delft expert judgment data base. Special issue on expert judgment. *Reliab. Eng. Syst. Safe* **93**, 657–674.
- de Vries, R. B. M., Wever, K. E., Avey, M. T., Stephens, M. L., Sena, E. S., and Leenaars, M. (2014). The usefulness of systematic reviews of animal experiments for the design of preclinical and clinical studies. *ILARJ.* **55**, 427–437.
- Evidence-based Toxicology Collaboration (2014). EBTC Workshop, November 21, 2014, Baltimore, MD. Available at: http://www.ebtox.com/nov2014_wksp/. Accessed November 13, 2015.
- Evidence-based Toxicology Collaboration (2015). Available at: <http://www.ebtox.com/>. Accessed November 13, 2015.
- European Food Safety Authority (2010). Application of systematic review methodology to food and feed safety assessments to support decision making. *EFSA J.* **8**, 1637–1727. doi:10.2903/j.efsa.2010.1637.
- European Food Safety Authority (2015a). EFSA Inventory of sources of relevance to systematic reviews of food and feed sources. Available at: <http://www.metaxis.com/EFSAInventory/>. Accessed November 13, 2015.
- European Food Safety Authority (2015b). Principles and process for dealing with data and evidence in scientific assessments. *EFSA J.* **13**, 4121–4157.
- Greenhalgh, T. (1997). Papers that summarise other papers (systematic reviews and meta-analyses). *BMJ* **315**, 672–675.
- Guyatt, G. H., Oxman, A. D., Vist, G. E., Kunz, R., Falck-Ytter, Y., Alonso-Coello, P., and Schünemann, H. J. and GRADE Working Group (2008). GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* **336**, 924–926.
- Guyatt, G., Oxman, A. D., Akl, E. A., Kunz, R., Vist, G., Brozek, J., Norris, S., Falck-Ytter, Y., Glasziou, P., DeBeer, H., et al. (2011). Introduction-GRADE evidence profiles and summary of findings tables. *J. Clin. Epidemiol.* **64**, 383–394.
- Guzelian, P. S., Victoroff, M. S., Halmes, N. C., James, R. C., and Guzelian, C. P. (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**, 241–251.
- Hartung, T., Hoffmann, S., and Stephens, M. (2013). Food for thought . . . Mechanistic validation. *Altex* **30**, 119–130.
- Higgins, J.P.T. and Green, S. (ed.) (2011). *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]*. The Cochrane Collaboration, 2011. www.cochrane-handbook.org. Accessed November 13, 2015.
- Hoffmann, S. and Hartung, T. (2005). Diagnosis: Toxic! – Trying to apply approaches of clinical diagnostics and prevalence in toxicology considerations. *Toxicol. Sci.* **85**, 422–428.
- Hoffmann, S., and Hartung, T. (2006). Toward an evidence-based toxicology. *Hum. Exp. Toxicol.* **25**, 497–513.
- Institute of Medicine (2011). *Finding What Works in Health Care: Standards for Systematic Reviews*. National Academies Press, Washington, DC.
- Johnson, P. I., Sutton, P., Atchley, D. S., Koustas, E., Lam, J., Sen, S., Robinson, K. A., Axelrad, D. A., and Woodruff, T. J. (2014). The Navigation Guide—evidence-based medicine meets environmental health: systematic review of human evidence for PFOA effects on fetal growth. *Environ. Health Perspect.* **122**, 1028–1039.

- Judson, R., Kavlock, R., Martin, M., Reif, D., Houck, K., Knudsen, T., Richard, A., Tice, R. R., Whelan, M., Xia, M., et al. (2013). Perspectives on validation of high-throughput assays supporting 21st century toxicity testing. *Altox* **30**, 51–66.
- Klimisch, H. J., Andreae, M., and Tillmann, U. (1997). A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data. *Regul. Toxicol. Pharmacol.* **25**, 1–5.
- Kousta, E., Lam, J., Sutton, P., Johnson, P. I., Atchley, D. S., Sen, S., Robinson, K. A., Axelrad, D. A., and Woodruff, T. J. (2014). The Navigation Guide—evidence-based medicine meets environmental health: systematic review of nonhuman evidence for PFOA effects on fetal growth. *Environ. Health Perspect.* **122**, 1015–1027.
- Krauth, D., Woodruff, T. J., and Bero, L. (2013). Instruments for assessing risk of bias and other methodological criteria of published animal studies: a systematic review. *Environ. Health Perspect.* **121**, 985–992.
- Lam, J., Kousta, E., Sutton, P., Johnson, P. I., Atchley, D. S., Sen, S., Robinson, K. A., Axelrad, D. A., and Woodruff, T. J. (2014). Review: The Navigation Guide—Evidence-Based Medicine Meets Environmental Health: integration of animal and human evidence for PFOA effects on fetal growth. *Environ. Health Perspect.* **122**, 1040–1051.
- Morgan, M. G. (1992). *Uncertainty: a Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. Cambridge University Press, Cambridge, New York.
- National Research Council (2011). *Review of the Environmental Protection Agency's Draft IRIS Assessment of Formaldehyde*. National Academies Press, Washington, DC.
- National Research Council (2014). *Review of the EPA's Integrated Risk Information System (IRIS) Process*. National Academies Press, Washington, DC.
- National Toxicology Program (2015). *Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration*. Available at: <http://ntp.niehs.nih.gov/go/38673>. Accessed November 13, 2015.
- Navas-Acien, A., Sharrett, A. R., Silbergeld, E. K., Schwartz, B. S., Nachman, K. E., Burke, T. A., and Guallar, E. (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., Clark, J. M., Burke, T. A., and Guallar, E. (2006). Arsenic exposure and Type 2 Diabetes: A systematic review of the experimental and epidemiologic evidence. *Environ. Health Perspect.* **114**, 641–648.
- Navas-Acien, A., Schwartz, B. S., Rothenberg, S. J., Hu, H., Silbergeld, E. K., and Guallar, E. (2008). Bone lead levels and blood pressure endpoints: a meta-analysis. *Epidemiology* **19**, 496–504.
- O'Connor, A. M., Lovei, G. L., Eales, J., Frampton, G. K., Glanville, J., Pullin, A. S., and Sargeant, J. (2012). *Implementation of systematic reviews in EFSA scientific outputs workflow*. Supporting Publications 2012: EN-367 <http://www.efsa.europa.eu/en/supporing/pub/367e.htm>. Accessed November 13, 2015.
- Rooney, A. A., Boyles, A. L., Wolfe, M. S., Bucher, J. R., and Thayer, K. T. (2014). Systematic review and evidence integration for literature-based environmental health science assessments. *Environ. Health Perspect.* **122**, 711–718.
- Sackett, D., Rosenberg, W. M., Gray, J. A., Haynes, R. B., and Richardson, W. S. (1996). Evidence-based medicine: what it is and what it isn't. *BMJ* **312**, 71.
- Samuel, G. O., Hoffmann, S., Wright, R. A., Lalu, M. M., Patlewicz, G., Becker, R. A., DeGeorge, G. L., Fergusson, D., Hartung, T., Lewis, R. J., et al. (2016). Guidance on assessing the methodological and reporting quality of toxicologically relevant studies: a scoping review. *Environ. Int.* <http://dx.doi.org/10.1016/j.envint.2016.03.010>.
- Schneider, K., Schwarz, M., Burkholder, I., Kopp-Schneider, A., Edler, L., Kinsner-Ovaskainen, A., Hartung, T., and Hoffmann, S. (2009). "ToxRTool", a new tool to assess the reliability of toxicological data. *Toxicol. Lett.* **189**, 138–144.
- Selderslaghs, I. W. T., Van Rompay, A. R., De Coen, W., and Witters, H. E. (2009). Development of a screening assay to identify teratogenic and embryotoxic chemicals using the zebrafish embryo. *Reprod. Toxicol.* **28**, 308–320.
- Silbergeld, E., and Scherer, R. W. (2013). Evidence-based toxicology: strait is the gate, but the road is worth taking. *Altox* **30**, 67–73.
- Stephens, M. L., Andersen, M., Becker, R. A., Betts, K., Boekelheide, K., Carney, E., Chapin, R., Devlin, D., Fitzpatrick, S., Fowle, J. R. 3rd., et al. (2013). Evidence-based toxicology for the 21st century: opportunities and challenges. *Altox* **30**, 74–103.
- US Environmental Protection Agency (2014). IRIS Stopping Rules. Available at: http://www2.epa.gov/sites/production/files/2014-06/documents/iris_stoppingrules.pdf. Accessed November 13, 2015.
- US Environmental Protection Agency (2015a). Background for Advancing Systematic Review Workshop (December 2015). Available at: <http://www2.epa.gov/iris/advancing-systematic-review-workshop-december-2015#background>. Accessed December 2, 2015.
- US Environmental Protection Agency (2015b). Systematic Review Workshop Materials. Available at: <http://www.epa.gov/iris/systematic-review-workshop-materials>. Accessed December 7, 2015.
- Viswanathan, M., Berkman, N. D., Dryden, D. M., and Hartling, L. (2013). Assessing Risk of Bias and Confounding in Observational Studies of Interventions or Exposures: Further Development of the RTI Item Bank. *Methods Research Report*. AHRQ Publication No. 13-EHC106-EF. Agency for Healthcare Research and Quality, Rockville, MD.
- Woodruff, T. J., Sutton, P., and The Navigation Guide Work Group, (2011). An evidence-based medicine methodology to bridge the gap between clinical and environmental health sciences. *Health Aff.* **30**, 5931–5937.
- Woodruff, T. J., and Sutton, P. (2014). The Navigation Guide systematic review methodology: A rigorous and transparent method for translating environmental health science into better health outcomes. *Environ. Health Perspect.* **122**, 1007–1014.