ehp ENVIRONMENTAL HEALTH PERSPECTIVES
http://www.ehponline.org

The Navigation Guide Systematic Review Methodology: A Rigorous and Transparent Method for Translating Environmental Health Science into Better Health Outcomes

Tracey J. Woodruff and Patrice Sutton

http://dx.doi.org/10.1289/ehp.1307175
Received: 4 June 2013
Accepted: 24 February 2014
Advance Publication: 25 June 2014

NIH National Institute of Environmental Health Sciences

The Navigation Guide Systematic Review Methodology: A Rigorous and Transparent Method for Translating Environmental Health Science into Better Health Outcomes

Tracey J. Woodruff and Patrice Sutton

Program on Reproductive Health and the Environment, University of California, San Francisco, Oakland, California, USA

Address correspondence to Tracey J. Woodruff, UCSF Program on Reproductive Health and the Environment. 1330 Broadway, Suite 1135, Oakland, CA, 94612, USA. Telephone 510-350-1241. E-mail: woodrufft@obgyn.ucsf.edu

Running title: The Navigation Guide Systematic Review Method

Acknowledgments: Funding for this commentary was provided to UCSF's Program on Reproductive Health and the Environment by the Clarence Heller Foundation, Forsythia Foundation, Fred Gellert Family Foundation, Passport Foundation, and New York Community Trust. Financial support for development of the PFOA case study was through grants from New York Community Trust and the U.S. Environmental Protection Agency through a contract with Abt Associates (GAIA-0-6-UCSF 17288). Support over the past 5 years for the development and dissemination of the Navigation Guide methodology was provided by: Clarence Heller Foundation, Passport Foundation, Forsythia Foundation, the Johnson Family Foundation, the Heinz Endowments, the Fred Gellert Foundation, Rose Foundation, Kaiser Permanente, New York Community Trust, University of California, San Francisco Institute for Health Policy Studies, Planned Parenthood Federation of America, National Institute for Environmental Health

Sciences (ES018135 and ESO22841) and U.S. Environmental Protection Agency STAR (RD83467801 and RD83543301). We are indebted to Dylan Atchley, Daniel Axelrad, Lisa Bero, Paula Johnson, Erica Koustas, and Juleen Lam for providing invaluable comments and suggestions on this commentary. Dylan Atchley also provided research assistance. The final paper is the responsibility of the authors.

Competing Financial Interests: We have no competing financial interests to disclose.

Abstract

Background: Synthesizing what is known about the environmental drivers of health is instrumental to taking prevention-oriented action. Methods of research synthesis commonly used in environmental health lag behind systematic review methods developed in the clinical sciences over the past 20 years.

Objectives: Develop proof of concept of the "Navigation Guide," a systematic and transparent method of research synthesis in environmental health.

Discussion: The Navigation Guide methodology builds on best practices in research synthesis in evidence-based medicine and environmental health. Key points of departure from current methods of expert-based narrative review prevalent in environmental health include: an *a priori* protocol; standardized and transparent documentation including expert judgment; a comprehensive search strategy; assessment of "risk of bias"; and separation of the science from values and preferences. Key points of departure from evidence-based medicine include: human observational studies were assigned a "moderate" quality rating; and diverse evidence streams were combined.

Conclusions: The Navigation Guide methodology is a systematic and rigorous approach to research synthesis that has been developed to reduce bias and maximize transparency in the evaluation of environmental health information. While novel aspects of the method will require further development and validation, our findings demonstrated that improved methods of research synthesis under development at the National Toxicology Program and under consideration by the U.S. Environmental Protection Agency are fully achievable. The institutionalization of robust methods of systematic and transparent review would provide a concrete mechanism for linking science to timely action to prevent harm.

Introduction

There is an urgent unmet need to shorten the time between scientific discovery and improved health outcomes. Population exposure to toxic environmental chemicals is ubiquitous (Centers for Disease Control and Prevention 2012; U.S. Environmental Protection Agency 2013c) and adverse health outcomes associated with exposure to such chemicals are prevalent and on the rise (Newbold and Heindel 2010; Olden et al. 2011; U.S. Environmental Protection Agency 2013c; Woodruff et al. 2010; World Health Organization and United Nations Environment Programme 2013). The health and economic benefits of translating scientific discoveries into actions to prevent harm and reap benefits have been clearly demonstrated. For example, global efforts to remove lead from gasoline have produced health and social benefits estimated at $2.4 trillion dollars annually (Tsai and Hatfield 2011); and the value of better air quality, including reductions in premature death and illness, and improved economic welfare and environmental conditions from the programs implemented pursuant to the 1990 Clean Air Act, will reach almost $2 trillion dollars in 2020 (U.S. Environmental Protection Agency 2011). However, many potential benefits have been squandered due to delays in acting on the available science (European Environment Agency 2012). Due to deficiencies in the current regulatory structure for manufactured chemicals, a failure or delay in acting on the science means that exposure to toxic chemicals persists while evidence of harm mounts (Vogel and Roberts 2011).

Failing or delaying to take action to prevent exposure to harmful environmental chemicals is not an inconsequential or neutral policy choice. For example, the costs in 2008 to the U.S. healthcare system for treatment of childhood illnesses linked to toxic environmental exposures has been estimated to be over $76 billion (Trasande and Liu 2011). Failure to prevent even low-level

environmental exposures can have large society-wide adverse consequences for health if exposures are ubiquitous (Bellinger 2012).

To the extent that science informs public policy to prevent harm, a robust method to synthesize what is known about the environmental drivers of health in a transparent and systematic manner is a necessary foundational step to making the science actionable. The body of science is voluminous, of variable quality and largely unfamiliar to decision-makers. Early warning signals of harm can be masked by the fragmented, complex, and at times, conflicting nature of the available information, undermining our capacity to act wisely. Yet consistently applied and transparent rules and descriptors about how environmental health science is translated into strength of evidence conclusions have been lacking (Beronius et al. 2010; Gee 2008; National Research Council 2009, 2011).

Today, methods of research synthesis prevalent in environmental health mirror that of clinical medicine over 40 years ago when the clinical sciences largely relied on a system of expert-based narrative reviews on which to recommend treatment decisions (Rennie and Chalmers 2009). A landmark paper by Antman and colleagues in 1992 in the *Journal of the American Medical Association* showed the superiority of systematic review methods by comparing expert opinion-based recommendations for treatment of myocardial infarction published in scientific reviews and clinical textbooks to statistical analyses of the combined results of randomized controlled trials (Antman et al. 1992). The paper by Antman et al. documented the lack of timely incorporation of experimental evidence into expert-based recommendations showing that some expert reviews did not mention effective therapies while others recommended therapies proven to be ineffective or even dangerous. From there, explicit approaches that harness expertise to a rigorous, transparent, and systematic methodology to evaluate a clearly formulated question were

advanced, and are now embodied in prominent empirically-demonstrated methods such as the Cochrane Collaboration and The Grading of Recommendations Assessment, Development and Evaluation (GRADE) (Guyatt et al. 2008b; Higgins and Green 2011). These methods are regularly relied on to inform billions of dollars of health care decisions in order to achieve cost savings and better health outcomes (Fox 2010). Howells et al. estimated that utilization of systematic review and meta-analysis of the pre-clinical evidence, i.e., animal studies undertaken prior to human drug trials, could reduce the cost of developing drugs for treating stroke by $1.1 to 7.9 billion dollars, the savings due to improving the validity of the evidence informing decisions on whether to advance drugs to clinical trials (Howells et al. 2012). It is anticipated that U.S. healthcare policy decisions will increasingly rely on systematic review methodologies; for example healthcare reform legislation has allocated $1.1 billion dollars for comparative effectiveness research (Centers for Disease Control and Prevention 2009).

The field of environmental health is now embarking on a similar journey. Reviews of the scientific evidence are as integral to decision-making about exposure to environmental chemicals in national and local government agencies and industry as they are for making treatment decisions in clinical medicine. But predominant approaches in use for evaluating the evidence in environmental health are more than 30 years old, based on expert opinion, and, with notable exceptions (National Toxicology Program 2013; U.S. Department of Health and Human Services 2006; U.S. Environmental Protection Agency 2013b) generally do not provide weight of evidence summaries for outcomes other than cancer. Improved methods of risk assessment that better reflect our current understanding of the science have been articulated by the National Academy of Sciences in its 2008 *Phthalates and Cumulative Risk Assessment: The Task Ahead* and in its 2009 *Science and Decisions* (National Research Council 2008, 2009). Systematic

approaches to evidence-based decision making that can improve our capacity to meet the needs of decision-makers are also currently underway at the National Toxicology Program (NTP) (Rooney et al. 2014) and under consideration at the U.S. Environmental Protection Agency (USEPA) (National Research Council 2011, 2014, 2014). Described below are the results of the application of a novel method for systematic and transparent review in environmental health that demonstrate that such advances are not only desirable but within our grasp.

Discussion

Overview of the Navigation Guide Methodology

With the goal of expediting the development of evidence-based recommendations for preventing harmful environmental exposures, beginning in 2009 a collaboration of scientists and clinicians undertook the development of the Navigation Guide methodology for systematic review. The Navigation Guide methodology was developed by coupling the rigor of systematic review methods being used by the clinical sciences to the "bottom line" approach to research synthesis being used by the International Agency for Research on Cancer (IARC) (International Agency for Research on Cancer 2006). Features of systematic reviews used in clinical medicine encompass specifying an explicit study question, conducting a comprehensive search, rating the quality and strength of the evidence according to consistent criteria, and meta and other statistical analyses; IARC's method allows for combining the results of human and non-human evidence into a single concise statement of health hazard (Woodruff et al. 2011).

As such, the Navigation Guide methodology translates the achievements of the past 20 years in evidence-based medicine into environmental health.

The Navigation Guide methodology involves 4 steps:

1. **Specify the Study Question**: frame a specific question relevant to decision-makers about whether human exposure to a chemical, class of chemicals or other environmental exposure is a health risk.

2. **Select the Evidence**: Conduct and document a systematic search for published and unpublished evidence.

3. **Rate the Quality and Strength of the Evidence:** Rate the quality of individual studies and the quality of the overall body of evidence based on *a priori* and transparent criteria. The Navigation Guide methodology conducts this process separately for human and non-human systems of evidence. As a consequence, the methodology involves an additional step of integrating the quality ratings of each of these two streams of evidence. The end result is one of five possible statements about the overall strength of the evidence: "known to be toxic", "probably toxic", "possibly toxic", "not classifiable", or "probably not toxic".

We were part of a team of scientists that developed the Navigation Guide method and applied steps 1-3 to the question "does developmental exposure to perfluorooctanoic acid (PFOA) affect fetal growth?" (Johnson PI et al. 2014; Koustas et al. 2014; Lam et al. 2014). Step 4 of the method, *Grade the Strength of the Recommendations*, involves integrating the strength of the evidence on toxicity (from step 3) with information about exposure, the availability of less toxic alternatives, and patient values and preferences. This step was not addressed in the PFOA case study due to the limitations of our resources. Below we highlight the features of the method that are new to environmental health, features that differ from methods used in evidence-based medicine, a comparison of the results of the Navigation Guide method to previous reviews of PFOA exposure and toxicity, limitations of the Navigation Guide method, and future directions.

Navigation Guide Features New to Environmental Health Reviews

To initiate the development of the Navigation Guide methodology we convened a novel inter-disciplinary team of 22 individuals from governmental and non-governmental organizations and academia (Woodruff et al. 2011). Two members of this team, Daniel Fox, President Emeritus of the Milbank Memorial Fund and Lisa Bero, currently Co-Chair of the Cochrane Collaboration, were world-renown experts on systematic review methodologies used in the clinical sciences. Seven members were scientists or environmental health advocates from international, national, state, and local government agencies and a non-governmental organization directly engaged in developing and/or employing strength of evidence conclusions in decision-making on environmental chemicals: David Gee, European Environmental Agency; Vincent James Cogliano, International Agency for Research on Cancer; Kathryn Guyton, USEPA; Lauren Zeise, California Environmental Protection Agency; Julia Quint, California Department of Public Health (retired); Karen Pierce, San Francisco Department of Public Health; and Heather Sarantis, Commonweal. Eleven were health professionals with expertise in women's, reproductive, pediatric and/or environmental health: Jeanne Conry, American Congress of Obstetricians and Gynecologists District IX and Kaiser Permanente; Mark Miller, UCSF's Pediatric Environmental Health Specialty Unit; Sarah Janssen, Natural Resources Defense Council; Beth Jordon and Rivka Gordon, Association of Reproductive Health Professionals; Sandy Worthington, Planned Parenthood Federation of America; Pablo Rodriguez, Brown Medical School and Women & Infants Hospital of Rhode Island; Michelle Ondeck and Judith Balk, University of Pittsburgh; Victoria Maizes, University of Arizona; and Ted Schettler, Science and Environmental Health Network. Finally our own expertise involved decades of work at the interface of environmental

and occupational health and public policy. At the time of publication of the method none of the collaborators reported a competing financial interest.

To conduct the first application of the Navigation Guide method we assembled a team of nine scientists from academia and USEPA that encompassed the multi-disciplinary expertise required to apply the methodology, including in environmental health sciences; epidemiology; toxicology; risk assessment; biostatistics; and the science of systematic reviews (Johnson PI et al. 2014; Koustas et al. 2014; Lam et al. 2014). One team member, Dr. Karen Robinson, Director of the Evidence Based Practice Center at Johns Hopkins University, was an expert on the identification, synthesis and presentation of evidence for informing healthcare decisions and research; three team members, Patrice Sutton, Erica Koustas, and Paula Johnson had formal training in Cochrane and/or GRADE methodologies. None of the review team reported a competing financial interest.

The method developed and applied through these inter-disciplinary teams builds on the best practices in research synthesis in evidence-based medicine and environmental health. Key points of departure of the Navigation Guide from current methods of expert-based narrative reviews in environmental health include:

1. A Protocol

The application of the Navigation Guide is guided by a detailed protocol developed prior to undertaking the review (Figure 1). In contrast, expert-based narrative review methods do not provide an *a priori* document that defines a specific question to be answered and sets up the "rules" of the evaluation. An *a priori* protocol is a staple of systematic reviews in the clinical sciences because it reduces the impact of review authors' biases, provides for transparency of

methods and processes, reduces the potential for duplication, and allows for peer review of the planned methods (Higgins and Green 2011). Notably, the protocol also provides a transparent forum to incorporate the expertise of non-scientists, including health-impacted populations and their advocates, in framing a meaningful study question. The protocol is developed around a "PECO" statement [**P**articipants, **E**xposure, **C**omparator, and **O**utcome(s)], which provides the framework from which studies are identified and selected for inclusion. The PECO statement is similar to recommendations by the National Academy of Sciences for improving the design of risk assessment through planning, scoping, and problem formulation to better meet the needs of decision-makers (National Research Council 2009).

2. Standardized and Transparent Documentation including Expert Judgment

Systematic reviews are not "automated" or "computerized" or otherwise conducted without applying judgment (Guyatt et al. 2011). The fundamental shift from existing methods of expert review in environmental health science is that each step of the Navigation Guide is conducted in a thorough, consistent, and transparent manner, and all information, including judgments, is documented and displayed in the same way. In short, the rationale for a decision is traceable, reproducible and comprehensible.

3. Assessment of "Risk of Bias"

The assessment of "risk of bias," defined as characteristics of a study that can introduce systematic errors in the magnitude or direction of the results (Higgins and Green 2011), is a new concept in environmental health. Systematic review methodologies distinguish between study quality criteria that can introduce a systematic error in the magnitude or direction of the result (i.e., risk of bias or "internal validity") from other methodological quality or reporting elements, which are related to important standards by which a study is conducted (e.g., adherence to human

subjects and animal welfare requirements) or reported (e.g., complete information provided), but that do not systematically influence study outcomes. A study conducted to the highest methodological standards can still have important risk of bias that will affect the magnitude or direction of a study outcome.

Risk of bias domains have been well developed and empirically shown to influence study outcomes in experimental human studies (Higgins et al. 2011; Roseman et al. 2011). However, risk of bias domains that are equally agreed upon for human observational studies are lacking. In the PFOA case study, we based our risk of bias domains for observational human studies on the domains used by the Cochrane Collaboration and the Agency for Healthcare Research and Quality (Higgins and Green 2011; Viswanathan et al. 2012) including recruitment strategy, blinding, confounding, incomplete outcome data, selective reporting, and exposure assessment.

Domains for risk of bias for animal studies are also under development. While thirty instruments have been identified in the environmental health literature for evaluating quality of animal studies, they are mostly comprised of domains related to reporting requirements, such as compliance with regulatory requirements, statistical model explained, and test animal details; importantly they do not include all the risk of bias domains in use in human experimental studies (Krauth et al. 2013).

To develop risk of bias domains for applying the Navigation Guide to animal studies, we adapted the risk of bias domains used in human experimental studies that have an empirical basis including: sequence generation; allocation concealment; blinding; incomplete outcome data; and selective reporting (See Figure 1 in Johnson et al. 2014 and Figure 1 in Koustas et al. 2014). According to GRADE, these five criteria address nearly all issues that bear on the quality of

human experimental evidence (Balshem et al. 2011). Further, these elements have been shown in the pre-clinical animal literature to influence study outcomes (Vesterinen et al. 2010). Our rationale was that risk of bias in a non-human experiment is comparable to risk of bias in human and pre-clinical animal experiments.

Further, in both human and animal studies, we included a "conflict of interest" risk of bias domain. This domain has been proposed---but not yet adopted---by Cochrane and GRADE as an important risk of bias (Bero 2013). This is based on empirical data from studies of the health effects of tobacco (Barnes and Bero 1997, 1998), the safety and efficacy of pharmaceuticals (Bero et al. 2007; Lexchin et al. 2003; Lundh et al. 2012), and medical procedures (Popelut et al. 2010; Shah et al. 2005) which have all shown that, on average, source of funding influences study outcome.

The assessment of risk of bias in the PFOA case study revealed worrisome truths about the conduct and reporting of experimental animal studies in environmental health. In particular, we found that included toxicology studies uniformly did not apply methodological approaches that are empirically recognized as minimizing bias in human experimental study outcomes. In particular, none of the studies reported how or if they used adequate allocation concealment, regardless of whether the studies were conducted through Good Laboratory Practices (GLP), by industry groups, or by independent research laboratories. Sub-optimal experimental animal study design and reporting is prevalent in the pre-clinical literature, and introduces bias into study findings (Bebarta et al. 2003; Landis et al. 2012; Macleod et al. 2004; McPartland et al. 2007; van der Worp et al. 2007; van der Worp and Macleod 2011; Vesterinen et al. 2011). For example, studies by the *Collaborative Approach to Meta Analysis and Review of Data from Experimental Studies* (*CAMARADES*) collaboration have shown that studies using randomization

and allocation concealment reported less improvement in heart response measures in animal models of focal ischemia treated with the pharmaceutical NXY059 (Macleod et al. 2008) and less improvement in neurobehavioral scores in animal models of intra-cerebral hemorrhage (Frantzias et al. 2011) than other studies.

In our outreach efforts related to the Navigation Guide we found that environmental health researchers in many and varied settings reported that methodological approaches to reduce bias in toxicological studies were not widely recognized or customary practices.

A second challenge to conducting risk of bias assessments and quantitative analyses in the PFOA case study was that the necessary data were not all reported in the published studies. Our efforts to contact study authors to get the needed data were moderately successful, i.e., 18 of 28 (64%) authors contacted responded, and were critical to our ability to conduct the review. We anticipate that contacting study authors will be a necessary step for those conducting systematic reviews until such time that steps are undertaken--by journals, funding agencies and through study registries---to standardize optimal reporting. Our findings underscore the urgency of calls for improved access to the data needed to conduct scientifically robust reviews of environmental health science (Goldman and Silbergeld 2013), and the importance to environmental health of nascent efforts in the pre-clinical arena to develop improved experimental animal study design and reporting (Landis et al. 2012; van der Worp and Macleod 2011; Vesterinen et al. 2011).

4. Comprehensive and Efficient Search Strategy

The outcome of the Navigation Guide search method demonstrated the potential for systematic reviews to be more comprehensive than traditional reviews. We evaluated four more human studies than an expert panel appointed to review the health effects of PFOA (C8 Science Panel

2011). The search strategy used to gather data for the C8 panel was not published. However, as these four papers did not present data that proved to be essential to the conclusions of the review, i.e. the data included was from small studies that did not weight heavily in the meta-analysis, they could have been identified by the C8 Panel's search but excluded from their reference list. Our comprehensive search strategy captured studies that measured PFOA exposure and fetal growth parameters but did not necessarily draw associations between the two. The four additional studies included in our review did not have birth weight or other fetal growth measures as the primary outcome or main topic of the paper (Fromme et al. 2010; S Kim et al. 2011; SK Kim et al. 2011; Wang et al. 2011). However, because our search identified these studies, we included them, contacted the study authors, and obtained additional relevant data to support our review from authors of two of these studies (Fromme et al. 2010; S Kim et al. 2011) and were referred by one author (Wang et al. 2011) to an article under peer review at the time on the same cohort with more relevant data (Chen et al. 2012). We also identified 10 more non-human studies compared to our own earlier nonsystematic literature review. Our adoption of a search filter for animal studies in use in the pre-clinical literature (Hooijmans et al. 2010) greatly expedited the development of a search for relevant animal studies.

We found that casting a wide net for relevant studies was feasible due to the *a priori* development of a PECO statement, from which we developed very explicit criteria used to efficiently screen titles and abstracts and due to the use of a software program that expedited the screening process. For this case study, our search strategy identified slightly over 2,000 non-human and 3,000 human potentially relevant studies. For the human data, it took one person-day to screen titles and abstracts (resulting in 248 articles eligible for full text review) and one week

to do a full text review which identified 18 relevant studies for evaluation. The time for the evaluation of the non-human data was similar.

Further, by applying a method that seeks to extract the exact same information, laid out in the same transparent way, our ability to interpret and understand the results was straightforward. As the application of systematic reviews expands, we anticipate greater efficiencies will be gained, for example, through the development of improved search filters, screening and management systems.

5. Separation of the Science From Values and Preferences

The PFOA case study demonstrated Steps 1-3 of the Navigation Guide methodology the result of which was a concise statement regarding PFOA's toxicity. However, toxicity is just one aspect of a risk management decision in environmental health. In step 4 of the Navigation Guide, which is modeled after GRADE's methods for rating treatment recommendations (Guyatt et al. 2008a), other important factors are brought to bear on recommendations for prevention, including values and preferences, extent of exposures, the availability of safer alternatives, and costs and benefits. Thus, the Navigation Guide transparently and explicitly delineates the science from other key considerations. While we did not have the resources to operationalize Step 4 in the PFOA case study, we hope to do so in future case studies.

Navigation Guide Features Different from Evidence-Based Medicine

Due to differences between environmental and clinical health sciences related to the evidence-base and decision-context, systematic review methodologies used in the clinical sciences were not seamlessly applicable to environmental exposures (Woodruff et al. 2011). Two key points of departure of the Navigation Guide methodology from evidence-based medicine are:

1. The Body of Human Observational Studies Is Assigned a "Moderate" Quality Rating

The Navigation Guide assigns *a priori* a "moderate" quality rating to the body of human observational evidence. This initial quality rating of "moderate" is independent of the specifics of the studies in the assessment. The actual quality of the body of human observational studies is then accounted for through upgrading or downgrading the "moderate" rating based on *a priori* criteria. In contrast, systematic reviews in the clinical sciences, which proceed from the availability of human experimental evidence, assign an *a priori* rating to the body of human observational studies of "low" quality. In particular, Cochrane and GRADE have been developed primarily based on evaluation of randomized controlled clinical trials (RCTs), and in this context, relative to RCTs, GRADE considers human observational studies to be "low" quality evidence (Balshem et al. 2011).

Our rationale to assign the body of human observational studies a rating of "moderate" and not "low" quality was based on the absolute and relative merit of human observational data in evidence-based decision-making in environmental and clinical health sciences. Overall, human observational studies are recognized as being a reliable source of evidence in the clinical sciences, as not all healthcare decisions are, or can be, based on RCTs. The contribution of observational studies to certain healthcare decisions is underscored by the conclusion of a 2008 Institute of Medicine (IOM) panel which found observational studies to be the preferred method for evaluating the causes of disease, which would include the contribution of environmental agents. The IOM panel noted that observational and experimental studies each can provide valid and reliable evidence, with their relative value dependent on the clinical question (Institute of Medicine et al. 2008). In this context, the IOM report states, "observational studies are generally the most appropriate for answering questions related to prognosis, diagnostic accuracy,

incidence, prevalence, and etiology" (Institute of Medicine et al. 2008). Moreover, recognition of the absolute value of human observational data to evidence-based clinical decision-making is increasing. Among the reasons for this is that the speed and complexity with which new medical interventions and scientific knowledge are being created make it unlikely that the evidence base required for treatment and cost effective health care delivery across subpopulations can be built using only RCTs (Peterson 2008). It is also expected that electronic medical records will revolutionize medical research by facilitating comprehensive longitudinal observational data in an instant (Halvorson 2008). Finally, ethical considerations virtually preclude experimental human data from the environmental health evidence stream. So, relative to the evidence available for decision-making in environmental health, human observational studies are the "gold standard" of the evidence base.

2. Diverse Evidence Streams Are Combined

In vitro, *in vivo, in silico* and human observational studies all inform decision-making on environmental chemical exposures. However, there is currently no agreed upon standard method in clinical medicine for evaluating evidence simultaneously across disparate evidence streams. We therefore adapted a mixture of IARC's method for integrating human and non-human evidence (International Agency for Research on Cancer 2006), linked to strength of evidence descriptions in use by USEPA (U.S. Environmental Protection Agency 1991, 1996). While this transparently produced a clear, concise and recognizable bottom line, i.e., "known to be toxic", "probably toxic", "possibly toxic", "not classifiable", or "probably not toxic", further development of precise criteria, definitions and nomenclature for strength of evidence that meets the needs of a wide-range of decision-makers will be an important undertaking as uptake of methodology moves forward.

Comparison of the Navigation Guide Method to Previous Reviews of PFOA and Fetal Growth

The authors of the review conducted with the Navigation Guide methodology concluded that "developmental exposure to PFOA adversely affects human health based on sufficient evidence of decreased fetal growth in both human and non-human mammalian species" (Lam et al. 2014). To compare these results to previous reviews, we searched PubMed without date or language restrictions for reviews of "PFOA" or "pefluorooctanoic acid". Of the 48 papers identified, 12 included discussions of reproductive or developmental health. Two additional reviews (Butenhoff et al. 2006; Stahl et al. 2011) were identified at the time we were embarking on this project and we also included those publications. Of 14 reviews, all but one (Stahl et al. 2011) which was not indexed in PubMed were also identified by our search strategy for the PFOA case study (Johnson PI et al. 2014).

Table 1 compares the 14 reviews of PFOA exposure and toxicity identified by our search to seven key features of systematic and transparent review methods, i.e., Cochrane and GRADE. All 14 reviews were conducted using non-systematic, expert-based narrative methods. Of the 14 reviews, 13 defined a study question, 9 included a summary of findings table, 3 specified criteria for included studies, 2 included limited information about their search strategy, 2 conducted data analysis, and 1 assessed the quality of individual studies. None of the 14 reviews systematically or transparently assessed risk of bias for individual studies and none integrated human and non-human evidence to produce an overall summary of the strength of the evidence (Butenhoff et al. 2004; DeWitt et al. 2009; Hekster et al. 2003; Jensen and Leffers 2008; Kennedy et al. 2004; Kudo and Kawashima 2003; Lau et al. 2004; Lau et al. 2007; Lindstrom et al. 2011; Olsen et al. 2009; Post et al. 2012; Stahl et al. 2011; Steenland et al. 2010; White et al. 2011). The 14

reviews produced vague or indeterminate answers to the question of PFOA's toxicity, or, presented a clear answer, i.e., "PFOA is a known developmental toxicant" (White et al. 2011) without specifying the search methods, study inclusion criteria, or statistical methods that produced the answer. Our comparison of the methods and results of these narrative reviews to the Navigation Guide method demonstrated that the application of the Navigation Guide provided more transparency about the steps taken in the review and a consistent path to a clear answer compared to the methods of expert-based narrative review that are currently employed in environmental health. Our results demonstrated that improved methods of research synthesis under-development at the National Toxicology Program (Birnbaum et al. 2013) and under consideration by the U.S. Environmental Protection Agency (National Research Council 2011; U.S. Environmental Protection Agency 2013a) are fully achievable.

Limitations

A limitation of the Navigation Guide systematic review method is that while its overall architecture is based on empirically proven and/or time-tested methods, i.e., methods in use by Cochrane, GRADE, IARC and EPA, novel aspects of the method need further development and validation including: rating the quality and strength of non-mammalian animal, *in vitro* and *in silico* evidence streams; consensus risk of bias domains for human observational studies and non-human studies; well-defined, measurable evidentiary bars for the factors used to downgrade the quality of environmental health evidence, i.e., indirectness, inconsistency, imprecision, and publication bias and for upgrading human evidence, i.e., dose-response and magnitude of effect; and exploring whether it makes a difference to the final quality rating if we assign the entire body of human observational studies a "moderate" rating and then downgrade for lesser quality study designs, or, as proposed in the NTP's framework, we assign different types of human

observational studies different ratings from the start, i.e., cross-sectional studies, case-control, and case series or reports are rated as "low" quality and cohort and nested case-control studies are rated as "moderate" quality (Birnbaum et al. 2013). Improved statistical tools for data analysis and integration will also advance the application of systematic review methods in environmental health. Whether the use of our nomenclature for the final strength of evidence ratings, i.e., "known to be toxic", "possibly toxic", etc., will be useful to decision-makers is also untested, and consensus methods for classifying strength of evidence for non-cancer health outcomes is a critical research and policy need (Gee 2008).

In addition, the application of the Navigation Guide method --- just like any expert-based narrative review --- can be poorly executed. For example, a systematic review can be conducted that does not specify a study question relevant to decision-making, or an incomplete search strategy can fail to uncover information pertinent to the review. However, a poorly performed systematic review is more readily detected since the methods are transparently displayed.

The capacity for improved methods of research synthesis in environmental health to spur timely health protective decision-making is also limited by the shortcomings of the available evidence stream that is produced by current systems of scientific knowledge generation. One key example is the need for an un-conflicted underlying evidence stream. As the Deputy Editor (West) of *JAMA* observed in 2010, "the biggest threat to [scientific] integrity [is] financial conflict of interest" (Rennie 2010). Moreover, risk of bias assessments leave unaddressed the inherent biases in environmental health science methodologies that generate false negatives and rely on strength of evidence criteria that are unequal to the task of addressing complex and multi-causal disease etiologies (Gee 2008). Finally, there are many other formidable non-scientific, social and

political barriers to prevention-oriented action (European Environment Agency 2012; Michaels 2008).

Future Directions

Shortening the time between scientific discovery and the prevention of exposures to toxic environmental chemicals is inextricably linked to the success of private and public sector efforts to advance safer and sustainable alternatives to toxic chemicals. The assessment of toxicity is an essential underpinning of such efforts (Edwards and The Lowell Center for Sustainable Production 2009; Malloy et al. 2013; Matus et al. 2012; Park et al. 2014; U.S. Environmental Protection Agency 2012). As such, the Navigation Guide methodology has broad applicability to support efforts by businesses, governments and consumers to compare and choose among various chemicals using a standardized and rigorous method.

As in the clinical application of systematic reviews, development of systematic and transparent methods of research synthesis in environmental health will be an on-going process. Some immediate methodological needs relate to how to routinely integrate critical concepts into the interpretation of data, including low dose effects; concordance in response across species; and human variability, including age and comorbidities. These issues were considered in the PFOA case study PECO question and statistical analyses, but a more thorough and overarching framework for how to integrate these concepts in systematic reviews is still needed. For example, failure to use animals with clinically relevant co-morbidities, such as hypertension in stroke models, has been shown to bias the assessment of drug efficacy (Macleod et al. 2008; Sena et al. 2010) and we would expect that including animals with chronic conditions may affect findings for environmental chemicals. Robust methods to assess publication bias in environmental health

science are also a need, as researchers can have financial and or other conflicts that can promote bias in opposite directions.

Uptake of methods of systematic and transparent review represents a new way of doing business in environmental health sciences. A realistic starting place is to recognize the potential for many or all the challenges related to using systematic reviews in clinical medicine to become our challenges, i.e., perceived threats to physician autonomy, patient choice, etc. We will need to overcome a lack of knowledge of environmental health science and research synthesis methods by every key target audience. The application of systematic reviews in environmental health is inherently an inter-disciplinary "team science" undertaking, and success will require formalizing the necessary expertise, and assembling and training review teams in these new methods and relevant communication skills.

Conclusion

Systematic and transparent methods of research synthesis are empirically based and can serve as a roadmap to more efficient and transparent decision-making using the available data. The use of systematic review methods allows decision-makers to act on any quality of evidence and in any direction. Moreover, the use of systematic reviews can prevent wasteful expenditures on studies that are duplicative or otherwise unnecessary for decision-making (Chalmers and Glasziou 2009).

In his 1965 address to the Royal Society of Medicine, Sir Austin Bradford Hill, the statistician who pioneered the RCT, admonished his audience that while science is always incomplete and subject to change, it "*does not confer upon us a freedom to ignore the knowledge we already have, or to postpone the action that it appears to demand at a given time"* (Hill 1965). Hill

emphasized that “strong evidence” does not imply “crossing every 't', and swords with every critic, before we act” (Hill 1965). He proposed differential standards of evidence for different actions, a recommendation echoed by the NAS a half-century later in *Science and Decisions* (National Research Council 2009).

Because systematic review methods transparently distinguish between science, values and preferences they can help sharpen the terms of debates as to whether we strive for more precision or more decisions about the meaning of the science to health.

This first case study of the Navigation Guide methodology demonstrated the successful application of a systematic and rigorous method for research synthesis designed to optimize transparency and reduce bias in the evaluation of environmental health information. Government agencies can use the Navigation Guide methodology to craft evidence-based statements regarding the relationship between an environmental exposure and health (Steps 1-3). Government agencies called on to make risk management decisions can also apply Step 4 of the Navigation Guide to grade the strength of recommendations for prevention. Professional societies, healthcare organizations, and other potential guideline developers working with toxicologists can use the Navigation Guide to craft consistent and timely recommendations to improve patient, and ultimately population, health outcomes (Steps 1-4). The institutionalization of robust methods of systematic and transparent review would provide a concrete mechanism for linking science to timely action to prevent harm. While simple in concept, it will require sustained visionary leadership harnessed to substantive investment, and the intellectual curiosity and commitment of environmental and clinical health scientists and advocates.

References

Antman EM, Lau J, Kupelnick B, Mosteller F, Chalmers TC. 1992. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts. Treatments for myocardial infarction. JAMA 268:240-248.

Balshem H, Helfand M, Schunemann HJ, Oxman AD, Kunz R, Brozek J, et al. 2011. Grade guidelines: 3. Rating the quality of evidence. J Clin Epidemiol 64:401-406.

Barnes DE, Bero LA. 1997. Scientific quality of original research articles on environmental tobacco smoke. Tob Control 6:19-26.

Barnes DE, Bero LA. 1998. Why review articles on the health effects of passive smoking reach different conclusions. JAMA 279:1566-1570.

Bebarta V, Luyten D, Heard K. 2003. Emergency medicine animal research: Does use of randomization and blinding affect the results? Acad Emerg Med 10:684-687.

Bellinger DC. 2012. Comparing the population neurodevelopmental burdens associated with children's exposures to environmental chemicals and other risk factors. Neurotoxicology 33:641-643.

Bero L, Oostvogel F, Bacchetti P, Lee K. 2007. Factors associated with findings of published trials of drug-drug comparisons: Why some statins appear more efficacious than others. PLoS Med 4:e184.

Bero L. 2013. Why the cochrane risk of bias tool should include funding source as a standard item [editorial]. Cochrane database of systematic reviews 12:ED000075.

Beronius A, Rudén C, Håkansson H, Hanberg A. 2010. Risk to all or none?:: A comparative analysis of controversies in the health risk assessment of bisphenol a. Reprod Toxicol 29:132-146.

Birnbaum LS, Thayer KA, Bucher JR, Wolfe MS. 2013. Implementing systematic review at the national toxicology program: Status and next steps. Environ Health Perspect 121:a108.

Butenhoff JL, Gaylor DW, Moore JA, Olsen GW, Rodricks J, Mandel JH, et al. 2004. Characterization of risk for general population exposure to perfluorooctanoate. Regulatory toxicology and pharmacology : RTP 39:363-380.

Butenhoff JL, Olsen GW, Pfahles-Hutchens A. 2006. The applicability of biomonitoring data for perfluorooctanesulfonate to the environmental public health continuum. Environ Health Perspect 114:1776.

C8 Science Panel. 2011. Probable link evaluation of preterm birth and low birth weight. Available: http://www.c8sciencepanel.org/pdfs/Probable_Link_C8_Preterm_and_LBW_birth_5Dec2011.pdf [accessed 5 December 2011].

Centers for Disease Control and Prevention. 2009. Cdc recovery act funding: Description of funded activities. Available: http://www.cdc.gov/fmo/topic/recovery_act/ [accessed 6 February 2014].

Centers for Disease Control and Prevention. 2012. Fourth national report on human exposure to environmental chemicals, updated tables. Available: http://www.cdc.gov/exposurereport/ [accessed 6 February 2014].

Chalmers I, Glasziou P. 2009. Avoidable waste in the production and reporting of research evidence. Lancet 374:86-89.

DeWitt JC, Shnyra A, Badr MZ, Loveless SE, Hoban D, Frame SR, et al. 2009. Immunotoxicity of perfluorooctanoic acid and perfluorooctane sulfonate and the role of peroxisome proliferator-activated receptor alpha. Crit Rev Toxicol 39:76-94.

Edwards S, The Lowell Center for Sustainable Production. 2009. A new way of thinking: The lowell center framework for sustainable products. Lowell, MA:University of Massachusetts Lowell.

European Environment Agency. 2012. Late lessons from early warnings: Science, precaution and politics 1824-2011. (EEA Reports). Copenhagen:European Environment Agency.

Fox DM. 2010. The convergence of science and governance: Research, health policy, and american states. Berkeley, CA:University of California Press.

Frantzias J, Sena ES, Macleod MR, Al-Shahi Salman R. 2011. Treatment of intracerebral hemorrhage in animal models: Meta-analysis. Ann Neurol 69:389-399.

Gee D. 2008. Establishing evidence for early action: The prevention of reproductive and developmental harm. Basic & clinical pharmacology & toxicology 102:257-266.

Goldman LR, Silbergeld EK. 2013. Assuring access to data for chemical evaluations. Environ Health Perspect 121:149-152.

Guyatt GH, Oxman AD, Kunz R, Falck-Ytter Y, Vist GE, Liberati A, et al. 2008a. Going from evidence to recommendations. BMJ 336:1049-1051.

Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. 2008b. Grade: An emerging consensus on rating quality of evidence and strength of recommendations. BMJ 336:924-926.

Guyatt GH, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. 2011. Grade guidelines: 1. Introduction-grade evidence profiles and summary of findings tables. J Clin Epidemiol 64:383-394.

Halvorson GC. 2008. Electronic medical records and the prospect of real time evidence development. In: Evidence-based medicine and the changing nature of health care: 2007 iom annual meeting summary. Washington, DC:The National Academies Press.

Hekster FM, Laane RW, de Voogt P. 2003. Environmental and toxicity effects of perfluoroalkylated substances. Rev Environ Contam Toxicol 179:99-121.

Higgins JPT, Altman DG, Sterne JAC. 2011. Chapter 8: Assessing risk of bias in included studies. In: Cochrane Handbook for Systematic Reviews of Interventions Version 510 [Updated March 2011], (Higgins J, Green S, eds):The Cochrane Collaboration. Available from http://www.cochrane-handbook.org

Higgins JPT, Green S. 2011. Cochrane handbook for systematic reviews of interventions. Version 5.1.0 [updated March 2011].The Cochrane Collaboration. Available from http://www.cochrane-handbook.org

Hill AB. 1965. The environment and disease: Association or causation? Proc R Soc Med 58:295-300.

Hooijmans CR, Tillema A, Leenaars M, Ritskes-Hoitinga M. 2010. Enhancing search efficiency by means of a search filter for finding all studies on animal experimentation in pubmed. Lab Anim 44:170-175.

Howells DW, Sena ES, O'Collins V, Macleod MR. 2012. Improving the efficiency of the development of drugs for stroke. Int J Stroke 7:371-377.

Institute of Medicine, Eden J, Wheatley B, McNeil B, Sox H. 2008. Knowing what works in health care: A roadmap for the nation:National Academies Press.

International Agency for Research on Cancer. 2006. Iarc monographs on the evaluation of carcinogenic risks to humans: Preamble (amended january 2006). Lyon, France.

Jensen AA, Leffers H. 2008. Emerging endocrine disrupters: Perfluoroalkylated substances. Int J Androl 31:161-169.

Johnson PI, Sutton P, Atchley DS, Koustas E, Lam J, Sen S, et al. 2014. The navigation guide—evidence-based medicine meets environmental health: Systematic review of human evidence for PFOA effects on fetal growth. Environ Health Perspect; http://dx.doi.org/10.1289/ehp.1307893 [Online 25 June 2014].

Kennedy GL, Jr., Butenhoff JL, Olsen GW, O'Connor JC, Seacat AM, Perkins RG, et al. 2004. The toxicology of perfluorooctanoate. Crit Rev Toxicol 34:351-384.

Koustas E, Lam J, Sutton P, Johnson PI, Atchley DS, Sen S, et al. 2014. The navigation guide—evidence-based medicine meets environmental health: Systematic review of nonhuman evidence for PFOA effects on fetal growth. Environ Health Perspect; http://dx.doi.org/10.1289/ehp.1307177 [Online 25 June 2014].

Krauth D, Woodruff TJ, 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.

Kudo N, Kawashima Y. 2003. Toxicity and toxicokinetics of perfluorooctanoic acid in humans and animals. J Toxicol Sci 28:49-57.

Lam J, Sutton P, E K, Johnson P, Atchley DS, Sen S, et al. 2014. The navigation guide—evidence-based medicine meets environmental health: Integration of animal and human evidence for PFOA effects on fetal growth. Environ Health Perspect; http://dx.doi.org/10.1289/ehp.1307929 [Online 25 June 2014].

Landis SC, Amara SG, Asadullah K, Austin CP, Blumenstein R, Bradley EW, et al. 2012. A call for transparent reporting to optimize the predictive value of preclinical research. Nature 490:187-191.

Lau C, Butenhoff JL, Rogers JM. 2004. The developmental toxicity of perfluoroalkyl acids and their derivatives. Toxicol Appl Pharmacol 198:231-241.

Lau C, Anitole K, Hodes C, Lai D, Pfahles-Hutchens A, Seed J. 2007. Perfluoroalkyl acids: A review of monitoring and toxicological findings. Toxicol Sci 99:366-394.

Lexchin J, Bero LA, Djulbegovic B, Clark O. 2003. Pharmaceutical industry sponsorship and research outcome and quality: Systematic review. BMJ 326:1167-1170.

Lindstrom AB, Strynar MJ, Libelo EL. 2011. Polyfluorinated compounds: Past, present, and future. Environ Sci Technol 45:7954-7961.

Lundh A, Sismondo S, Lexchin J, Busuioc OA, Bero L. 2012. Industry sponsorship and research outcome. Cochrane Database of Systematic Reviews; doi: 10.1002/14651858.MR000033.pub2.

Macleod MR, O'Collins T, Howells DW, Donnan GA. 2004. Pooling of animal experimental data reveals influence of study design and publication bias. Stroke; a journal of cerebral circulation 35:1203-1208.

Macleod MR, van der Worp HB, Sena ES, Howells DW, Dirnagl U, Donnan GA. 2008. Evidence for the efficacy of nxy-059 in experimental focal cerebral ischaemia is confounded by study quality. Stroke 39:2824-2829.

Malloy TF, Sinsheimer PJ, Blake A, Linkov I. 2013. Use of multi - criteria decision analysis in regulatory alternatives analysis: A case study of lead free solder. Integrated Environmental Assessment and Management 9:652-664.

Matus KJ, Clark WC, Anastas PT, Zimmerman JB. 2012. Barriers to the implementation of green chemistry in the united states. Environmental Science & Technology 46:10892-10899.

McPartland JM, Glass M, Pertwee RG. 2007. Meta-analysis of cannabinoid ligand binding affinity and receptor distribution: Interspecies differences. Br J Pharmacol 152:583-593.

Michaels D. 2008. Doubt is their product: How industry's assault on science threatens your health. New York, NY:Oxford University Press.

National Research Council. 2008. Phthalates and cumulative risk assessment: The task ahead. Washington, D.C.:National Academies Press.

National Research Council. 2009. Science and decisions: Advancing risk assessment. Washington, D.C.:National Academies Press.

National Research Council. 2011. Review of the environmental protection agency's draft iris assessment of formaldehyde. Washington, D.C.:National Academies Press.

National Research Council. 2014. Review of the environmental protection agency's state-of-the-science evaluation of nonmonotonic dose–response relationships as they apply to endocrine disruptors. Washington, DC.

National Research Council. 2014 Review of EPA's Integrated Risk Information System (IRIS) process. Washington, DC.

National Toxicology Program. 2013. OHAT nominations under consideration and evaluations. Available: http://ntp.niehs.nih.gov/?objectid=497BF6E6-D00C-C4E6-423E8917D64B6A20 [accessed September 24 2013].

Newbold R, Heindel J. 2010. Developmental exposures and implications for early and latent disease. In: Environmental impacts on reproductive health and fertility, (Woodruff TJ JS, Guillette Jr LJ, Giudice LC, ed). Cambridge, UK:Cambridge University Press, 92-102.

Olden K, Freudenberg N, Dowd J, Shields AE. 2011. Discovering how environmental exposures alter genes could lead to new treatments for chronic illnesses. Health Aff (Millwood) 30:833-841.

Olsen GW, Butenhoff JL, Zobel LR. 2009. Perfluoroalkyl chemicals and human fetal development: An epidemiologic review with clinical and toxicological perspectives. Reprod Toxicol 27:212-230.

Park SJ, Ogunseitan OA, Lejano RP. 2014. Dempster-shafer theory applied to regulatory decision-making for safer alternatives to toxic chemicals in consumer products. Integrated Environmental Assessment and Management 10:12-21.

Peterson ED. 2008. Research methods to speed the development of better evidence- the registries example. In: Evidence-based medicine and the changing nature of health care:2007 iom annual meeting summary. Washington, DC:The National Academies Press.

Popelut A, Valet F, Fromentin O, Thomas A, Bouchard P. 2010. Relationship between sponsorship and failure rate of dental implants: A systematic approach. PLoS One 5:e10274.

Post GB, Cohn PD, Cooper KR. 2012. Perfluorooctanoic acid (pfoa), an emerging drinking water contaminant: A critical review of recent literature. Environ Res 116:93-117.

Rennie D, Chalmers I. 2009. Assessing authority. JAMA 301:1819-1821.

Rennie D. 2010. Integrity in scientific publishing. Health Serv Res 45:885-896.

Rooney AA, Boyles AL, Wolfe MS, Bucher JR, Thayer KA. 2014. Systematic review and evidence integration for literature-based environmental health science assessments. Environ Health Perspect.

Roseman M, Milette K, Bero LA, Coyne JC, Lexchin J, Turner EH, et al. 2011. Reporting of conflicts of interest in meta-analyses of trials of pharmacological treatments. JAMA 305:1008-1017.

Sena ES, Briscoe CL, Howells DW, Donnan GA, Sandercock PA, Macleod MR. 2010. Factors affecting the apparent efficacy and safety of tissue plasminogen activator in thrombotic occlusion models of stroke: Systematic review and meta-analysis. Journal of cerebral blood flow and metabolism : official journal of the International Society of Cerebral Blood Flow and Metabolism 30:1905-1913.

Shah RV, Albert TJ, Bruegel-Sanchez V, Vaccaro AR, Hilibrand AS, Grauer JN. 2005. Industry support and correlation to study outcome for papers published in spine. Spine (Phila Pa 1976) 30:1099-1104; discussion 1105.

Stahl T, Mattern D, Brunn H. 2011. Toxicology of perfluorinated compounds. Environmental Sciences Europe 23:1-52.

Steenland K, Fletcher T, Savitz DA. 2010. Epidemiologic evidence on the health effects of perfluorooctanoic acid (pfoa). Environ Health Perspect 118:1100-1108.

Trasande L, Liu Y. 2011. Reducing the staggering costs of environmental disease in children, estimated at $76.6 billion in 2008. Health Aff (Millwood) 30:863-870.

Tsai PL, Hatfield TH. 2011. Global benefits from the phaseout of leaded fuel. J Environ Health 74:8-14.

U.S. Department of Health and Human Services. 2006. The health consequences of involuntary exposure to tobacco smoke: A report of the surgeon general. Atlanta, GA:U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, Coordinating Center for Health Promotion, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health.

U.S. Environmental Protection Agency. 1991. Guidelines for developmental toxicity risk assessment. Available: http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=23162#Download [accessed 27 February 2014].

U.S. Environmental Protection Agency. 1996. Guidelines for reproductive toxicity risk assessment. Available: http://www.epa.gov/raf/publications/pdfs/REPRO51.PDF [accessed 27 February 2014].

U.S. Environmental Protection Agency. 2011. The benefits and costs of the clean air act from 1990 to 2020. Available: http://www.epa.gov/air/sect812/feb11/fullreport_rev_a.pdf [accessed 6 February 2014].

U.S. Environmental Protection Agency. 2012. Design for the environment master criteria for safer ingredients. Available: http://epa.gov/dfe/pubs/projects/gfcp/index.htm#Master [accessed 6 February 2014].

U.S. Environmental Protection Agency. 2013a. Systematic review workshop. Available: http://www.epa.gov/IRIS/irisworkshops/systematicreview/ [accessed May 21 2013].

U.S. Environmental Protection Agency. 2013b. Air quality: Epa's integrated science assessments Available: http://www.epa.gov/ncea/isa/ [accessed 24 September 2013].

U.S. Environmental Protection Agency. 2013c. America's children and the environment. Available: http://www.epa.gov/ace/ [accessed 6 February 2014].

van der Worp HB, Sena ES, Donnan GA, Howells DW, Macleod MR. 2007. Hypothermia in animal models of acute ischaemic stroke: A systematic review and meta-analysis. Brain : a journal of neurology 130:3063-3074.

van der Worp HB, Macleod MR. 2011. Preclinical studies of human disease: Time to take methodological quality seriously. J Mol Cell Cardiol 51:449-450.

Vesterinen HM, Sena ES, ffrench-Constant C, Williams A, Chandran S, Macleod MR. 2010. Improving the translational hit of experimental treatments in multiple sclerosis. Mult Scler 16:1044-1055.

Vesterinen HM, Egan K, Deister A, Schlattmann P, Macleod MR, Dirnagl U. 2011. Systematic survey of the design, statistical analysis, and reporting of studies published in the 2008 volume of the journal of cerebral blood flow and metabolism. Journal of cerebral blood flow and metabolism : official journal of the International Society of Cerebral Blood Flow and Metabolism 31:1064-1072.

Viswanathan M, Ansari M, Berkman N, Chang S, Hartling L, McPheeters L, et al. 2012. Assessing the risk of bias of individual studies in systematic reviews of health care interventions. (Agency for Healthcare Research and Quality Methods Guide for Comparative Effectiveness Reviews). AHRQ Publication No. 12-EHC047-EF.

Vogel SA, Roberts JA. 2011. Why the toxic substances control act needs an overhaul, and how to strengthen oversight of chemicals in the interim. Health Aff (Millwood) 30:898-905.

White SS, Fenton SE, Hines EP. 2011. Endocrine disrupting properties of perfluorooctanoic acid. The Journal of Steroid Biochemistry and Molecular Biology 127:16-26.

Woodruff TJ, Janssen SJ, Guillette LJ, Jr., Giudice LC. 2010. Environmental impacts on reproductive health and fertility. New York, NY:Cambridge University Press.

Woodruff TJ, Sutton P, The Navigation Guide Work Group. 2011. An evidence-based medicine methodology to bridge the gap between clinical and environmental health sciences. Health Aff (Millwood) 30:931-937.

World Health Organization, United Nations Environment Programme. 2013. Sate of the science of endocrine disrupting chemicals - 2012. Available: http://www.who.int/ceh/publications/endocrine/en/ [accessed 6 February 2014].

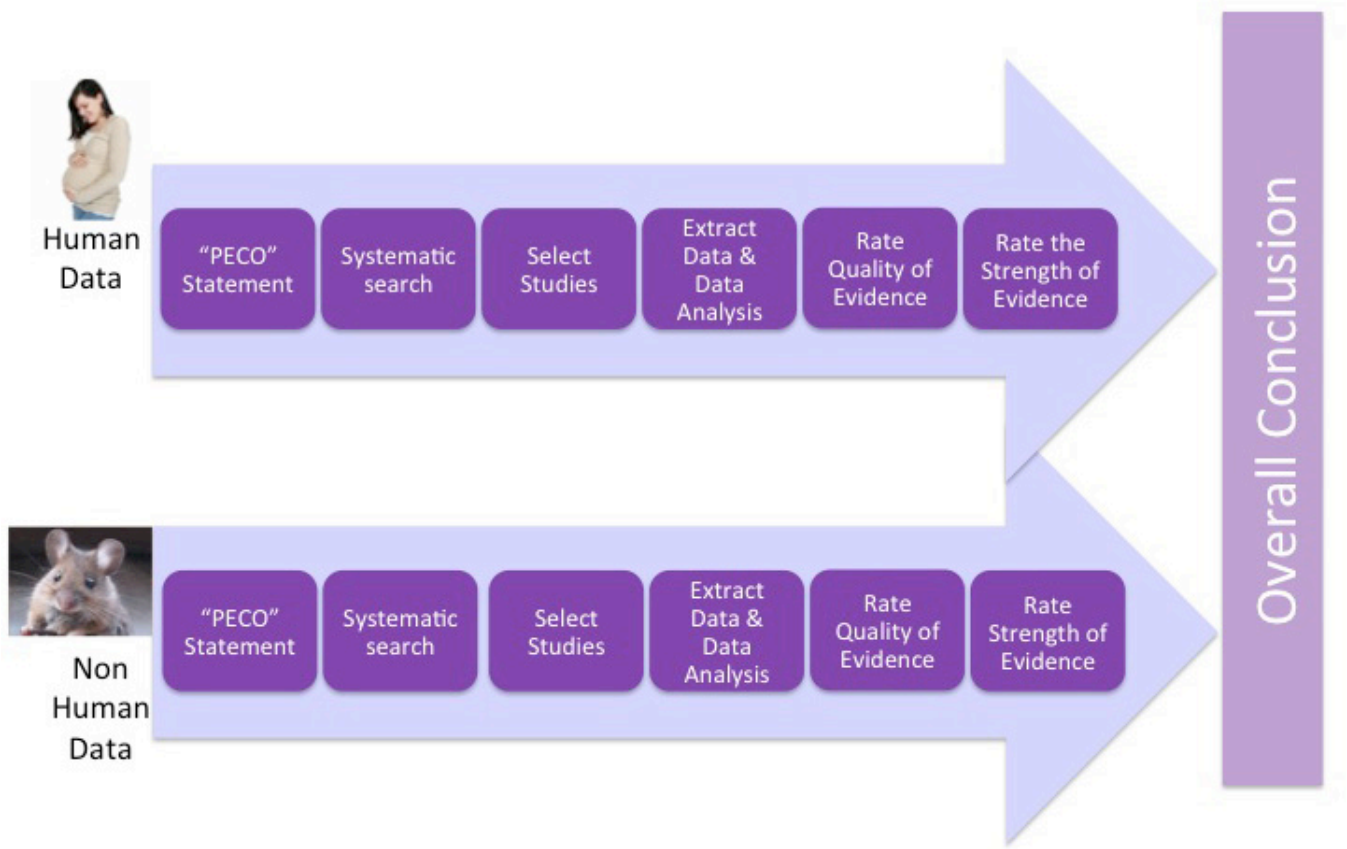
Table 1. Comparison of PFOA review methods according to key features of Cochrane and GRADE systematic and transparent review methods.

Reference	Specify study question	Specify inclusion/exclusion criteria	Conduct reproducible search	Assess 'Risk of Bias'	Data analysis and/or meta-analyses	Summary of findings table	Assess quality and strength of body of evidence
Navigation Guide PFOA Case Study 2013	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Post et al 2012	Yes	No	No	No	Some data analysis (BMD, BMDL)	Yes	No
Lindstrom et al 2011	Yes	No	No	No	No	No	No
Stahl et al 2011	Yes	No	No	No	No	Yes	No
White et al 2011	Yes	No	No	No	No	Yes	No
Steenland et al 2010	Yes	No	No	No	No	Yes	No
DeWitt et al 2009	Yes	No	No	No	No	No	No
Olsen et al 2009	Yes	Inclusion criteria	No	No	No	Yes	Assess methodological weaknesses of included studies
Jensen and Leffers 2008	No	No	No	No	No	No	No
Lau et al 2007	Yes	No	No	No	No	Yes	No
Butenhoff et al 2004	Yes	Yes	No	No	Some data analysis (MOE, $LBMIC_{10}$)	Yes	No
Kennedy et al 2004	Yes	No	No	No	No	Yes	No
Lau et al 2004	Yes	No	Limited discussion of literature search	No	No	No	No
Hekster et al 2003	Yes	Some inclusion criteria described in cited report by same authors	Limited discussion of literature search	No	No	Yes	No
Kudo and Kawashima 2003	Yes	No	No	No	No	No	No

Figure Legend

Figure 1. Steps in the Navigation Guide protocol.

Figure 1.


Human Data
"PECO" Statement
Systematic search
Select Studies
Extract Data & Data Analysis
Rate Quality of Evidence
Rate the Strength of Evidence
Non Human Data
"PECO" Statement
Systematic search
Select Studies
Extract Data & Data Analysis
Rate Quality of Evidence
Rate Strength of Evidence
Overall Conclusion