



Food for Thought ...

How Evidence-Based Methodologies Can Help Identify and Reduce Uncertainty in Chemical Risk Assessment

Sebastian Hoffmann¹, Paul Whaley^{1,2} and Katya Tsaïoun¹

¹Evidence-based Toxicology Collaboration (EBTC) at Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA; ²Lancaster Environment Centre, Lancaster University, Lancaster, UK

Abstract

Evidence-based methodology, in particular systematic review, is increasingly being applied in environmental, public, and occupational health to increase the transparency, comprehensiveness, and objectivity of the processes by which existing evidence is gathered, assessed, and synthesized in answering research questions. This development is also changing risk assessment practices and will impact the assessment of uncertainties in the evidence for risks to human health that are posed by exposure to chemicals. The potential of evidence-based methodology for characterizing uncertainties in risk assessment has been widely recognized, while its contribution to uncertainty reduction is yet to be fully elucidated. We therefore present some key aspects of the evidence-based approach to risk assessment, showing how they can contribute to the identification and the assessment of uncertainties. We focus on the pre-specification of an assessment methodology in a protocol, comprehensive search strategies, study selection using predefined eligibility criteria, critical appraisal of individual studies, and an evidence integration and uncertainty characterization process based on certainty of evidence frameworks that are well-established in health care research. We also provide examples of uncertainty in risk assessment and discuss how evidence-based methodology could address those. This perspective, which neither claims to be comprehensive nor complete, is intended to stimulate discussion of the topic and to motivate detailed exploration of how evidence-based methodology contributes to characterization of uncertainties, and how it will lead to uncertainty reduction in the conduct of health risk assessment.

1 Introduction

Risk assessment of chemicals is increasingly stimulating scientific debate and subsequently public interest, especially when discussed controversially. A lack of transparency has been identified as a major reason for such controversies (Schreider et al., 2010). In addition, increased objectivity and scientific rigor in risk assessment methodology have been called for (NRC, 2014). The potential of evidence-based approaches, as exemplified in systematic review methods, to address those shortcomings has been recognized by various institutions, including governmental agencies such as the World Health Organization, the European Food Safety Authority (EFSA), the US Environmental Protection Agency, and other research organizations, e.g., the US National Institute for Environmental Health Sciences and the Evidence-based Toxicology Collaboration. Such approaches have been developed to provide a crit-

ical summary of what often complex evidence is actually saying in response to a research question, to render processes more transparent, objective and reproducible, and to move away from tradition-based, subjective and sometimes poorly justified decisions, most of which are associated with an unknown level of uncertainty (Eddy, 2005). Evidence-based methodologies are well established in healthcare, but also in other areas such as social, education and environmental sciences, where they serve as the foundation of evidence-based practice and policy decisions. Application and adaptation of evidence-based methodology to the assessment of health risks of chemicals is being pioneered by a steadily growing community (EFSA, 2010; Hoffmann et al., 2017; Rooney et al., 2014; Stephens et al., 2013, 2016; Thayer et al., 2014; Whaley et al., 2016; Woodruff and Sutton, 2014).

Consideration of uncertainties is an integral element of such assessments. By “uncertainty” we mean a lack of confidence in a

Received January 13, 2022; Accepted January 31, 2022;
Epub January 31, 2022; © The Authors, 2022.

ALTEX 39(2), 175-182. doi:10.14573/altex.2201131

Correspondence: Katya Tsaïoun, PhD
Evidence-based Toxicology Collaboration (EBTC)
at Johns Hopkins Bloomberg School of Public Health
615 N Wolfe St., Baltimore, MD, 21205, USA
(ktsaiou1@jhu.edu)

This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 International license (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is appropriately cited.



knowledge claim, primarily due to other informational deficiencies, which are context and time-dependent (EFSA Scientific Committee, 2018b).

In 2018, EFSA provided a general and broad list of uncertainty types relevant to risk assessment, to which we refer here to illustrate our perspective (EFSA Scientific Committee, 2018a). The EFSA guidance identified and grouped common sources of uncertainty into those associated with assessment input and those associated with assessment methodology.

Of the eight specific types of uncertainty related to the inputs (EFSA Scientific Committee, 2018a), ambiguity, i.e., the quality of allowing more than one data interpretation, measurement accuracy and precision, missing studies, missing data within studies, as well as extrapolation uncertainty are addressed here. Sampling uncertainty, assumptions about inputs, and statistical estimates appear less relevant in our context. Extrapolation uncertainty, which subsumes sources of uncertainty such as inter- and intra-species extrapolation, exposure route, and exposure duration, is integral to chemical risk assessment. Depending on the available evidence, the respective uncertainty is typically accounted for in a chemical-specific approach or by default uncertainty (or assessment) factors to determine toxicity reference values of a chemical to be compared to exposure (Dankovic et al., 2015; NRC, 2014).

Of the ten specific uncertainties that EFSA associated with assessment methodology, we address only uncertainties introduced by the methods for processing evidence from the literature and uncertainties associated with expert judgement, especially when integrating frequently heterogeneous and inconsistent evidence (EFSA Scientific Committee, 2018a). The latter type may affect the risk assessment result, such as when it is biased by disregarded evidence which potentially undermines its acceptance.

In addition to the uncertainties related to evidence and assessment methods, uncertainties related to a specific assessment context need to be considered. For example, a lack of transparency and objectivity or the perception of potential conflicts of interest may create further potential uncertainties and impact on stakeholder trust in the assessment process as occurred, for example, in the complex controversies surrounding the carcinogenicity assessment of glyphosate (see, e.g., Kogevinas, 2019, including the rapid response by Kabat, 2019, and analysis by Robinson et al., 2020) and the risk assessment of aspartame (Kass and Lodi, 2020; Millstone and Dawson, 2019), which both provide ample illustration of stakeholder concerns in this area.

The assessment of uncertainties, i.e., their identification, prioritization and qualification or quantification, can be conducted in various manners. Uncertainties related to the evidence are often based on generic assumptions, e.g., as specified in (prescriptive) guidance documents. More chemical-specific uncertainties are being explored by employing comprehensive and systematic approaches that also consider methodological aspects (e.g., Schenk and Johanson, 2011; Beck et al., 2016; Bhat et al., 2017; Wikoff et al., 2020a). The potential of evidence-based approaches for the assessment and reduction of uncertainty has been recognized (EFSA Scientific Committee, 2018b; Hoffmann et al., 2017; Wikoff et al., 2020b; Wolffe et al., 2019). For example, in their proposal of an evidence-based risk assessment framework, Wikoff et al. (2020b)

reflected upon input-related uncertainty and how evidence-based methods can inform its characterization.

We intend here to further stimulate the exploration of the potential of how evidence-based methodology, focusing particularly on the use of systematic review methods as a component of EBT approaches, can contribute to uncertainty assessment. Because uncertainties in risk assessment are manifold, we select those that in our opinion are particularly interesting or instructive and do not claim to present a comprehensive or complete view of whether or how EBT can resolve all uncertainties.

2 Evidence-based approaches and uncertainty

Guided by the systematic review steps described by Hoffmann et al. (2017), we discuss how evidence-based approaches to specific risk assessment tasks can help identify and make apparent various sources of uncertainty. This may eventually lead to the reduction of important uncertainties relating to risk assessment, and improved qualification and characterization of the confidence in the evidence that supports risk assessment conclusions and subsequent risk management intervention.

2.1 Protocol

The pre-specification of the methodology to conduct research is pivotal to ensure transparency of the process and to reduce bias (Whaley et al., 2020a). This is a required step in the systematic review process. In chemical risk assessment, comprehensive documentation of a thorough planning process, issued in the form of a protocol, allows for decisions to be planned in advance of seeing the evidence in detail. This helps reassure stakeholders that the results of an SR are derived from methods, rather than methods being determined ad-hoc that are at risk of being fitted around expectations and opinions. Since some foreknowledge of data is inevitable on the part of domain experts involved in the review, protocols can minimize the risk that bias is introduced, e.g., by approaches favoring conclusions in line with the experts' expectations.

How does this work? An openly accessible protocol that is registered in advance of study conduct allows comparison of the applied methodology with what was originally planned. Examples of systematic review protocols for chemical risk assessment can increasingly be found in the peer-reviewed literature (e.g., Matta et al., 2019; van Luijk et al., 2019). The use of such protocols discourages ad-hoc decision-making that may lead to bias in a review, for example the omission of studies based on their results or changing the data analysis plan in such a way that it generates more favorable or interesting results. Conducting risk assessment using a protocol helps researchers to reflect on the potential bias introduced by a deviation and prompts them to provide a justification for amendments. In this way, uncertainty related to the review methodology can be made transparent, assessed, and possibly be reduced. A lack of transparency in inhalation toxicity reference values has, for example, been identified for chemicals regulated in the EU, leaving sometimes substantial differences in values derived by governmental agencies and industry unexplained (Schenk et al., 2015). More clarity on the study methods prior to conduct of the research may help address this.

Registration and publication of the protocol also allows external review of methods to be applied, helping to prevent critical errors in methods before a study is conducted. It informs stakeholders and enables them to provide feedback on methodology, potentially increasing the trust in the evidence and the decisions made based on it, by those affected by its findings. This early involvement of stakeholders potentially increases their engagement and consent with the risk assessment approach and subsequently its results (EFSA, 2015). In this way, the likelihood of post-hoc criticism, which may make the risk assessment appear less trustworthy, can potentially be lowered. In terms of the EFSA uncertainties, the development and publication of protocols prior to conduct of assessment has most impact on the assessment methodology uncertainties related to ambiguity, to the process for dealing with evidence from the literature, and to expert judgement. This is via definition of processes for dealing with evidence, including the use of expert opinion, being made transparent, justified, and improved upon in advance of conducting the assessment (e.g., EFSA et al., 2017).

2.2 Search for the evidence

A common criticism of chemical risk assessments is that not all available and pertinent evidence has been identified (Chvátalová, 2019; Deveau et al., 2015; Rudén, 2001; Schenk, 2010). The main reason for missing critical evidence is the way it is searched for, which may be influenced by subjectivity, as well as time and resources available for conducting the risk assessment. In addition, relevant evidence may be hard to find, particularly when not available in readily searchable databases. Such sources are commonly referred to as “gray literature” and include, inter alia, government reports, regulatory databases and conference proceedings. Efforts are underway to bring this evidence into the public domain and incorporate it into systematic reviews, building upon experience with gray literature in healthcare research (Mahood et al., 2014). EFSA considers the uncertainty associated with missing evidence as a methodology issue (EFSA Scientific Committee, 2018b). The potential for it can be limited by the use of systematic search strategies that comprehensively cover relevant literature sources and terms used for retrieving literature from those sources (Whaley et al., 2020b). While literature databases such as PubMed, Web of Science, Embase and BIOSIS previews are standard sources, the identification of relevant sources of gray literature usually requires domain and information retrieval expertise. In this way, the risk of missing relevant evidence, either due to the search approach or, worse, by selective choices, which possibly results in an evidence base that can be biased in either direction, can be reduced (Gusenbauer and Haddaway, 2021). While it can be challenging to implement a comprehensive search, e.g., due to the amount of evidence retrieved, demonstrating that (almost) all potentially relevant evidence has been identified and considered should increase trust in the findings of the risk assessment. Comprehensive search strategies are fundamental for evidence-based approaches and systematic reviews (Hoffmann et al., 2017), which have the additional advantage of being readily reproducible, amendable, and updateable. Such searches help address the input-related uncertainties of missing studies.

2.3 Selection of the evidence

Any search approach that is more comprehensive and objective than an expert-driven selection of to-be-considered studies will most likely retrieve studies that are not relevant for the risk assessment. If a comprehensive search strategy is conducted, the number of studies ultimately excluded will in most, if not all, cases be substantially higher than the number included (see, e.g., Dorman et al., 2018; Lam et al., 2014). Therefore, a selection process is needed. In systematic reviews, an evidence-based, i.e., transparent, *a priori* defined and reproducible, two-tier approach using eligibility criteria for inclusion and exclusion of studies is applied. The predefined criteria describe primarily the population and intervention/exposure of the underlying risk assessment question and are applied in a consistent manner. Text mining and machine learning tools are becoming increasingly available to assist the evidence selection (e.g., Marshall and Wallace, 2019; Howard et al., 2020).

In the first tier, studies retrieved through the search are reduced by excluding those that can be unambiguously identified as irrelevant based on their title and/or abstract. In the second tier, the remaining studies are assessed in detail for their relevance based on the full texts. The selection process, often using appropriate web-based software tools, is usually summarized using the PRISMA statement flow chart (Moher et al., 2015), while the details, e.g., the reasons for exclusion, are also documented, so that they are available if needed. The transparency established in this way reduces uncertainty caused by excluding relevant or including irrelevant evidence based on selective choices, which is one of the issues in the glyphosate controversy, where the studies/evidence considered in the three evaluations (IARC, Monsanto, EPA) differed (Benbrook, 2019), resulting in opposing conclusions being drawn. In particular, it provides a rationale why evidence that others might consider relevant was excluded. This opens the opportunity to systematically explore the effect of adjustments to the eligibility criteria on the risk assessment outcome. Ultimately, a transparent selection process increases trust because it insulates an assessment from criticism for being selective (see, e.g., Robinson et al., 2020). In this way, the definition of explicit eligibility criteria, and the screening of potentially relevant studies against those criteria, helps address uncertainties associated with missing studies and the process for dealing with evidence from the literature.

2.4 Assessment of the evidence

2.4.1 Critical appraisal of the individual studies

Once the relevant evidence has been selected, the next step in a chemical risk assessment is to understand the uncertainty inherent in the individual studies to determine how informative each study is with regard to developing the assessment findings. There are several sources of such uncertainty, comprising the level of detail in the reporting of the research, the internal validity of the study (i.e., the extent to which a study minimizes systematic errors or biases), the external validity of the study (i.e., the extent to which the findings of a study can be generalized to other circumstances) as well as other aspects, such as the appropriateness of statistics, conflict of interest, and study sensitivity (Cooper et al., 2016; Rooney et al., 2016; Samuel et al., 2016). The most fundamental issue is the reporting quality of studies. If the reporting of studies is incom-



plete or unclear, it becomes difficult to assess the other uncertainty sources (Schenk et al., 2015). Poor reporting especially affects the assessment of internal study validity, not least as authors, but also reviewers, are often not aware of study design, conduct and analysis details that relate to internal validity. Contacting authors for unreported information may resolve the issue but is likely to fail in many cases for several reasons, including failing to obtain contact details, poor response rates, time-constraints of the assessment, and incomplete or inappropriate record keeping (see, e.g., supplementary data of Li et al., 2020). In the recent WHO ILO (International Labour Organization) systematic reviews to support the estimation of global burden of disease from occupational environmental exposures, one third of requests for missing data were not met (Pega et al., 2021).

The internal validity of studies is the extent to which the results of a study are subject to systematic error, either over- or underestimating the true effect. Originally developed for randomized controlled trials (RCT) for health care interventions, empirical evidence demonstrating the effects of biases is available (Berkman et al., 2014; Schulz et al., 1995). While it is reasonable to assume that some biases relevant for RCT may also be relevant for the type of studies used in chemical risk assessment, empirical evidence supporting the adaptation is still scarce, as, e.g., observed by Bero et al. (2018) for observational exposure studies. Assessing internal study validity is further complicated by the fact that biases to be considered depend on the types of studies encountered for chemical risk assessment, comprising observational, *in vivo* and *in vitro* studies, and by the fact that broader agreement on approaches and tools still needs to emerge (Lynch et al., 2016; Samuel et al., 2016).

While systematic review methodology cannot on its own fix experimental issues with studies, a pre-defined and systematic approach to critical study appraisal results in a consistent and traceable assessment of each included study. In this way, it reduces uncertainty by providing a thorough and transparent quality assessment of all included evidence pieces that supports their interpretation and integration, thus avoiding subjective or inappropriate approaches. In addition, it reduces the likelihood of misleading the risk assessment by missing to identify biases in studies, e.g., when reasons for possible over- or under-estimation in studies go unnoticed. For example, Wikoff et al. (2018) found in a case study on trichloroethylene that the only oral exposure study that showed an effect also had the strongest risk of bias concern. They further stated that this study was considered in regulatory setting of reference values. Assuming that the biases in that study resulted in an overestimation of the effect, the confidence in the risk assessment outcome is decreased by uncertainties introduced by a flawed assessment of the individual study quality. While the interpretation of the evidence in this case study is certainly a matter of expert debate, the systematic assessment approach permits such debate by providing a well-defined and transparent basis for it.

Assessing the external validity of the pieces of evidence supports the elucidation of their relevance to the research question. Depending on the study quality assessment approach and tool, this aspect may be analyzed on the level of the individual study (NTP, 2015; see, e.g., Wikoff et al., 2019) or when considering the body of evidence. Examples for aspects to be considered for external validity are relevance of various administration routes in experimen-

tal animal studies for the human exposure covered by the risk assessment, or if the mode of action underlying the observed toxicological effects is relevant for humans, e.g., in terms of species differences. In many chemical risk assessment frameworks, such considerations are addressed in a weight-of-evidence (WoE) process when assessing all evidence, as elaborated below.

Critical appraisal of individual studies can help address uncertainty in assessment inputs, especially accuracy and precision of measures and missing data within studies, as well as with assessment methodology-related uncertainty of expert judgement.

2.4.2 Evidence integration

A broad spectrum of additional uncertainties becomes apparent when the evidence for a chemical risk assessment is considered across studies. At this stage, inconsistencies between individual studies, strains, species or cell types, the human relevance of non-human studies, and the characterization of effects in terms of size, precision, severity/adversity, reversibility, and plausibility are assessed (see, e.g., Rhomberg et al., 2013). This process is often termed WoE and used in many chemical risk assessment contexts and regulations. As definitions and descriptions of WoE differ, also the methodology applied and the processes differ (Martin et al., 2018; OECD, 2019). While some WoE approaches are well-characterized, others are less specific and allow for flexibility, which may lead to substantial differences in chemical risk assessments conducted in different regulatory contexts, even when based on the same evidence (Hassauer and Roosen, 2020).

The adoption and adaptation of evidenced-based methodology to chemical risk assessments has, at the level of the body of evidence to be integrated, led to exploring the application of the Grading of Recommendations Assessment, Development and Evaluation (GRADE) working group approach (Cano-Sancho et al., 2019; Koustas et al., 2014; NTP, 2016; Yost et al., 2019). The GRADE working group has developed an approach to rating certainty in the body of evidence, originally developed for health care questions, that considers several quality-determining factors: five potentially downgrading (risk of bias, inconsistency, indirectness, imprecision, and publication bias) and three potentially upgrading (large effect size, plausible confounders likely to reduce effect size, and dose-response gradient) (Guyatt et al., 2008; Schünemann et al., 2013). A pre-defined level of certainty in a body of evidence is then modified by applying each of the factors, resulting in an overall certainty rating that can be considered as a categorical integration of the uncertainties addressed by the various factors.

It has been suggested that these factors also should be applicable to the type of evidence encountered for chemical risk assessments, while acknowledging that additional guidance still needs to be developed (Morgan et al., 2016). Some GRADE factors can be associated with uncertainty sources related to the evidence, e.g., interspecies differences would fall under indirectness, unexplained intra-species differences would be covered by inconsistency, and study quality limitations would at least partly be addressed under risk-of-bias assessment. Others, especially publication bias, i.e., systematic differences between the findings of published and unpublished research, and possibly related biases, such as multiple publication and citation bias, are typically not considered in chemical risk assessment, so that their roles and impact remain largely to

Tab. 1: The contribution of evidence-based methodology to uncertainty assessment

| Evidence-based methodology | Contribution to assessment of uncertainty associated with | | | Relation to EFSA types of uncertainty |
|---|---|---|---|---|
| | Evidence/Input | Methodology | Context | |
| Predefinition of the risk assessment methodology in an openly accessible protocol | – discourages subjective selection of studies | – lower likelihood of introducing biases by subjective and unjustified decisions – avoidance of critical errors | – stakeholder engagement leading to a higher acceptance of the assessment outcome | – ambiguity – process for dealing with evidence from the literature – expert judgement |
| Comprehensive search | – n.a. | – reduction of uncertainty caused by missing relevant evidence or selective choices | – increased trust by identifying all relevant evidence | – missing studies |
| Selection using eligibility criteria | – n.a. | – reduction of uncertainty through selective choices | – increased trust by preventing criticism for being selective | – missing studies – process for dealing with evidence from the literature |
| Critical appraisal of individual studies | – lower likelihood of misleading the risk assessment by failing to identify biases in studies | – reduction of uncertainty due to insufficient quality assessment by a transparent and comprehensive appraisal approach | – n.a. | – accuracy and precision of measures – missing data within studies – expert judgement |
| Evidence integration | – structured identification of uncertainty, e.g., inconsistency and indirectness | – increased transparency and objectivity by applying a structured and consistent framework | – reduction of uncertainties due to inconsistent risk assessment outcomes | – accuracy and precision – missing studies – extrapolation uncertainty – process for dealing with evidence from the literature – expert judgement |

n.a., not applicable

be explored. Likewise, it needs to be explored if all sources of uncertainty considered when integrating the evidence in a chemical risk assessment can be accommodated in the GRADE framework.

Some of these challenges are addressed, for example, in the GRADE Environmental Health Project Group (Morgan et al., 2019), which explores how the concept of biological plausibility links to external validity and indirectness and how GRADE applies in the context of development of adverse outcome pathways (AOP) (De Vries et al., 2021; Whaley et al., 2022), a pragmatic, chemical-agnostic framework to describe our knowledge of causally linked events at different levels of biological organization that lead to adverse health effects. As a WoE approach to assess the causality in AOP, and also chemical-specific mode of action, is often based on modified Bradford-Hill considerations (Meek et al., 2014), a discussion is taking place on how the two approaches relate to each other (De Vries et al., 2021; Hoffmann et al., 2022). It is important to note that the GRADE framework itself has roots in Bradford Hill criteria and has been further developed, operationalized, and validated for healthcare research over the last decades (Schünemann et al., 2011).

Incorporation of a GRADE-like approach in chemical risk assessment would facilitate consistent, transparent and comprehensive assessment of various important and frequently encountered

sources of uncertainty. In addition, it could contribute to reducing methodological and contextual uncertainties due to different risk assessment outcomes when considering the same evidence, as, e.g., contributed to differences in the setting of an occupational exposure level for n-methyl-2-pyrrolidone in the European Union (RAC and SCOEL, 2016). While the use of GRADE does not eliminate uncertainty, it can help considerably in its characterization of a range of uncertainties across assessment inputs, especially the accuracy and precision of measured data, missing studies, extrapolation uncertainty, the process of dealing with evidence from the literature, and expert judgement.

3 Conclusions

Application and adaptation of evidence-based methodology, cumulating in systematic reviews, to chemical risk assessment is being pioneered by a steadily growing community. This is starting to change risk assessment practices, including the assessment of uncertainties related to the data. While this impact on uncertainty assessment has been acknowledged (EFSA Scientific Committee, 2018b; Hoffmann et al., 2017; Wikoff et al., 2020b; Wolffe et al., 2019), detailed evaluations of this connection are still to be con-



ducted. As a first step to further instruct such evaluations, we here present some considerations of how evidence-based methodology, in the form of systematic review methods, links and can contribute to the assessment of uncertainties related to the evidence, the methodology, and the context of a chemical risk assessment. We also provide examples of issues encountered in chemical risk assessment due to uncertainty and lack of confidence and explain how evidence-based methodology can make the sources of uncertainty apparent and reduce, and possibly avoid, the associated uncertainty. Evidence-based methodology holds potential to inform all three types of uncertainty, while methodological and contextual uncertainty can be reduced via the pre-definition of the assessment methodology in a protocol, a comprehensive search strategy, study selection using predefined eligibility criteria, critical appraisal of individual studies, and an evidence integration process adopting and adapting an operationalized framework (Tab. 1).

We note that our perspective is not complete. For example, we did not address the potential role the framing of the risk assessment questions, in particular following the PECO (population, exposure, comparator, outcome) framework, could play (EFSA, 2010; Morgan et al., 2018); rather, this commentary is intended as a starting point to explore the topic in greater detail to identify the opportunities it presents and also the limitations that it may entail. As the community of early adopters continues to work on the development of evidence-based methodology for chemical risk assessment questions, the Evidence-Based Toxicology Collaboration (EBTC) provides a platform for exchange and collaboration to all stakeholders to harmonize approaches and increase impact.

References

- Beck, N. B., Becker, R. A., Erraguntla, N. et al. (2016). Approaches for describing and communicating overall uncertainty in toxicity characterizations: U.S. Environmental Protection Agency's integrated risk information system (IRIS) as a case study. *Environ Int* 89-90, 110-128. doi:10.1016/j.envint.2015.12.031
- Benbrook, C. M. (2019). How did the US EPA and IARC reach diametrically opposed conclusions on the genotoxicity of glyphosate-based herbicides? *Environ Sci Eur* 31, 2. doi:10.1186/s12302-018-0184-7
- Berkman, N. D., Santaguida, P. L., Viswanathan, M. et al. (2014). The Empirical Evidence of Bias in Trials Measuring Treatment Differences. <https://www.ncbi.nlm.nih.gov/sites/books/NBK253184/>
- Bero, L., Chartres, N., Diong, J. et al. (2018). The risk of bias in observational studies of exposures (ROBINS-E) tool: Concerns arising from application to observational studies of exposures. *Syst Rev* 7, 242. doi:10.1186/s13643-018-0915-2
- Bhat, V. S., Meek, M. E. B., Valcke, M. et al. (2017). Evolution of chemical-specific adjustment factors (CSAF) based on recent international experience; increasing utility and facilitating regulatory acceptance. *Crit Rev Toxicol* 47, 729-749. doi:10.1080/10408444.2017.1303818
- Cano-Sancho, G., Ploteau, S., Matta, K. et al. (2019). Human epidemiological evidence about the associations between exposure to organochlorine chemicals and endometriosis: Systematic review and meta-analysis. *Environ Int* 123, 209-223. doi:10.1016/j.envint.2018.11.065
- Chvátalová, V. (2019). A critical evaluation of EFSA's environmental risk assessment of genetically modified maize MON810 for honeybees and earthworms. *Environ Sci Eur* 31, 52. doi:10.1186/s12302-019-0238-5
- Cooper, G. S., Lunn, R. M., Agerstrand, M. et al. (2016). Study sensitivity: Evaluating the ability to detect effects in systematic reviews of chemical exposures. *Environ Int* 92-93, 605-610. doi:10.1016/j.envint.2016.03.017
- Dankovic, D. A., Naumann, B. D., Maier, A. et al. (2015). The Scientific Basis of Uncertainty Factors Used in Setting Occupational Exposure Limits. *J Occup Environ Hyg* 12, Suppl 1, S55-S68. doi:10.1080/15459624.2015.1060325
- De Vries, R. B. M., Angrish, M., Browne, P. et al. (2021). Applying evidence-based methods to the development and use of adverse outcome pathways. *ALTEX* 38, 336-347. doi:10.14573/altex.2101211
- Deveau, M., Chen, C.-P., Johanson, G. et al. (2015). The global landscape of occupational exposure limits – Implementation of harmonization principles to guide limit selection. *J Occup Environ Hyg* 12, Suppl 1, S127-144. doi:10.1080/15459624.2015.1060327
- Dorman, D. C., Chiu, W., Hales, B. F. et al. (2018). Systematic reviews and meta-analyses of human and animal evidence of prenatal diethylhexyl phthalate exposure and changes in male anogenital distance. *J Toxicol Environ Health B Crit Rev* 21, 207-226. doi:10.1080/10937404.2018.1505354
- Eddy, D. M. (2005). Evidence-based medicine: A unified approach. *Health Aff (Millwood)* 24, 9-17. doi:10.1377/hlthaff.24.1.9
- EFSA (2010). Application of systematic review methodology to food and feed safety assessments to support decision making. *EFSA J* 8, 1637. doi:10.2903/j.efsa.2010.1637
- EFSA (2015). Principles and process for dealing with data and evidence in scientific assessments. *EFSA J* 13, 4121. doi:10.2903/j.efsa.2015.4121
- EFSA, Gundert-Remy, U., Bodin, J. et al. (2017). Bisphenol A (BPA) hazard assessment protocol. *EFSA Supporting Publication* 14, 1354E. doi:10.2903/sp.efsa.2017.EN-1354
- EFSA Scientific Committee (2018a). Guidance on uncertainty analysis in scientific assessments. *EFSA J* 16, e05123. doi:10.2903/j.efsa.2018.5123
- EFSA Scientific Committee (2018b). The principles and methods behind EFSA's guidance on uncertainty analysis in scientific assessment. *EFSA J* 16, e05122. doi:10.2903/j.efsa.2018.5122
- Gusenbauer, M. and Haddaway, N. R. (2021). What every researcher should know about searching – Clarified concepts, search advice, and an agenda to improve finding in academia. *Res Synth Methods* 12, 136-147. doi:10.1002/jrsm.1457
- Guyatt, G. H., Oxman, A. D., Vist, G. E. et al. (2008). GRADE: An emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 336, 924-926. doi:10.1136/bmj.39489.470347.AD
- Hassauer, C. and Roosen, J. (2020). Toward a conceptual framework for food safety criteria: Analyzing evidence practices using the case of plant protection products. *Saf Sci* 127, 104683. doi:10.1016/j.ssci.2020.104683
- Hoffmann, S., De Vries, R. B. M., Stephens, M. L. et al. (2017).

- A primer on systematic reviews in toxicology. *Arch Toxicol* 9, 2551-2575. doi:10.1007/s00204-017-1980-3
- Hoffmann, S., Aiassa, E., Angrish, M. et al. (2022). Application of evidence-based methods to construct mechanism-driven chemical assessment frameworks. *ALTEX*, Online ahead of print. doi:10.14573/altex.2202141
- Howard, B. E., Phillips, J., Tandon, A. et al. (2020). SWIFT-active screener: Accelerated document screening through active learning and integrated recall estimation. *Environ Int* 138, 105623. doi:10.1016/j.envint.2020.105623
- Kabat, G. C. (2019). Rapid response to: Probable carcinogenicity of glyphosate. *BMJ* 365, 11613. doi:10.1136/bmj.11613
- Kass, G. E. N. and Lodi, F. (2020). Letter to the editor regarding the article 'EFSA's toxicological assessment of aspartame: Was it even-handedly trying to identify possible unreliable positives and unreliable negatives?' *Arch Public Health* 78, 14. doi:10.1186/s13690-020-0395-4
- Kogevinas, M. (2019). Probable carcinogenicity of glyphosate. *BMJ* 365, 11613. doi:10.1136/bmj.11613
- Kousta, E., Lam, J., Sutton, P. 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* 122, 1015-1027. doi:10.1289/ehp.1307177
- Lam, J., Kousta, E., Sutton, P. 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* 122, 1040-1051. doi:10.1289/ehp.1307923
- Li, J., Pega, F., Ujita, Y. et al. (2020). The effect of exposure to long working hours on ischaemic heart disease: A systematic review and meta-analysis from the WHO/ILO joint estimates of the work-related burden of disease and injury. *Environ Int* 142, 105739. doi:10.1016/j.envint.2020.105739
- Lynch, H. N., Goodman, J. E., Tabony, J. A. et al. (2016). Systematic comparison of study quality criteria. *Regul Toxicol Pharmacol* 76, 187-198. doi:10.1016/j.yrtph.2015.12.017
- Mahood, Q., Van Eerd, D. and Irvin, E. (2014). Searching for grey literature for systematic reviews: Challenges and benefits. *Res Synth Methods* 5, 221-234. doi:10.1002/jrsm.1106
- Marshall, I. J. and Wallace, B. C. (2019). Toward systematic review automation: a practical guide to using machine learning tools in research synthesis. *Syst Rev* 8, 163. doi:10.1186/s13643-019-1074-9
- Martin, P., Bladier, C., Meek, B. et al. (2018). Weight of evidence for hazard identification: A critical review of the literature. *Environ Health Perspect* 126, 076001. doi:10.1289/EHP3067
- Matta, K., Ploteau, S., Coumoul, X. et al. (2019). Associations between exposure to organochlorine chemicals and endometriosis in experimental studies: A systematic review protocol. *Environ Int* 124, 400-407. doi:10.1016/j.envint.2018.12.063
- Meek, M. E., Boobis, A., Cote, I. et al. (2014). New developments in the evolution and application of the WHO/IPCS framework on mode of action/species concordance analysis. *J Appl Toxicol* 34, 1-18. doi:10.1002/jat.2949
- Millstone, E. P. and Dawson, E. (2019). EFSA's toxicological assessment of aspartame: Was it even-handedly trying to identify possible unreliable positives and unreliable negatives? *Arch Public Health* 77, 34. doi:10.1186/s13690-019-0355-z
- Moher, D., Shamseer, L., Clarke, M. et al. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev* 4, 1. doi:10.1186/2046-4053-4-1
- Morgan, R. L., Thayer, K. A., Bero, L. et al. (2016). GRADE: Assessing the quality of evidence in environmental and occupational health. *Environ Int* 92-93, 611-616. doi:10.1016/j.envint.2016.01.004
- Morgan, R. L., Thayer, K. A., Bero, L. et al. (2018). Identifying the PECO: A framework for formulating good questions to explore the association of environmental and other exposures with health outcomes. *Environ Int* 121, 1027-1031. doi:10.1016/j.envint.2018.07.015
- Morgan, R. L., Beverly, B., Ghersi, D. et al. (2019). GRADE guidelines for environmental and occupational health: A new series of articles in Environment International. *Environ Int* 128, 11-12. doi:10.1016/j.envint.2019.04.016
- NRC – National Research Council (2014). *Review of EPA's Integrated Risk Information System (IRIS) Process*. Washington, DC, USA: The National Academies Press.
- NTP – National Toxicology Program (2015). *Handbook for Preparing Report on Carcinogens Monographs*. https://ntp.niehs.nih.gov/ntp/roc/handbook/roc_handbook_508.pdf
- NTP (2016). *Monograph on Immunotoxicity Associated With Exposure to Perfluorooctanoic Acid (PFOA) and Perfluorooctane Sulfonate (PFOS)*. https://ntp.niehs.nih.gov/ntp/ohat/pfoa_pfos/pfoa_pfosmonograph_508.pdf
- OECD (2019). Guiding Principles and Key Elements for Establishing a Weight of Evidence for Chemical Assessment. *OECD Series on Testing and Assessment, No. 311*. OECD Publishing, Paris. <https://doi.org/10.1787/f11597f6-en>
- Pega, F., Momen, N.C., Ujita, Y. et al. (2021). Systematic reviews and meta-analyses for the WHO/ILO joint estimates of the work-related burden of disease and injury. *Environ Int* 155, 106605. doi:10.1016/j.envint.2021.106605
- RAC and SCOEL (2016). ECHA Committee for Risk Assessment (RAC) and Scientific Committee on Occupational Exposure Limits (SCOEL) Joint Opinion to resolve differences in scientific opinion as regards exposure levels for N-methyl-2-pyrrolidone. https://echa.europa.eu/documents/10162/13579/rac_joint_nmp_opinion_en.pdf/e4b4f43b-a3bd-a7c0-08be-16c3886593e7
- Rhomberg, L. R., Goodman, J. E., Bailey, L. A. et al. (2013). A survey of frameworks for best practices in weight-of-evidence analyses. *Crit Rev Toxicol* 43, 753-784. doi:10.3109/10408444.2013.832727
- Robinson, C., Portier, C. J., Cavoski, A. et al. (2020). Achieving a high level of protection from pesticides in Europe: Problems with the current risk assessment procedure and solutions. *Eur J Risk Regul* 11, 450-480. doi:10.1017/err.2020.18
- Rooney, A. A., Boyles, A. L., Wolfe, M. S. et al. (2014). Systematic review and evidence integration for literature-based environmental health science assessments. *Environ Health Perspect* 122, 711-718. doi:10.1289/ehp.1307972
- Rooney, A. A., Cooper, G. S., Jahnke, G. D. et al. (2016). How credible are the study results? Evaluating and applying internal



- validity tools to literature-based assessments of environmental health hazards. *Environ Int* 92-93, 617-629. doi:10.1016/j.envint.2016.01.005
- Rudén, C. (2001). The use and evaluation of primary data in 29 trichloroethylene carcinogen risk assessments. *Regul Toxicol Pharmacol* 34, 3-16. doi:10.1006/rtph.2001.1482
- Samuel, G. O., Hoffmann, S., Wright, R. A. et al. (2016). Guidance on assessing the methodological and reporting quality of toxicologically relevant studies: A scoping review. *Environ Int* 92-93, 630-646. doi:10.1016/j.envint.2016.03.010
- Schenk, L. (2010). Comparison of data used for setting occupational exposure limits. *Int J Occup Environ Health* 16, 249-262. doi:10.1179/107735210799160255
- Schenk, L. and Johanson, G. (2011). A quantitative comparison of the safety margins in the European indicative occupational exposure limits and the derived no-effect levels for workers under REACH. *Toxicol Sci* 121, 408-416. doi:10.1093/toxsci/kfr056
- Schenk, L., Deng, U. and Johanson, G. (2015). Derived no-effect levels (DNELs) under the European chemicals regulation REACH – An analysis of long-term inhalation worker-DNELs presented by industry. *Ann Occup Hyg* 59, 416-438. doi:10.1093/annhyg/meu103
- Schreider, J., Barrow, C., Birchfield, N. et al. (2010). Enhancing the credibility of decisions based on scientific conclusions: Transparency is imperative. *Toxicol Sci* 116, 5-7. doi:10.1093/toxsci/kfq102
- Schulz, K. F., Chalmers, I., Hayes, R. J. et al. (1995). Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 273, 408-412. doi:10.1001/jama.273.5.408
- Schünemann, H., Hill, S., Guyatt, G. et al. (2011). The GRADE approach and Bradford Hill's criteria for causation. *J Epidemiol Community Health* 65, 392-395. doi:10.1136/jech.2010.119933
- Schünemann, H. J., Brozek, J., Guyat, G. et al. (2013). GRADE handbook for grading quality of evidence and strength of recommendations. <https://gdt.gradepro.org/app/handbook/handbook.html>
- Stephens, M. L. et al. (2013). Evidence-based toxicology for the 21st century: Opportunities and challenges. *ALTEX* 30, 74-103. doi:10.14573/altex.2013.1.074
- Stephens, M. L., Andersen, M., Becker, R. A. et al. (2016). The emergence of systematic review in toxicology. *Toxicol Sci* 152, 10-16. doi:10.1093/toxsci/kfw059
- Thayer, K. A., Wolfé, M. S., Rooney, A. A. et al. (2014). Intersection of systematic review methodology with the NIH reproducibility initiative. *Environ Health Perspect* 122, A176-177. doi:10.1289/ehp.1408671
- van Luijk, J., Popa, M., Swinkels, J. et al. (2019). Establishing a health-based recommended occupational exposure limit for nitrous oxide using experimental animal data – A systematic review protocol. *Environ Res* 178, 108711. doi:10.1016/j.envres.2019.108711
- Whaley, P., Halsall, C., Agerstrand, M. et al. (2016). Implementing systematic review techniques in chemical risk assessment: Challenges, opportunities and recommendations. *Environ Int* 92-93, 556-564. doi:10.1016/j.envint.2015.11.002
- Whaley, P., Aiassa, E., Beausoleil, C. et al. (2020a). Recommendations for the conduct of systematic reviews in toxicology and environmental health research (COSTER). *Environ Int* 143, 105926. doi:10.1016/j.envint.2020.105926
- Whaley, P., Edwards, S. W., Kraft, A. et al. (2020b). Knowledge organization systems for systematic chemical assessments. *Environ Health Perspect* 128, 125001. doi:10.1289/EHP6994
- Whaley, P., Piggott, T., Morgan, R. L. et al. (2022). Biological plausibility in environmental health systematic reviews: A GRADE concept paper. *J Clin Epidemiol*, S0895-4356(22)00050-6. Online ahead of print. doi:10.1016/j.jclinepi.2022.02.011
- Wikoff, D., Urban, J. D., Harvey, S. et al. (2018). Role of risk of bias in systematic review for chemical risk assessment: A case study in understanding the relationship between congenital heart defects and exposures to trichloroethylene. *Int J Toxicol* 37, 125-143. doi:10.1177/1091581818754330
- Wikoff, D., Haws, L., Ring, C. et al. (2019). Application of qualitative and quantitative uncertainty assessment tools in developing ranges of plausible toxicity values for 2,3,7,8-tetrachlorodibenzo-p-dioxin. *J Appl Toxicol* 39, 1293-1310. doi:10.1002/jat.3814
- Wikoff, D., Urban, J. D., Ring, C. et al. (2020a). Development of a range of plausible noncancer toxicity values for 2,3,7,8-tetrachlorodibenzo-p-dioxin based on effects on sperm count: Application of systematic review methods and quantitative integration of dose response using meta-regression. *Toxicol Sci* 179, 162-182. doi:10.1093/toxsci/kfaa171
- Wikoff, D., Lewis, R. J., Erraguntla, H. et al. (2020b). Facilitation of risk assessment with evidence-based methods – A framework for use of systematic mapping and systematic reviews in determining hazard, developing toxicity values, and characterizing uncertainty. *Regul Toxicol Pharmacol* 118, 104790. doi:10.1016/j.yrtph.2020.104790
- Wolffe, T. A. M., Whaely, P., Halsall, C. et al. (2019). Systematic evidence maps as a novel tool to support evidence-based decision-making in chemicals policy and risk management. *Environ Int* 130, 104871. doi:10.1016/j.envint.2019.05.065
- 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. doi:10.1289/ehp.1307175
- Yost, E. E., Euling, S. Y., Weaver, J. A. et al. (2019). Hazards of diisobutyl phthalate (DIBP) exposure: A systematic review of animal toxicology studies. *Environ Int* 125, 579-594. doi:10.1016/j.envint.2018.09.038

Conflict of interest

For conduct of the immediate work, Sebastian Hoffmann and Paul Whaley declare personal fees from the Evidence-based Toxicology Collaboration. In addition, Paul Whaley declares in-kind support from Lancaster University. Outside the immediate work, Paul Whaley declares personal fees from Elsevier Ltd as Systematic Reviews Editor for the journal *Environment International*, the Cancer Prevention and Education Society, Yordas Group, the University of Central Lancashire, and grants from Lancaster University that relate to the development and promotion of systematic review methods in environmental health research, delivering training around these methods, and providing editorial services.