Implementing systematic review techniques in chemical risk assessment: challenges, opportunities and recommendations

Authors: Paul Whaley (1), Crispin Halsall (1)*, Marlene Ågerstrand (2), Diane Benford (3), Elisa Aiassa (4), Gary Bilotta (5), David Coggon (7), **Chris Collins (8)**, Ciara Dempsey (19), Raquel Duarte-Davidson (10), Rex FitzGerald (11), **Malyka Galay Burgos (12), David Gee (13)**, **Andy Hart (14)**, Sebastian Hoffmann (15), **Juleen Lam (16)**, Toby Lassersson (17)**, Leonard Levy (18)**, Steven Lipworth (19), Sarah Mackenzie Ross (20), Olwenn Martin (21), Catherine Meads (22), Monika Meyer-Baron (23), James Miller (24), **Patrice Mongelard (25)**, **Camilla Pease (26)**, Andrew Rooney (27), **Alison Sapiets (28)**, Gavin Stewart (29), David Taylor (19), **Didier Verloo (31)**.

(1) Lancaster Environment Centre, Lancaster University, Lancaster LA1 4YQ, UK

(2) Department of Environmental Science and Analytical Chemistry, Stockholm University, SE-106 91 Stockholm, Sweden

(3) Food Standards Agency, Aviation House, 125 Kingsway, London, WC2B 6NH, UK

(5) Aquatic Research Centre, University of Brighton, Lewes Road, Brighton BN2 4GJ, UK

(11) Swiss Centre for Applied Human Toxicology, University of Basel, Missionsstrasse 64, 4055 Basel, Switzerland

(15) Evidence-Based Toxicology Collaboration (EBTC), Stembergring 15, 33106 Paderborn, Germany

(17) Cochrane Editorial Unit, Cochrane Central Executive, St Albans House, 57-9 Haymarket, London, SW1Y 4QX , UK

(19) Royal Society of Chemistry, Burlington House, Piccadilly, London, W1J 0BA, UK

(20) Research Department of Clinical, Educational and Health Psychology, University College London, Gower Street, London WC1E 6BT, UK

(21) Institute for the Environment, Health and Societies, Brunel University London, Kingston Lane, Uxbridge UB8 3PH, UK

(22) Health Economics Research Group, Brunel University London, Kingston Lane, Uxbridge, UB8 3PH, UK

(23) Leibniz Research Centre for Working Environment and Human Factors (IfADo). Neurobehavioural Toxicology, Ardeystr. 67, D - 44139 Dortmund, Germany

(24) Centre for Ecology and Hydrology, Wallingford, Oxfordshire, 0X10 8BB, UK

(27) National Institute of Environmental Sciences (NIEHS), National Institutes of Health (NIH), Department of Health and Human Services (DHHS), Research Triangle Park, North Carolina, USA

(29) Centre for Rural Economy, School of Agriculture, Food and Rural Development, University of Newcastle upon Tyne, UK.

*corresponding author: c.halsall@lancaster.ac.uk

Disclaimer: The views expressed in this manuscript are those of the authors and do not necessarily represent the views or policies of their employers or otherwise affiliated organisations.

Abstract

Systematic review (SR) is a rigorous, protocol-driven approach designed to minimise error and bias when summarising the body of research evidence relevant to a specific scientific question. Taking as a comparator the use of SR in synthesising research in healthcare, we argue that SR methods could also pave the way for a "step change" in the transparency, objectivity and communication of chemical risk assessments (CRA) in Europe and elsewhere. We suggest that current controversies around the safety of certain chemicals are partly due to limitations in current CRA procedures which have contributed to ambiguity about the health risks posed by these substances. We present an overview of how SR methods can be applied to the assessment of risks from chemicals, and indicate how challenges in adapting SR methods from healthcare research to the CRA context might be overcome. Regarding the latter, we report the outcomes from a workshop exploring how to increase uptake of SR methods, attended by experts representing a wide range of fields related to chemical toxicology, risk management and SR. Priorities which were identified include: the conduct of CRA-focused prototype SRs; the development of a recognised standard of reporting and conduct for SRs in toxicology and CRA; and establishing a network to facilitate research, communication and training in SR methods. We see the workshop and this paper as a logical step in the creation of a research climate that fosters communication between experts in CRA and SR and facilitates wider uptake of SR methods into CRA.

Keywords: evidence; human health; toxicology; risk; environment

1. Introduction

Systematic review (SR) is a rigorous, protocol-driven approach to minimising error and bias[1] in the aggregation and appraisal of evidence relevant to answering a research question. SR techniques were initially developed in the fields of psychology, social science and health care and have, since the 1980s, provided a valuable tool for evidence-informed decision-making across many domains (Lau et al. 2013). In medicine, SRs have provided a valuable response to the need for consistent, transparent and scientifically-robust interpretations of the results of increasing numbers of often conflicting studies of the efficacy of healthcare interventions. SRs have taken on an increasingly fundamental role both in supporting decision-making in healthcare and, by channelling resources towards questions for which the answers are not yet known, reducing waste in research (Chalmers, Glasziou 2009; Salman et al. 2014). It is now accepted practice in healthcare to use SR methods to assess evidence not only for the efficacy of interventions, but diagnostic tests, prognostics and adverse outcomes.

The extension of SR techniques to other fields is based on a mutual need across disciplines to make the best use of existing evidence when making decisions, a move for which momentum has been growing for several decades. For example, the What Works Clearinghouse was established in 2002 to apply SR techniques in support of American educational policy (US Institute of Education Sciences 2015), and in 2000 the international Campbell Collaboration research network was convened to undertake and disseminate systematic reviews on the effects of social interventions in diverse fields such as crime and justice, education, international development and social welfare (Campbell Collaboration 2015). Meta-analysis and SR in ecology have contributed to evidence-based environmental policy since the mid-1990s (Stewart 2010); more recently, the Collaboration for Environmental Evidence (CEE) has been established to encourage conduct of SRs on a wide range of environmental topics (Collaboration for Environmental Evidence 2015).

The potential advantages of adapting SR methodology to the field of chemical risk assessment (CRA) have also been recognised and multiple research groups and organisations have adopted (Woodruff, Sutton 2014; Birnbaum et al. 2013; European Food Safety Authority 2010; Rooney et al. 2014) or recommended (US National Research Council 2014a, 2014b; US

[1] It is worth drawing a distinction between three sources of bias in the review process. There is potential for bias in the conduct of a review (e.g. because of inappropriate methods for identifying and selecting evidence for inclusion in the review); bias because the material available for the review is not representative of the evidence base as a whole (due to selective publication); and bias arising from flaws in the design, conduct, analysis and reporting of individual studies included in the review that can cause the effect of an intervention or exposure to be systematically under- or over-estimated. One of the major functions of SRs is to minimise bias in the conduct of a review and prevent, as far as possible, bias from selective publication and methodological flaws in the evidence from giving a misleading impression of what is known in relation to a particular research question.

Environmental Protection Agency 2013; Silbergeld, Scherer 2013; Hoffmann, Hartung 2006; Zoeller et al. 2015) the use of SR methods for evaluating the association between health effects and chemical exposures to inform decision-making. There are, however, a number of recognised challenges in extending SR methods to CRA, many of which derive from key differences in the evidence base between the healthcare and toxicological sciences.

SRs in medicine often focus on direct evidence for benefits and adverse effects of healthcare interventions derived from randomised controlled trials (RCTs) in humans. The evidence base for CRA is generally more complex, with a need to extrapolate from investigations in animals, *in vitro* and *in silico*, and then to synthesise findings with those from human studies if available. Furthermore, the human data tend to come from observational studies with greater and more varied potential for bias and confounding than RCTs. Also, the range of outcomes to be considered is usually much wider than in the assessment of healthcare interventions. Thus, when the various types of toxicological research are combined into a single overall conclusion about the health risks posed by a chemical exposure, reviewers are challenged with integrating the results from a broad and heterogeneous evidence base.

In spite of these differences, there is reason for thinking that SR methods can be applied successfully to CRA. For example, techniques for aggregating the results of different study types are already addressed in various frameworks already in use in toxicology. These include: International Agency of Research on Cancer (IARC) monographs (International Agency for Research on Cancer 2006); the Navigation Guide (Woodruff, Sutton 2014); and the US Office for Health Assessment and Translation (OHAT) (Rooney et al. 2014; US National Toxicology Panel 2015). Heterogeneous sources of evidence are a familiar challenge in all domains including clinical medicine (Lau et al. 1998), and SR of observational studies has a crucial role in identifying complications and side-effects of healthcare interventions (Sterne et al. 2014; Higgins, Green 2011). The need for SR of pre-clinical animal trials of healthcare interventions, in order to better anticipate benefits and harms to humans, is another area in which methods being developed and utilised by a number of groups including SYRCLE (Hooijmans et al. 2012; van Luijk et al. 2014) and CAMARADES (Macleod et al. 2005; Sena et al. 2014). Stewart and Schmid (2015) argue that research synthesis methods (including systematic review) are generic and applicable to any domain if appropriately contextualised.

Given the sometimes controversial outcome of CRAs and the increasingly high public and media profile of the risks that chemicals may pose to humans and the environment, SR is increasingly viewed as a potentially powerful technique in assessing and communicating how likely it is that a chemical will cause harm. SR methods add transparency, rigour and objectivity

to the process of collecting the most relevant scientific evidence with which to inform policy discussions and could provide a critical tool for organising and appraising the evidence on which chemical policy decisions are based.

Consequently, in November 2014 a group of 35 scientists and researchers from the fields of medicine, toxicology, epidemiology, environmental chemistry, ecology, risk assessment, risk management and SR participated in a one-day workshop to consider the application of SR in CRA. The purpose was three-fold:

1. Identify from expert practitioners in risk assessment and SR the obstacles, in terms of practical challenges and knowledge gaps, to implementing SR methods in CRA.

2. Develop a "roadmap" for overcoming those obstacles and expediting the implementation of SR methods, where appropriate, by the various stakeholders involved in CRA.

3. Establish the foundations of a network to co-ordinate research and activities relating to the implementation of SR methods. The aim would be to promote the wider adoption of SR in CRA, both in Europe and elsewhere, and to support best practice in the application of SR techniques.

Participants heard seven presentations about recent developments in SR methods, their application to the risk assessment process, and their potential value to policy-makers. There were two break-out sessions in which participants were divided into three facilitated groups, firstly to discuss challenges to implementing SR methods in CRA, and then to suggest ways in which the obstacles could be overcome. These ideas were discussed in plenary before being summarised, circulated for comment, and then published in this paper. The workshop was conducted under the "Chatham House Rule" such that participants were free to refer to the information presented and discussed, provided they did not attribute it to identifiable individuals or organisations.

The purpose of this overview paper is to present the rationale for exploring the application of SR methods to CRA, the various experts' views on the challenges to implementing SR methods in CRA, and their suggestions for overcoming them. The remaining goals of the meeting are ongoing work, including the development of the roadmap concept for publication and the establishment of a network for supporting the use of SR in CRA.

2. The appeal of SR methods in CRA

Chemical risk assessment is a multi-step process leading to a quantitative characterisation of risk, which can then be used to inform the management of chemical substances so as to ensure that any risks to human health or the environment are managed optimally. CRAs entail four fundamental steps: hazard identification; hazard characterisation (often a dose-response assessment); exposure assessment; and risk characterisation (see Figure 1). These steps draw on various fields of scientific research including chemistry, exposure sciences, toxicology (encompassing *in vivo*, *in vitro*, ecotoxicological and *in silico* methods), ecotoxicology, human epidemiology, and mathematical modelling.

There are many ways in which errors can occur in the interpretation of evidence from these varied disciplines, including failure to consider all relevant data, failure to allow appropriately for the strengths and limitations of individual studies, and over- or underestimating the relevance of experimental models to real-world scenarios (to name a few). Whether the appraisal of evidence is based on objective processes or on subjective expert judgement and opinion may also be an important factor in accurate interpretation of evidence. The assessment process always requires input from technical experts, which inevitably brings an element of subjectivity to the interpretation of the scientific evidence. Different experts may have varying degrees of practical and cognitive access to relevant information, place differing weight on individual studies and/or strands of evidence that they review, and when working in committee, may be more or less influenced by dominant personalities. This can result in misleading conclusions in which the potential for health risks is overlooked, underestimated or overstated. Furthermore, if the factors determining their assessment of evidence are undocumented, when expert opinions are in conflict it can be very challenging to distinguish which opinion is likely to represent the most valid synthesis of the totality of available evidence.

A recent illustrative example (see Figure 2) of when expert scientists and reputable organisations have come to apparently contradictory conclusions about the likelihood of a chemical causing harm is the case of bisphenol-A (BPA). BPA is a monomer used in the manufacture of the resinous linings of tin cans and other food contact materials such as polycarbonate drinks bottles. It has been banned from use in infant-feed bottles across the EU (European Commission 1/28/2011) because of "uncertainties concerning the effect of the exposure of infants to Bisphenol A" (European Commission 5/31/2011).

The European Food Safety Authority (EFSA) considers current exposure levels to BPA to present a low risk of harm to the public (European Food Safety Authority 2015a). The French food regulator ANSES takes a seemingly different stance on the risks to health posed by BPA

(French Agency for Food, Environmental and Occupational Health & Safety 4/7/2014), determining there to be a "potential risk to the unborn children of exposed pregnant women". ANSES has on this basis proposed classifying BPA as toxic to reproduction in humans (French Agency for Food, Environmental and Occupational Health & Safety 2013), a proposal which has contributed to the French authorities' decision to implement an outright ban on BPA in all food packaging materials (France 12/24/2012). While the ban has been challenged by some stakeholders as being disproportionate under EU law (Tošenovský 2014, 2015; Plastics Europe 1/15/2015), the Danish National Food Institute has argued that EFSA has overestimated the safe daily exposure to BPA and that some populations are exposed to BPA at levels higher than can be considered safe (National Food Institute, Denmark 2015); a view reflected in the conclusions of some researchers, e.g. Vandenberg et al. (2014) but not others, e.g. US Food and Drug Administration (2014).

The example of BPA shows that while all these bodies ostensibly have access to the same evidence base regarding the potential toxicity of BPA, there is a lack of consensus on how best to interpret it, either in terms of what is known and what is uncertain about the risks to health posed by BPA, and/or what response is appropriate to managing those risks and uncertainties. It also shows how, in the absence of that consensus, there is a danger that policy on BPA may become disconnected from the evidence base, either risking harm to health through continued exposure to BPA or incurring unnecessary economic costs through restricting the use of a chemical which is in fact sufficiently safe. It also suggests that if the reasons for disagreement about health risks posed by a chemical are not accessible to various stakeholders in the debate, then it becomes much more difficult for regulators to credibly resolve controversies about chemical safety, potentially undermining their authority in the long term.

This example highlights the potential for differences in the interpretation of evidence when assessing chemical toxicity and the need for a process that is not only scientifically robust but also transparent, so the reasons for any disagreement can be easily recognised – including giving stakeholders greater opportunity to understand when differences in policy stem from divergent assessments of risk, and when they stem from divergent opinions as to how those risks are best managed. It also suggests the importance of the following characteristics in risk assessments that are used to inform risk management decisions:

1. *Transparency*, in that the basis for the conclusions of the risk assessment should be clear (otherwise they may not be trusted and errors may go undetected).

2. *Validity*, in that CRAs should be sufficiently (though not necessarily maximally) scientifically robust in their methodology and accurate in their estimation of risks and

characterisation of attendant uncertainties as to optimise the decisions that must be made in risk management.

3. *Confidence,* providing the user with a clear statement as to the overall strength of evidence for the conclusions reached and a characterisation of the utility of the evidence for decision-making (e.g. "appropriate for hazard identification but inappropriate for identification of a reference dose").

4. *Utility*, in that the output of the risk assessment should be in a form that is convenient and intelligible to those who will use it (outputs that are too detailed and complex may be inaccessible, leading to inefficiency and possibly erroneous decisions).

5. *Efficiency*, providing a clear justification of the choice of research question in the context of efficiently solving a CRA problem. Resources for CRA are limited and it is wasteful to expend unnecessary effort on aspects of an assessment that will not be critical to decision-making (although for the purposes of transparency and validity, the reasons for focusing on a particular outcome or otherwise restricting the evaluation should be explained).

6. *Reproducibility*, in that the conclusions of the SR process when applied to the same question and data should ideally produce the same answer even when undertaken by different individuals (also described as "consistency"). In practice, different experts may reach difference conclusions because they will not all make the same value judgments about scope, quality and interpretation of evidence. Therefore, the process should be sufficiently rigorous such that it is highly likely that scientific judgment would result in the same conclusion independent of the experts involved, and as a minimum the SR process should render transparent any reasons for all conclusions.

It may be perceived that the value of SR methods lies in their provision of unequivocal assessments of whether or not a chemical will induce specific harm to humans and/or wildlife in given circumstances. In practice, however, this will happen only if the evidence base is sufficiently extensive and there is also unanimity in identification of the problem, the quality of the evidence base, and how it is to be interpreted in answering the review question (without this, SRs will also produce different results). Often, the consensus and/or information may be relatively limited; in such circumstances, a SR will instead clearly state the limitations of the available data and consequent uncertainties. The value here is in the provision of a comprehensive and transparent assessment of what is *not* known and insight into the drivers of divergent opinion. From a research perspective, this yields valuable information about how

research limitations and knowledge gaps contribute to ongoing uncertainty about environmental and health risks, allowing the subsequent efforts of researchers to be more clearly focused. From a policy perspective, SRs offer a transparent explanation as to why there are differences in opinion which can then be communicated to stakeholders.

Overall, SR contributes to achieving consensus not by eliminating expert judgement, nor by eliminating conflicting opinions about whether a compound should be banned, but by providing a robust, systematic and transparent framework for reviewing evidence of health risks, such that when there is disagreement, the reasons for it are clearly visible and the relative merits of differing opinions can be appraised. In this way, it may help to resolve controversies in the interpretation of the science which informs the risk management process.

3. SR and its application to CRA

3.1. Traditional vs. SR methods

SR methods are often contrasted with "traditional", non-systematic narrative approaches to describing what is and is not already known in relation to a research question. In reality, the distinction between systematic and narrative review is a crude one, with narrative reviews encompassing a number of different approaches to reviewing evidence, from the caricature of one researcher writing about "my field, from my standpoint […] using only my data and my ideas, and citing only my publications" (Caveman 2000), to thorough narrative critiques of comprehensively identified evidence relevant to answering an explicitly articulated question, as conducted by organisations such as IARC.

Nonetheless, it is worth noting that only relatively recently has it been recognised that traditional narrative reviews are, to varying degrees, vulnerable to a range of methodological shortcomings which are likely to bias their summarisation of the evidence base (Chalmers et al. 2002). These include selective rather than comprehensive retrieval of evidence relevant to the review topic, inconsistent treatment of differences in the methodological quality of included studies, and even an absence of clear review objectives or conclusions which are drawn directly from the strengths and limitations of the evidence base (Mulrow 1987).

The presence of these shortcomings seriously challenges the reader's ability to determine the credibility of a review. When there exist multiple competing reviews, each using opaque methods, it becomes almost impossible to judge their relative merits and therefore to base decisions on current best available evidence. The consequence is a proliferation of conflicting

opinions about best practice that fail to take proper account of the body of research evidence. In the healthcare sciences, this was initially shown by Antman and colleagues when they found that, in comparison to recommendations of clinical experts, systematic aggregation of data from existing clinical trials of streptokinase to treat myocardial infarction would have demonstrated benefit some years before recommendations for its use became commonplace (Antman et al. 1992). More recently, cumulative meta-analyses have been shown to be more accurate in estimating current understanding of the size of effect of a wide range of healthcare interventions than researchers planning new clinical trials who have not used these methods (Clarke et al. 2014).

A SR is an approach to reviewing evidence which specifically sets out to avoid these problems, by methodically attempting "to collate all empirical evidence that fits pre-specified eligibility criteria in order to answer a specific research question," using "explicit, systematic methods that are selected with a view to minimizing bias" (Higgins & Green, 2011).

In detail, this amounts to the pre-specification, in a written protocol, of the objective and methods of the SR, in which the aim of conducting the review is clearly stated as a structured question (for a SR of the effects of an intervention or exposure, this can establish a testable hypothesis or quantitative parameter that is to be estimated), along with the articulation of appropriate methods. The methods specified should include the methods for identifying literature of potential relevance to the research question, the criteria for inclusion of the studies of actual relevance to the research question, how the internal validity[2] of the included studies will be appraised, and the analytical techniques used for combining the results of the included studies. The purposes of the protocol are to discourage potential biases from being introduced via adjustments being made to the review methodology in the course of its conduct, to allow any justifiable changes to be tracked, and also to allow the protocol to be peer-reviewed in order to help ensure utility and validity of the proposed objective and approach.

The final SR itself consists of a statement of the objective, the search method, the criteria for including relevant studies for analysis, and the results of the appraisal of internal validity of the included studies, e.g. implemented as a "risk of bias" assessment in Cochrane Reviews of randomised trials (Higgins et al. 2011). The evidence is then synthesised using statistical meta-

[2] "Internal validity" is a term used in Cochrane Collaboration guidance on conduct of SRs specifically intended to supersede the use of terms such as "methodological quality" or their equivalents, which are considered ambiguous (Higgins, Green 2011). The internal validity of a piece of research is appraised in a "risk of bias" assessment. The target of the risk of bias assessment is the likelihood, magnitude and direction of systematic error in estimated size of effect of an intervention, as caused by flaws in the design, conduct, analysis and reporting of a study. Throughout this document, we follow Cochrane Collaboration conventions in using "internal validity" as a technical term in place of "methodological quality".

analytical techniques, narrative methods or both (depending on the extent to which meta-analysis is possible) into an overall answer to the research question. An assessment is then made of the strength of the evidence supporting the answer; in Cochrane reviews, this typically follows the GRADE methodology (Atkins et al. 2004), taking into account overall features of the evidence base including risk of bias across the included studies, publication bias in the evidence base, external validity or directness of the evidence to the population of interest, heterogeneity of the evidence, and the overall precision of the evidence. This is finally followed by a concluding interpretation of what the SR as a whole determines is and is not known in relation to its objective.

In this, we draw a distinction between a SR and a meta-analysis, whereby a meta-analysis pools the results of a number of studies and may be a component of a SR but does not incorporate the full set of methodological features which define the SR process (e.g. a meta-analysis may or may not include an assessment of the internal validity of included studies). While we acknowledge that some researchers use the terms "systematic review" and "meta-analysis" interchangeably, we believe the two approaches should be disambiguated. It is also worth noting that many reviews employ a combination of narrative and systematic methods; there were differing opinions among workshop participants as to the extent to which it is reasonable to expect all reviews to fully incorporate SR methods.

3.2. The current status of SR in environmental health, toxicology and CRA

While the use of SR methodologies is well established in healthcare to determine the effect of interventions on health outcomes or the accuracy of a diagnostic test, application of SR is relatively new in the fields of toxicology and environmental health. Workshop participants heard how methods for systematic review of medical interventions have in the United States been adapted in both academic and federal contexts to the gathering and appraising of evidence for the effects of chemical exposures on human health: researchers at the University of California have developed the *Navigation Guide* (Woodruff, Sutton 2014), and the US Office of Health Assessment and Translation (OHAT) at the US National Toxicology Program has developed the OHAT Framework for systematically reviewing environmental health research for hazard identification (Rooney et al. 2014).

The two approaches adapt the key elements of SR methods to questions in environmental health (which is directly relevant to the CRA process but does not include assessment of dose-response). Features the two approaches have in common include: conducting a SR according to a pre-specified protocol; the development of a specific research question and use of "PECO" statements in systematising review objectives and the methods that will be used to answer that

question; an approach to appraising the internal validity of included studies adapted from the risk of bias appraisal tool developed by the Cochrane Collaboration (Higgins et al. 2011); an adaptation of the GRADE methodology (Atkins et al. 2004) for describing the certainty or strength of a body of evidence, incorporating risk of bias elements with other criteria such as for the assessment of relevance or external validity; and a methodology for combining the results of human and animal research into a statement of confidence about the hazard which a chemical poses to health.

"PECO" is an acronym representing: Population (the exposure group of interest, e.g. people of a certain age or rats in laboratory studies); Exposure (the compounds or exposure scenarios of interest, e.g. respiratory exposure to fine particulate matter); Comparator (the group to which the exposure group is being compared, e.g. vehicle-exposed controls in laboratory experiments or less exposed groups in epidemiological studies); Outcome (a deleterious change or marker thereof hypothesised to be brought about by the exposure). The purpose of a PECO statement is to provide a framework for developing the key question which a SR will answer, and also to determine the rationale for the inclusion and exclusion criteria that explicitly define which studies are relevant for the review.

Other tools are being developed to contribute to the systematic assessment of *in vivo* and ecotoxicity studies which have not been directly derived from Cochrane Collaboration methods. Presented at the workshop was SciRAP (Science in Risk Assessment and Policy), a system developed to improve the consistency with which the relevance and reliability of studies are appraised in the context of conducting a chemical risk assessment for regulatory purposes. It is also intended to reduce the risk of selection bias in the risk assessment process by providing a mechanism for including non-standardised study methods yielding potentially valuable data (Beronius et al. 2014; SciRAP 2014).

There are a number of other initiatives promoting and developing the use of SR methodologies in environmental and chemical risk assessment. Participants heard about how the European Food Safety Authority is integrating SR methods into its assessments of food and feed safety (see e.g. European Food Safety Authority 2015b, 2015c), and about the UK Joint Water Evidence Group methods for rapid and systematic assessments of evidence (see e.g. Collins et al. 2014). Other coordinated initiatives include the Evidence-Based Toxicology Collaboration (Hoffmann & Hartung 2006); the Collaboration for Environmental Evidence (Bilotta et al. 2014a; Land et al. 2015); and the Systematic Review Centre for Laboratory Animal Experimentation (SYRCLE).

3.3. Overcoming the challenges in implementing SR methods in CRA

Risk assessment for a chemical or group of chemicals is a multi-faceted process that normally requires consideration of multiple endpoints in relation to a variety of exposure scenarios, integrating evidence from epidemiological studies, bioassays in animals, mechanistic studies and studies on the distribution and determinants of exposure by different pathways and routes. In addition to resolving methodological issues relating to underdeveloped methods (e.g. how SR methods can be used as part of dose-response assessment or how they can be applied to exposure assessment), it is important to consider how SR should fit into the CRA process. The principal challenge going forward is to explore the circumstances in which it would be worth applying more rigorous methods to assess scientific evidence than have been used to date, and to evaluate the practicality and cost-effectiveness of applying such methods in those situations.

In principle, SRs could be conducted on any aspect of a CRA. Given the success in employing SR methods to support evidence-based practice in healthcare, it is intuitive that SRs be applied to address specific questions arising within toxicology, human epidemiology and environmental health (e.g. hazard assessment within a CRA) and this view appears to be gaining momentum within the environmental health literature. The SR method may also lend itself to answering questions concerning e.g. the accuracy of the reported physical-chemical properties of a substance, doses predicted by quantitative exposure assessment, concentrations of a chemical in the environment and biota, and the derivation of a No Observed Adverse Effect Level (NOAEL) or Benchmark Dose Lower 95% confidence limit (BMDL).

Depending on scope, the resources (time and cost) to undertake an SR can be considerable. Currently there is a lack of empirical evidence relating to the resource-effectiveness of SR approaches in CRA and there was a difference of opinion among workshop participants as to whether the effort required for conducting a SR tends to be under- or overestimated. It was suggested that, where effort is likely to be substantial, efficient use of resources may be achieved by focusing on high-value questions developed through initial scoping exercises. For example, a low-dose adverse effect may be evident in animal models and supported to some extent by human epidemiology and hence a question may be formulated around this initial evidence; there may be little point, however, in pursuing a question related to non-carcinogenic toxicity in wildlife if a substantial part of the literature points towards that substance being a potential human carcinogen. There is also growing interest in rapid reviews, when full SR methods are considered overly onerous (Collins et al. 2014; Schünemann, Moja 2015).

The priorities for expediting the adaptation of SR methods to CRA identified at the workshop are as follows:

1. The development of a number of prototype CRA-focused SRs to explore how readily SR procedures can be integrated into the CRA process, to:

    a. identify additional methodological challenges in adapting SR methods to the CRA context and develop techniques to address them;

    b. acquire practical experience in managing resources when conducting SRs in CRA, including the conduct of scoping exercises for identifying high-value review questions, and the further development and/or application of novel "rapid evidence review" methods (UK Civil Service 2015).

2. Technical development of SR methodologies for CRA purposes, in particular the further advancement of techniques for incorporating both animal toxicology and human epidemiological studies into the SR and CRA process, to include:

    a. refining tools for more consistent and scientifically robust appraisal of the internal validity of individual studies included in a CRA and the implications for interpretation of their findings; see e.g. Bilotta et al. (2014b). This might include further development and validation of tools such as the SYRCLE methodology for assessing the internal validity of animal studies (Hooijmans et al. 2014), for SR of observational studies, see e.g. Sterne et al. (2014), the methods employed in the NTP/OHAT and Navigation Guide protocols, and the applicability of other assessment methods such as SciRAP (Beronius et al. 2014);

    b. the further development of software akin to the Cochrane Collaboration's Review Manager (Nordic Cochrane Centre 2014) and the Systematic Review Data Repository (Ip et al. 2012), and tools such as DRAGON (ICF International 2015) and the Health Assessment Workspace Collaborative (Rusyn, Shapiro 2013) to support extraction, analysis and sharing of data from studies included in reviews.

3. The development an empirical evidence base for the different types of bias that operate in the CRA domain, including their direction and potential magnitude, and the extent to which any methods being adopted to address them are appropriate and effective.

4. The development of a recognised "gold standard" for SRs in toxicology and risk assessment equivalent to the Cochrane Collaboration in evidence-based medicine, to address the growing number of purported SRs of unclear methodological robustness which are increasingly prevalent in the environmental health literature.

5. The creation of a climate of constructive discussion that fosters advancement of methods whereby chemical risk practitioners, industry, competent authorities, academic researchers and policy makers can research, discuss and evaluate SR methods and the potential advantages they can bring.

6. The establishment of a network of scientists and CRA practitioners to pursue research into and discussion of SR methodologies and facilitate their implementation.

7. The implementation of training programmes for risk assessment practitioners and stakeholders, focusing specifically on application of SR methods to CRA rather than current courses which focus largely on SR methods in healthcare.

4. Conclusions

While SR methods have proved highly influential in healthcare, they have yet to make significant impact on the CRA process. There are a number of challenges to implementing SR methods in CRA, with particular concerns about approaches to assessing bias and confounding in observational studies, the effort involved in conducting SRs and the subsequent benefits of conforming to SR standards. There is also much promise in the concept of adapting SR methods to CRA via its potential either to give definitive answers to specified research questions, or to enable identification of the reasons for failure to resolve debate. Recent experience from both regulatory agencies and academics already yields some clear recommendations which would expedite the implementation of SR methods in CRA, with the potential to increase the efficiency, transparency and scientific robustness of the CRA process.

5. Acknowledgements

Funding for the workshop was provided through the Economic & Social Science Research Council grant “Radical Futures in Social Sciences” (Lancaster University) and Lancaster Environment Centre. CH, PW, AR are grateful to Lancaster University’s Faculty of Science & Technology “Distinguished Visitors” funding programme. The Royal Society of Chemistry is acknowledged for generously providing a meeting room, refreshments and facilitating the workshop proceedings. The PhD studentship of PW is partly funded through Lancaster Environment Centre. The contribution of non-author workshop participants to the development of the manuscript is also greatly appreciated.

6. References

Antman, E.M.; Lau, J.; Kupelnick, B.; Mosteller, F.; Chalmers, T.C. (1992): A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts. Treatments for myocardial infarction. *JAMA* 268 (2), pp. 240–248.

Atkins, D.; Best, D.; Briss, P.A.; Eccles, M.; Falck-Ytter, Y.; Flottorp, S. et al. (2004): Grading quality of evidence and strength of recommendations. *BMJ (Clinical research ed.)* 328 (7454), p. 1490. DOI: 10.1136/bmj.328.7454.1490.

Beronius, A.; Molander, L.; Rudén, C.; Hanberg, A. (2014): Facilitating the use of non-standard in vivo studies in health risk assessment of chemicals: a proposal to improve evaluation criteria and reporting. *Journal of applied toxicology : JAT* 34 (6), pp. 607–617. DOI: 10.1002/jat.2991.

Bilotta, G.S.; Milner, A.M.; Boyd, I. (2014a): On the use of systematic reviews to inform environmental policies. *Environmental Science & Policy* 42, pp. 67–77. DOI: 10.1016/j.envsci.2014.05.010.

Bilotta, G.S.; Milner, A.M.; Boyd, I.L. (2014b): Quality assessment tools for evidence from environmental science. *Environ Evid* 3 (1), p. 14. DOI: 10.1186/2047-2382-3-14.

Birnbaum, L.S.; Thayer, K.A.; Bucher, J.R.; Wolfe, M.S. (2013): Implementing systematic review at the National Toxicology Program: status and next steps. *Environ. Health Perspect.* 121 (4), pp. A108-9. DOI: 10.1289/ehp.1306711.

French Agency for Food, Environmental and Occupational Health & Safety (4/7/2014): Bisphenol A: ANSES publishes its comments in response to the EFSA draft Opinion for consultation. Available online at https://www.anses.fr/en/content/bisphenol-anses-publishes-its-comments-response-efsa-draft-opinion-consultation, accessed 2/18/2015.

Campbell Collaboration (2015): The Campbell Collaboration. Available online at http://www.campbellcollaboration.org/, accessed 6/13/2015.

Caveman (2000): The invited review - or, my field, from my standpoint, written by me using only my data and my ideas, and citing only my publications. *J. Cell. Sci.* 113 (Pt 18), pp. 3125–3126.

Chalmers, I.; Glasziou, P. (2009): Avoidable waste in the production and reporting of research evidence. *The Lancet* 374 (9683), pp. 86–89. DOI: 10.1016/S0140-6736(09)60329-9.

Chalmers, I.; Hedges, L.V.; Cooper, H. (2002): A Brief History of Research Synthesis. *Evaluation & the Health Professions* 25 (1), pp. 12–37. DOI: 10.1177/0163278702025001003.

Clarke, M.; Brice, A.; Chalmers, I. (2014): Accumulating research: a systematic account of how cumulative meta-analyses would have provided knowledge, improved health, reduced harm and saved resources. *PLoS ONE* 9 (7), pp. e102670. DOI: 10.1371/journal.pone.0102670.

Collaboration for Environmental Evidence (2015): The Collaboration for Environmental Evidence. Available online at http://www.environmentalevidence.org/, accessed 6/13/2015.

Collins, A.; Miller, J.; Coughlin, D.; Kirk, S. (2014): The Production of Quick Scoping Reviews and Rapid Evidence Assessments: A How to Guide (Beta Version 2). Joint Water Evidence Group. Available online at https://sbri.innovateuk.org/documents/3058188/3918930/The+Production+of+QSRs+and+REAs-+A+How+to+guide.pdf/45975020-be7d-4788-b74b-f3b6ed32c73a.

European Commission (1/28/2011): Directive 2011/8/EU of 28 January 2011 amending Directive 2002/72/EC as regards the restriction of use of Bisphenol A in plastic infant feeding bottles, Directive 2011/8/EU. In : Official Journal of the European Union. Available online at http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2011:026:0011:0014:EN:PDF, accessed 6/15/2015.

European Commission (5/31/2011): Bisphenol A: EU ban on baby bottles to enter into force tomorrow. Brussels. Available online at http://europa.eu/rapid/press-release_IP-11-664_en.htm, accessed 2/17/2015.

European Food Safety Authority (2010): Application of systematic review methodology to food and feed safety assessments to support decision making. *EFSA Journal 2010; 8(6):1637. DOI:* 10.2903/j.efsa.2010.1637.

European Food Safety Authority (2015a): No consumer health risk from bisphenol A exposure. Press Release 21 Jan 2015. Parma. Available online at http://www.efsa.europa.eu/en/press/news/150121.htm, accessed 2/18/2015.

European Food Safety Authority (2015b): Outcome of the targeted consultation of the EFSA Journal editorial on increasing openness, robustness and transparency of scientific assessments. Available online at http://www.efsa.europa.eu/en/supporting/pub/785e.htm, accessed 8/4/2015.

European Food Safety Authority (2015c): Principles and process for dealing with data and evidence in scientific assessments. *EFSA Journal* (13(5):4121). DOI: 10.2903/j.efsa.2015.4121.

French Agency for Food, Environmental and Occupational Health & Safety (2013): Bisphenol A: ANSES demonstrates potential health risks and confirms the need to reduce exposure.

Available online at https://www.anses.fr/en/content/bisphenol-anses-demonstrates-potential-health-risks-and-confirms-need-reduce-exposure.

Plastics Europe (1/15/2015): French ban on the use of Bisphenol A (BPA) in food contact: In conflict with European law and risk assessment - severe distortion of the market - no safety benefit for consumers. Jasmin Bird. Available online at http://www.bisphenol-a-europe.org/uploads/Modules/Mediaroom/stm_re_french-bpa-ban-being-enforced-01-01-2015.pdf, accessed 6/15/2015.

Higgins, J.P.T.; Altman, D.G.; Gøtzsche, P.C.; Jüni, P.; Moher, D.; Oxman, A.D. et al. (2011): The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 343, pp. d5928. DOI: 10.1136/bmj.d5928.

Higgins, Julian P. T.; Green, S. (Eds.) (2011): Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration. Available online at http://handbook.cochrane.org/, accessed 2/18/2015.

Hoffmann, S.; Hartung, T. (2006): Toward an evidence-based toxicology. *hum exp toxicol* 25 (9), pp. 497–513. DOI: 10.1191/0960327106het648oa.

Hooijmans, C.R.; Rovers, M.; Vries, R.B. de; Leenaars, M.; Ritskes-Hoitinga, M. (2012): An initiative to facilitate well-informed decision-making in laboratory animal research: report of the First International Symposium on Systematic Reviews in Laboratory Animal Science. *Lab. Anim.* 46 (4), pp. 356–357. DOI: 10.1258/la.2012.012052.

Hooijmans, C.R.; Rovers, M.M.; de Vries, Rob B M; Leenaars, M.; Ritskes-Hoitinga, M.; Langendam, M.W. (2014): SYRCLE's risk of bias tool for animal studies. *BMC medical research methodology* 14, p. 43. DOI: 10.1186/1471-2288-14-43.

ICF International (2015): DRAGON: An Online Tool for Systematic Review. Available online at http://www.icfi.com/insights/products-and-tools/dragon-online-tool-systematic-review, accessed 8/4/2015.

International Agency for Research on Cancer (2006): IARC Monographs on the Evaluation of Carcinogenic Risks to Humans: Preamble.

Ip, S.; Hadar, N.; Keefe, S.; Parkin, C.; Iovin, R.; Balk, E.M.; Lau, J. (2012): A Web-based archive of systematic review data. *Systematic reviews* 1, p. 15. DOI: 10.1186/2046-4053-1-15.

Land, M.; Wit, C.A. de; Cousins, I.T.; Herzke, D.; Johansson, J.; Martin, J.W. (2015): What is the effect of phasing out long-chain per- and polyfluoroalkyl substances on the concentrations of perfluoroalkyl acids and their precursors in the environment? A systematic review protocol. *Environ Evid* 4 (1), p. 3. DOI: 10.1186/2047-2382-4-3.

Lau, J.; Ioannidis, J.P.A.; Schmid, C.H. (1998): Summing up evidence. One answer is not always enough. *The Lancet* 351 (9096), pp. 123–127. DOI: 10.1016/S0140-6736(97)08468-7.

Lau, J.; Rothstein, H.R.; Stewart, G.B. (2013): History & progress of meta-analysis. In Julia Koricheva, Jessica Gurevitch, Kerrie Mengersen (Eds.): Handbook of meta-analysis in ecology and evolution. Princeton: Princeton University Press, pp. Chapter 25.

France (12/24/2012): LOI n° 2012-1442 du 24 décembre 2012 visant à la suspension de la fabrication, de l'importation, de l'exportation et de la mise sur le marché de tout conditionnement à vocation alimentaire contenant du bisphénol A. *Legifrance.gouv.fr*. Available online at http://legifrance.gouv.fr/affichTexte.do;jsessionid=F6553AACC19D178279D8DF154EAC8558.tpdila17v_1?cidTexte=JORFTEXT000026830015, accessed 6/15/2015.

Macleod, M.R.; Ebrahim, S.; Roberts, I. (2005): Surveying the literature from animal experiments: systematic review and meta-analysis are important contributions. *BMJ* 331 (7508), p. 110. DOI: 10.1136/bmj.331.7508.110-b.

Mulrow, C.D. (1987): The medical review article: state of the science. *Annals of internal medicine* 106 (3), pp. 485–488.

National Food Institute, Denmark (2015): Evaluation of EFSA's new Scientific Opinion on Bisphenol A. Søborg, Denmark (REG-no. DK 30 06 09 46). Available online at http://www.food.dtu.dk/english/~/media/Institutter/Foedevareinstituttet/Publikationer/Pub-2015/Evaluation_BisphenolA.ashx?la=da.

Nordic Cochrane Centre (2014): Review Manager (RevMan). Version 5.3: Cochrane Collaboration. Available online at http://tech.cochrane.org/revman, accessed 6/18/2015.

Rooney, A.A.; Boyles, A.L.; Wolfe, M.S.; Bucher, J.R.; Thayer, K.A. (2014): Systematic review and evidence integration for literature-based environmental health science assessments. *Environmental health perspectives* 122 (7), pp. 711–718. DOI: 10.1289/ehp.1307972.

Rusyn, I.; Shapiro, A. (2013): Health Assessment Workspace Collaborative (HAWC). Version Solid Hammer: UNC-CH Software. Available online at https://hawcproject.org/, accessed 8/4/2015.

Salman, R.A.-S.; Beller, E.; Kagan, J.; Hemminki, E.; Phillips, R.S.; Savulescu, J. et al. (2014): Increasing value and reducing waste in biomedical research regulation and management. *The Lancet* 383 (9912), pp. 176–185. DOI: 10.1016/S0140-6736(13)62297-7.

Schünemann, H.J.; Moja, L. (2015): Reviews: Rapid! Rapid! Rapid! …and systematic. *Systematic reviews* 4, p. 4. DOI: 10.1186/2046-4053-4-4.

SciRAP (2014): Science in Risk Assessment and Policy. Department of Applied Environmental Science at Stockholm University; Institute of Environmental Medicine at Karolinska Institutet in Stockholm; MistraPharma. Available online at http://www.scirap.org/, accessed 3/11/2015.

Sena, E.S.; Currie, G.L.; McCann, S.K.; Macleod, M.R.; Howells, D.W. (2014): Systematic reviews and meta-analysis of preclinical studies: why perform them and how to appraise them critically. *J. Cereb. Blood Flow Metab.* 34 (5), pp. 737–742. DOI: 10.1038/jcbfm.2014.28.

Silbergeld, E.; Scherer, R.W. (2013): Evidence-based toxicology: strait is the gate, but the road is worth taking. *ALTEX* 30 (1), pp. 67–73.

Sterne, J.A.C.; Higgins, J.P.T.; Reeves, B.C. (2014): A Cochrane Risk Of Bias Assessment Tool for Non-Randomized Studies of Interventions (ACROBAT-NRSI). The Cochrane Collaboration. Available online at https://sites.google.com/site/riskofbiastool/, accessed 9/29/2014.

Stewart, G. (2010): Meta-analysis in applied ecology. *Biol. Lett.* 6 (1), pp. 78–81. DOI: 10.1098/rsbl.2009.0546.

Stewart, G.B.; Schmid, C.H. (2015): Lessons from meta-analysis in ecology and evolution: the need for trans-disciplinary evidence synthesis methodologies. *Research synthesis methods* 6 (2), pp. 109–110. DOI: 10.1002/jrsm.1152.

Tošenovský, E. (2014): Question for written answer to the Commission, Rule 130. European Parliament, Parliamentary questions, P-008546/2014, Subject: Possible negative impact on the internal market of measures concerning BPA adopted by the French authorities. 30 October 2014. Available online at http://www.europarl.europa.eu/sides/getDoc.do?type=WQ&reference=P-2014-008546&language=EN, accessed 6/15/2015.

Tošenovský, E. (2015): Question for written answer to the Commission, Rule 130. European Parliament, Parliamentary questions, E-004315-15, Subject: Measures concerning Bisphenol A. 17 March 2015. Available online at http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+WQ+E-2015-004315+0+DOC+XML+V0//EN&language=en, accessed 6/15/2015.

UK Civil Service (2015): What is a Rapid Evidence Assessment? Available online at http://www.civilservice.gov.uk/networks/gsr/resources-and-guidance/rapid-evidence-assessment/what-is, accessed 6/13/2015.

US Environmental Protection Agency (2013): Process for Developing IRIS Health Assessments. Available online at http://www.epa.gov/IRIS/process.htm, accessed 6/16/2015.

US Food and Drug Administration (2014): Bisphenol A (BPA): Use in Food Contact Application. Update on Bisphenol A (BPA) for Use in Food Contact Applications. Available online at http://www.fda.gov/NewsEvents/PublicHealthFocus/ucm064437.htm.

US Institute of Education Sciences (2015): What Works Clearinghouse. Available online at http://ies.ed.gov/ncee/wwc/default.aspx, accessed 6/13/2015.

US National Research Council (2014a): A framework to guide selection of chemical alternatives. Washington, D.C.: The National Academies Press.

US National Research Council (2014b): Review of EPA's Integrated Risk Information System (IRIS) Process. Washington, D.C.: The National Academies Press.

US National Toxicology Panel (2015): Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration. Available online at http://ntp.niehs.nih.gov/ntp/ohat/pubs/handbookjan2015_508.pdf, accessed 1/13/2015.

van Luijk, J.; Bakker, B.; Rovers, M.M.; Ritskes-Hoitinga, M.; de Vries, Rob B M; Leenaars, M. (2014): Systematic reviews of animal studies; missing link in translational research? *PLoS ONE* 9 (3), pp. e89981. DOI: 10.1371/journal.pone.0089981.

Vandenberg, L.N.; Ehrlich, S.; Belcher, S.M.; Ben-Jonathan, N.; Dolinoy, D.C.; Hugo, E.R. et al. (2014): Low dose effects of bisphenol A. *Endocrine Disruptors* 1 (1), pp. e26490. DOI: 10.4161/endo.26490.

Woodruff, T.J.; Sutton, P. (2014): The Navigation Guide systematic review methodology: a rigorous and transparent method for translating environmental health science into better health outcomes. *Environmental health perspectives* 122 (10), pp. 1007–1014. DOI: 10.1289/ehp.1307175.

Zoeller, R.T.; Bergman, Å.; Becher, G.; Bjerregaard, P.; Bornman, R.; Brandt, I. et al. (2015): A path forward in the debate over health impacts of endocrine disrupting chemicals. *Environ Health* 14, p. 118. DOI: 10.1186/1476-069X-13-118.

Figure 2: Diagram showing the CRA process, whereby risk is a function of hazard and exposure. While SR methods could in principle be applied to all steps of the CRA process, it is the view of the workshop participants that up to this point in time most attention has been focused on the hazard identification and hazard characterisation steps. There are issues around conducting a systematic review for exposure assessment which were not discussed at the workshop, such as the requirement for a very different tool for assessing risk of bias in exposure studies which may necessitate very specialised knowledge of analytical chemistry.

Figure 3: The potential utility of SR methods in application to REACH registrations

Figure 4: Examples of conflicting opinions from scientists and government agencies about the risks to health posed by bisphenol-A at current exposure levels.