Lancaster Environment Centre

Evidence-Based Approaches to Chemical Risk Assessment and Risk Management Decision-Making

by

Taylor Ann Maree Wolffe (MChem)

in collaboration with

Yordas Group

This project was supported by the Centre for Global Eco-Innovation and is part financed by the European Regional Development Fund.

European Union
European Regional Development Fund

Abstract

Chemicals policy is designed to protect human and ecological health from the adverse effects that can result from exposure to manufactured chemical substances. It entails a complex process of regulatory chemical risk assessment and risk-management decision-making, drawing expertise from a diverse range of fields including toxicology and environmental health. However, these decision-making processes have come under increased scrutiny in recent years – criticized for bias, lack of transparency, rigor and a failure to identify unacceptable risks *before* widespread exposure occurs. This has resulted in calls for a more "evidence-based" approach, in which all relevant, available evidence is analyzed in a robust, transparent and reproducible manner. There is thus a growing need to incorporate methodological frameworks capable of facilitating evidence-based approaches to chemical risk assessment and regulatory decision-making.

Such frameworks have been successfully developed in the field of medicine, which underwent a similar paradigm shift to that currently shaping chemical risk assessment, in the early 1990s. The gold-standard for evidence-based decision-making championed by the evidence-based medicine movement takes the form of systematic review. Systematic review describes a prescriptive and transparent method for collating, appraising and analyzing all available, relevant evidence in answer to a specific research question. By pooling the results of individual (independent) studies, systematic reviews synthesize conclusions which are not only more precise but are representative of an entire evidence-base. Now well established within clinical decision-making, the application of systematic review to chemical risk assessment is beginning to gain prominence.

However, several challenges and barriers threaten to slow the uptake and quality of systematic review for chemical risk assessment. These include the prohibitively narrow focus of systematic reviews, which are at odds with the information requirements of regulatory decisions, and a mismatch in the resource availability within chemical risk assessment compared to the resource demands associated with systematic review.

This thesis explores the challenges associated with implementing evidence-based approaches such as systematic review for chemical risk assessment, and identifies key methodological solutions:

Chapter 1 examines the *risk of bias assessment* process – one of the most important but also most challenging aspects of systematic review methodology to adapt for environmental health. It examines the rationale for eschewing seemingly objective, quantitative approaches to assessing risk of bias in favour of seemingly more subjective, qualitative approaches. Through illustrative models, this thesis uncovers the mismatch between the mechanics of quantitative risk of bias assessment methods and the fundamental mechanics of risk of bias itself. Promoting understanding of this issue is increasingly important as systematic review gains prominence within chemical risk assessment – a field traditionally reliant on quantitative scoring methods for assessing the quality of included evidence.

Chapter 2 considers the wider challenges to uptake of systematic review in environmental health, and proposes "systematic evidence mapping" as a methodological solution. A systematic evidence map is a queryable database of systematically gathered evidence which facilitates the broader identification of trends across the evidence-base. In this thesis, the potential utility of systematic mapping for existing and future chemical risk assessment workflows is characterized and critically assessed. A hypothetical but representative example (in which legacy flame retardants are prioritized for further regulatory assessment) is used to demonstrate the trend-spotting capacity of the methodology.

Chapter 3 further explores the methodological adaptions required for effective implementation of systematic evidence mapping in chemical risk assessment and wider environmental health. By surveying current evidence mapping practice in environmental-management (a field where the methodology is more mature), and qualitatively appraising this practice against the concepts of "*data storage technology*", "*data integrity*", "*data accessibility*", and "*transparency*", this thesis reveals the ill-suited nature of conventional tabular data structures for housing complex and highly connected environmental health/toxicology data. It identifies graph-based storage technologies as the most flexible and optimally suited data structures for the varied needs of chemical risk assessment workflows, and makes recommendations for their uptake in systematic evidence mapping.

Chapter 4 of this thesis explores the practical implementation of graph-based solutions to evidence mapping in environmental health by conducting a proof-of-concept evidence mapping exercise, in which trends in the study of exposure-outcome associations for National Health and Nutrition Examination Survey (NHANES) datasets in the academic literature are explored. By contrasting this graph-based evidence mapping exercise to an equivalent tabular scoping review, this chapter demonstrates how significant gains in resolution and complexity can be achieved by adopting the graph data model – leading to greater insights than can be offered by traditional evidence-surveillance methods. The transparency, accessibility, interoperability and potential to expand graph-based evidence maps is also highlighted in this chapter by providing data models and methods which can be further adapted e.g. for the development of a suitable controlled vocabulary ontology.

Finally, this thesis concludes by discussing the future direction of evidence-based chemical risk assessment and the role of graph-based evidence mapping within it, highlighting the need for further advances in automation and the uptake of data standards.

Acknowledgements

I am immensely grateful for the expertise, interest and/or enthusiasm of so many people who have supported me in one way or another along this journey. Of special mention is my supervisor, Professor Crispin Halsall – for his reassurances, helpful insight and unwavering confidence in me. I also extend special thanks to Paul Whaley for his mentorship over the past three years; for sharing his knowledge, experience and infectious interest in the field of evidence-based toxicology; and for introducing me to a network of excellent evidence-synthesis researchers and professionals.

I would also like to thank Dr Rosalinda Gioia, and especially Dr Neil Hunt at Yordas Group for imparting their valuable regulatory and chemical risk-assessment expertise; and for their support of, and genuine interest in, my research.

Thanks to all the team at the Centre for Global Eco-Innovation for their support, and especially to Jan Parkinson for her endless patience and speedy processing of admin queries.

Thanks to my wonderful husband, Mike Wolffe – for his patience and fierce love and support. Thanks also to my family for their support; to my parents, Isobel and Rob Harrison for cheering me on; and to my brother, Mitchell Harrison for help with troubleshooting broken code.

Finally, thanks to all the animal companions that have kept my spirits high over the past few years. To Toffee. To Pecan, Pumpkin and Poppyseed. And to Pig. I miss you terribly, thank you.

Table of Contents

Abstract ..... **i**

Acknowledgements ..... **iv**

Contents ..... **v**

List of Abbreviations ..... **vii**

List of Tables ..... **ix**

List of Figures ..... **xi**

Introduction **1**

Background 1

Chemicals regulation, risk assessment and risk-management 1

Regulatory failings, data availability and regulatory reform 2

Continued discrepancies in chemical risk assessment 5

Evidence-Based Approaches 7

Systematic review 7

Aims and structure of this thesis 11

Objective 1: Understand the challenges associated with implementing systematic review in chemicals policy and wider environmental health 12

Objective 2: Seek methodological solutions which facilitate the uptake of systematic review and other evidence-based approaches in chemicals policy 13

Objective 3: Characterise methodological solutions in the context of environmental health and toxicology, identifying specific adaptations required for chemicals policy settings..... 14

Objective 4: Explore the practical application of these methodological solutions to environmental health research problems, identifying remaining challenges and clarifying the direction of future work..... 15

Chapter 1: Scales, Scoring and Subjectivity: Assessing risk of bias when conducting systematic reviews of the environmental health literature **17**

Chapter 2: Systematic evidence maps as a novel tool to support evidence-based decision-making in chemicals policy and risk management **51**

Chapter 3: A survey of systematic evidence mapping practice and the case for knowledge graphs in environmental health & toxicology..... **62**

Chapter 4: A graph-based approach to mapping human exposure-outcome associations for chemical contaminants **82**

Conclusion **116**

Future work **120**

References **124**

Appendix: PFAS health effects database: Protocol for a systematic evidence map **156**

List of Abbreviations

ANSES	Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail (French Agency for Food, Environmental and Occupational Health & Safety)
BFR	Brominated Flame Retardant
BPA	Bisphenol-A
CDC	Centers for Disease Control and Prevention
CEE	Collaboration for Environmental Evidence
DDT	Dichlorodiphenyltrichloroethane
EBM	Evidence-based medicine
EBT	Evidence-based toxicology
ECHA	European Chemicals Agency
EFSA	European Food Safety Authority
EH	Environmental Health
EPA	Environmental Protection Agency
EU	European Union
FR	Flame Retardant
IUCLID	International Uniform Chemical Information Database
NHANES	National Health and Nutrition Examination Survey (United States)
NTP-OHAT	National Toxicology Program's Office for Health Assessment and Translation (United States)
PAH	Polycyclic aromatic hydrocarbons
PCBs	Polychlorinated biphenyls
PECO	Population, Exposure, Comparator, Outcome

PFAS	Per- and polyfluoroalkyl substances
PFBS	Perfluorobutane sulfonic acid
PFC	Perfluorinated compound
PFOA	Perfluorooctanoic acid
PFOS	per-fluorooctanesulfonic acid
REACH	Registration, Evaluation, Authorisation and Restriction of Chemicals
RDF	Resource Description Framework
RM	Risk Management
SEM	Systematic evidence map
SPARQL	SPARQL Protocol and RDF Query Language
SR	Systematic review
SQL	Structured Query Language
TDI	Tolerable Daily Intake
TEDX	The Endocrine Disruption Exchange
TSCA	Toxic Substances Control Act (United States)
UMLS	Unified Medical Language System
US	United States
VOC	Volatile organic compound
WHO	World Health Organization
WoE	Weight of Evidence

List of Tables

Chapter One: Assessing risk of bias when conducting systematic reviews of the environmental health literature

Table 1: A simple quality scale comprised of nine quality criteria, (A to I), which each represent a different aspect of quality within a study (e.g. use of controls, blinding, etc.). One point is awarded for every criterion fulfilled, and points are subsequently summed to produce an overall score (e.g. 4 out of 9)..... 28

Table 2: The assumption that the magnitude of bias introduced by failing to meet a criterion is equal for all criteria allows a summary score to mean the same thing in all scenarios (e.g. Study A compared to Study B above), and allows the score to accurately scale studies; those with a higher score suffer a smaller magnitude of bias (i.e. Study C compared to Study A or B)..... 30

Table 3: Removing the assumption that the magnitude of bias introduced by failing to meet a criterion is equal for all criteria breaks the ability of the same score to represent the same magnitude of bias (e.g. Study A compared to Study B above). Additionally, removing the assumption that bias is unidirectional breaks the ability of a score to scale with quality, as it is no longer true that a higher score necessarily accounts for a lower sum magnitude of bias compared to a lower score (e.g. Study C compared to Study A). 33

Chapter Two: Systematic evidence maps as a novel tool to support evidence-based decision-making in chemicals policy and risk management

Table 1: The key features of systematic reviews and their primary advantages. PECO=Population-Exposure-Comparator-Outcome. 54

Table 2: A comparison of systematic review and systematic evidence mapping methodology and their respective roles in risk management decision-making (adapted from James et al., 2016). SR=systematic review, SEM=systematic evidence map, RM=risk management, TDI=tolerable daily intake. 55

Chapter Three: A survey of systematic evidence mapping practice and the case for knowledge graphs in environmental health & toxicology

Table 1: The concepts used to guide data extraction and subsequent assessment and discussion of the outputs of CEE systematic mapping exercises. 65

Table 2: Description of the storage technologies used by CEE systematic maps. 67

List of Figures

Introduction

Figure 1: Brief outline of the methodological stages involved in conducting a systematic review. All proposed methods are clearly defined in a pre-published protocol, holding reviewers accountable to their methods and creating an opportunity for peer-review and stake-holder input. Several stages of the review process are conducted in duplicate by two independent reviewers – minimizing the influence of human error and ensuring consistent understanding of the research methods and objectives. 9

Chapter One: Assessing risk of bias when conducting systematic reviews of the environmental health literature

Figure 1: Key phases of the systematic review process. 24

Figure 2: The Simple Model of scoring study quality assumes a linear, proportional relationship between magnitude of error and study score. As study score increases, bias decreases. 31

Figure 3: Revising the assumptions underpinning the Simple Model. Expanding on Table 3, the total magnitude of bias associated with every possible way of achieving a score on the artificially generated scale is displayed, illustrating the range of meaning a single score might serve to mask. 34

Chapter Two: Systematic evidence maps as a novel tool to support evidence-based decision-making in chemicals policy and risk management

Figure 1: The process of identifying trends and exploring the evidence landscape involves querying the SEM database and visualizing the results of the query. Queries may start by asking broader questions which consider a wider range and volume of data (e.g. Queries 1 and 2). Users may then further explore any trends of interest discovered in the results of these broad queries by running narrower queries which consider a more specific subset of data (e.g. Queries 3 and 4). Data displayed in this Figure have been artificially generated to illustrate a hypothetical use case for SEMs. FR = flame retardant, TDI = tolerable daily intake, SEM = systematic evidence map. 57

Chapter Three: A survey of systematic evidence mapping practice and the case for knowledge graphs in environmental health & toxicology

Figure 1: Publication history of CEE SEMs indicating the number of maps published per year. The year in which the CEE guidance on systematic mapping methods was published (2016) is marked on the corresponding bar (James et al., 2016). 66

Figure 2: The number of CEE systematic maps that are structured with expanding rows and/or expanding columns as a means of preserving data relationships. Maps using the relational database storage technology are presented on the left, while maps using the spreadsheet storage technology are presented on the right. In addition, the numbers of systematic maps which store multiple values within a single cell of their data table/s are indicated by solid shading, whereas those that do not are indicated by patterned shading. 67

Figure 3: Illustrative example of how “expanding-rows” (A) and “expanding columns” (B) are used in long-form (A) and wide-form (B) tabular data structures,

respectively. 67

Figure 4: A, The relationship between attribute A and entity 3 is explicit in the formal structure of the array. However, the relationship between attribute A and attribute C is implicit and has to be inferred by the user from features external to the table eg, conventions around interpreting tabular data. The external conventions are not part of, or known to, the table and may not be known to the user. B, For example, a user may (in this case, correctly) infer that "sex" is a property of "species" and not "outcome," but this inference is made using external conventions and contextual understanding—the relationship is not in fact known to the table. All the table can assert is that each entity 1 through 6 has a relationship to properties of sex, age, species, and outcome, respectively. 69

Figure 5: A, Loss of referential integrity resulting from the column expansion of more than 1 study attribute (data field). The recording of multiple populations and multiple outcomes on a single row compromises the ability of users to decipher which population was affected by which outcome. The table asserts that both populations (mice and rats) were affected by all 3 outcomes (reduced birth weight, tumors, and behavioral changes), respectively—which may not be truly representative of the raw data, compromising data integrity. B, This is similarly observed when multivalued cells are used for more than 1 study attribute. 70

Figure 6: (A) Knowledge captured in unstructured, textual formats e.g. scientific articles, is distributed and programmatically inaccessible. (B) This knowledge can be structured in an intuitive and machine-readable way as a series of semantic subject-predicate-object triples – where entities are the subjects and/or objects and the relationships between entities are the predicates. (C) Entities can be stored as the nodes of a graph. The semantic value of the relationships between entities are

preserved and stored as edges. The graph can continue to grow to produce a queryable representation of all knowledge on a topic (see Figure 7). 71

Figure 7: Storing relationships as first class entities allows knowledge graphs to continue to grow and expand without needing to revise schema and migrate data. This flexibility is particularly useful when relationships between entities cannot be characterised a priori. 72

Chapter Four: A graph-based approach to mapping human exposure-outcome associations for chemical contaminants

Figure 1: Graph data model describing the relationships between publications and the exposure-outcome associations they report. Neo4j's labelled property graph model was exploited to represent "RefID", "Title", "Authors" and "Year" as properties of Publication nodes. 92

Figure 2: (A) Chemical exposure nodes are categorised by a "CODED_AS" relationship to nodes housing controlled vocabulary terms. (B) Similarly, health outcome nodes are categorised by a "CODED_AS" relationship to nodes housing controlled vocabulary terms. Due to the variation and complexity of health outcomes, multiple controlled vocabulary terms may be used to categorise a single outcome. 93

Figure 3: The number of associations (x axis) which investigated a chemical exposure within each of the exposure categories (y axis), across all included publications. ... 95

Figure 4: The number of associations (x axis) which investigated a health outcome within each of the outcome categories (y axis), across all included publications. . .. 96

Figure 5: (A) Association-level heat map illustrating the frequency with which exposure and outcome categories have been associated, across all included publications i.e. a single publication reporting associations between "cadmium, blood" and "lipid levels" as well as "lead, blood" and "lipid levels" would count as

two associations between “metals/metalloids” and “body weight/metabolism”. (B) Publication-level heat map illustrating the frequency with which independent publications report associations between exposure and outcome categories i.e. a single publication reporting associations between “cadmium, blood” and “lipid levels” as well as “lead, blood” and “lipid levels” would count only once as reporting associations between “metals/metalloids” and “body weight/metabolism”.. 98

Figure 6: Brief outline of a potential workflow for devising a controlled vocabulary ontology which describes the availability of data within the NHANES datasets. Variables which are more specific to particular survey cycles (e.g. dates) might also be incorporated into the organisational structure of the ontology. 107

Introduction

Background

Chemicals regulation, risk assessment and risk-management

Manufactured chemicals are ubiquitous to all aspects of modern life. Designed to perform a range of functions, they are integral to the consumer goods and industrial processes on which society relies. The chemicals industry is beneficial for improving standards of living and life expectancy, as well as promoting economic growth. However, many of its products can additionally have unintended and/or unforeseen negative impacts on human and environmental health (Egeghy et al., 2012; Koch & Ashford, 2006; Schwarzman & Wilson, 2009). It is therefore vital to implement a system able to identify, weigh and control the risk of such adverse outcomes. This function is served by the chemicals regulation system, which promotes maximum benefit of manufactured chemicals by minimising the harmful consequences of their use, driving industry towards safer alternatives.

Chemicals regulation achieves these aims by setting limits that control a population's exposure to a chemical through restricting its manufacture, distribution and disposal or controlling its approved uses. In the European Union (EU), the European Chemicals Agency (ECHA) implements such legislation through the REACH regulation (Registration, Evaluation, Authorisation and Restriction of Chemicals).

The severity of regulatory action taken against a chemical substance is determined by assessing the risk associated with exposure to that substance.

Chemical risk assessment is a complex process comprised of four key stages (Beronius & Vandenberg, 2015; National Academy of Sciences, 1983):

- *hazard identification* - assessing the hazardous properties intrinsic to a chemical;
- *hazard characterisation* – charactering the relationship between chemical dose and biological response, i.e. investigating the mode and level of exposure required for the substance's hazardous properties to affect an adverse outcome;
- e*xposure assessment* - investigating and/or estimating the potential sources and severity of a population's exposure to a chemical;
- *Risk characterisation* – combining hazard and exposure data to determine the magnitude of the risk posed by a chemical substance.

A variety of heterogeneous data sources are relevant to each of these stages, encompassing physio-chemical, *in-vitro*, *in-vivo*, *in-silico* and human epidemiological studies, as well as environmental- or bio-monitoring and exposure studies. Chemical risk assessment, and the subsequent risk-management process, draw together these varied scientific disciplines when reaching overall conclusions on the safety of a chemical substance. It is the challenges associated with drawing together, managing and synthesising data from disparate sources, and the consequences of failing to meet those challenges, which motivate the research discussed in this thesis.

Regulatory failings, data availability and regulatory reform

Chemicals regulation is designed to protect human and environmental health (Abelkop & Graham, 2014). However, regulatory decisions have not always been successful in meeting this aim. Chemical policy's brief history is marred by case studies of regulatory failure, where substances allowed to market are later confirmed to be of significant harm to public and/or

environmental health (Commission of the European Communities, 2001; European Environment Agency, 2013). The adverse effects which result from such widespread exposures can be severe, irreversible but also long-lasting – as persistent and/or bio-accumulative substances continue to cause harm many years after reactive regulatory action reduces or eliminates sources of exposure. This is well illustrated by case studies concerning exposures to infamous legacy chemicals such as dichlorodiphenyltrichloroethane (DDT) (Commission of the European Communities, 2001) and polychlorinated biphenyls (PCBs) (Silbergeld et al., 2015).

Consequently, regulatory frameworks and the chemical risk assessment process have come under increased scrutiny in recent times (Whaley et al., 2016). Lack of sufficient toxicological data, and failure to adopt a precautionary approach in light of such data gaps, have been cited as key flaws of regulatory frameworks (Applegate, 2008; Eckley & Selin, 2004). In a 1998 report by the US EPA, it was estimated that a full set of basic toxicity data was only available for 7% of the high production volume chemicals (produced or imported at or above 1 million pounds per year) produced in the US – with 43% of those chemicals lacking any human or environmental toxicity data at all (EPA's Office of Pollution Prevention and Toxics, 1998). Similarly, in a white paper published by the European Commission, it was indicated that some 80,000 legacy chemicals (released to market prior to 1981) in use across Europe had undergone no formal risk assessment (Brown, 2003; Commission of the European Communities, 1998, 2001).

These issues are being addressed by initiatives to reform reactive chemicals regulation systems toward more proactive systems, where the risks associated with exposure to chemical substances can be assessed and managed prior to their release (Abelkop & Graham, 2014; Commission of the European Communities, 2001; United States Environmental Protection Agency, 2016).

The largest, most ambitious and complex of these reforms is that of the European Union's REACH (Registration, Evaluation, Authorisation and Restriction of Chemicals) regulation (Commission of the European Communities, 2001; European Comission, 2016). In contrast to the frameworks which it replaces, REACH operates under a "no data, no market" ethos (European Comission, 2016). It shifts the burden of demonstrating safety away from regulators and onto manufacturers. Entered into force in June 2007, REACH has seen data gaps for tens of thousands of chemicals filled within its 10 year registration period (ECHA, 2019a). While not as extensive as REACH, the recent reform of the US Toxic Substances Control Act (TSCA) has similarly introduced measures designed to fill data gaps (Schmidt, 2016; United States Environmental Protection Agency, 2016).

However, with so many legacy chemicals in commerce (Abelkop & Graham, 2014), and with new chemicals continually approaching market, filling data gaps represents a considerably resource intensive task. A key feature of REACH designed to avoid redundant repeat toxicity testing is the requirement that applicants registering a chemical substance share toxicity data with other manufacturers of that substance. This data is made available in the International Uniform Chemical Information Database (IUCLID), which has been described as the "world's largest database on the properties of chemical substances" (Buxton, 2017; European Comission, 2016).

To further improve resource efficiency and the minimisation of expensive and/or unethical *in-vivo* toxicity testing, REACH and other reformed chemicals policy workflows emphasise the need to identify, and make best use of, *pre-existing* data. For example, pre-existing data can be used within REACH in read-across applications, where predictions regarding the toxicological behaviour of data-poor substances can be made by evaluating structurally similar data-rich substances (Schaafsma et al., 2009; Vink et al. 2010); or in weight-of-evidence assessments (WoE) – where, although data for a specific toxicological endpoint may be

insufficient, the data gap can be addressed by combining related data from several independent pre-existing sources (ECHA, 2019b; Schoeters, 2010).

Continued discrepancies in chemical risk assessment

In tackling the toxicity data gap, reformed regulatory processes such as REACH serve to demonstrate how lack of data is a solvable and diminishing challenge for chemicals policy. Despite this, concerns over the generation, identification and/or synthesis of toxicologically relevant data continue to be raised (Hartung, 2009; Hoffmann & Hartung, 2006). Consequently, the chemical risk assessment process has been criticised for issues concerning conflicts of interest, poor transparency (Ingre-Khans et al., 2016), poor reproducibility and a continuing tendency to miss "early-warnings" (European Environment Agency, 2013; Hoffmann & Hartung, 2006). These concerns are in-part founded by the fact that while *more* data is available for chemical risk assessment (i.e. endpoints for a larger suite of toxicological, chemical and exposure testing) the frameworks for collecting, managing and appraising this data remain unsystematic and opaque. This is a key criticism of the current REACH registration and assessment frameworks (Ingre-Khans et al., 2016), whereby the methods used to collate and select the evidence presented in registration dossiers and chemical risk assessments are inaccessible.

This makes it difficult to determine whether the data selected for chemical risk assessment is in fact representative of *all available evidence*, or whether this data has been cherry picked; potentially by an industry with vested interests in a substance's regulatory approval (a threat of particular relevance to REACH (Ingre-Khans et al., 2016)). Appraising the assessment process itself, in which selected data are analysed and evaluated to reach conclusions on a chemical's safety, is similarly difficult. This process has traditionally relied upon expert elicitation (Ingre-Khans et al., 2016; Morgan, 2014) – where a panel of specialists with varied expertise interpret primary toxicity and exposure data for the wider human or environmental

context. However, without an objective, consistent and robust framework to guide expert assessment, the conclusions of this process may be biased by the variation in methodological choices made by assessors, and by the variation in individual knowledge, experiences and opinions which exist from one expert to the next (Rudén, 2001b; Whaley et al., 2016).

Several case studies demonstrate the discrepancies which can arise from an unsystematic and opaque chemical risk assessment process, where different assessors reach conflicting conclusions regarding a chemical's safety despite access to the same evidence base (Hoffmann & Hartung, 2006; Whaley et al., 2016). Examples range from inconsistent and/or contradictory conclusions between two risk assessments (e.g. PCBs (Golden et al., 2003)) to multiple risk assessments (e.g. trichloroethylene, for which 29 assessments reached varied conclusions (Rudén, 2001b, 2001a)). Similarly, discrepant risk assessments lead to contradictory and conflicting regulatory action from one authority to the next (e.g. between European Food Safety Authority (EFSA) and the French Agency for Food (ANSES) regarding the regulation of Bisphenol-A (BPA) (Whaley et al., 2016)).

Such discrepancy causes uncertainty, confusion and a lack of trust in the ability of chemicals policy to protect human and environmental health. Thus, robust, transparent and systematic methodological frameworks are required to ensure that chemical risk assessments avoid bias and discrepancy. At the very least, increased methodological rigour and transparency would allow sources of discrepancy to be identified and assessed by *all* stakeholders. Such methodological frameworks are offered by evidence-based approaches such as systematic review (SR) and systematic evidence mapping.

Evidence-Based Approaches

Systematic review

Systematic review is a method of systematically gathering, appraising and synthesising all relevant and available evidence such that a single, representative answer to a specific research question can be derived from the pooled results of individual, independent studies.

The steps of the methodology are organised within a consistent and prescriptive framework (summarised briefly in Figure 1 and elaborated further in Table 2 of Chapter 2). Each step advocates transparency, and is designed to ensure the rigour of the review and/or the representativeness of its pooled finding/s. Briefly, systematic review builds on the methods of traditional narrative reviews in several key ways (see Chapters 1 and 2 for further detailed discussion of systematic review methods):

- All methodological decisions are planned ahead of commencing the review and are specified in a pre-published protocol. This holds reviewers accountable to their methods and prevents the kind of ad-hoc analyses which introduce bias and discrepancy. Pre-published protocols also increase transparency and reproducibility, allowing others to critically appraise the methods via which a review conclusion has been reached and to update the review in the future. Pre-publication of systematic review protocols also offers an opportunity for peer-review of planned methods, allowing any potential issues or sources of bias to be amended prior to conducting the review itself.
- Systematic searches form the basis of the evidence gathering step of systematic reviews. A systematic search consists of a series of search strings formatted for specific bibliographic databases. These search strings are designed to cover all key concepts relevant to the review question and ensure that the search returns as much relevant

information as is available. Reporting the search strategy (i.e. the combination of search strings and bibliographic databases searched) in the review's protocol ensures the search can be updated in the future.

- In contrast to narrative reviews, which rely on ad-hoc processes of literature searching and selection, often shaped by the reviewer's own expertise and interests – systematic reviews ensure that the *all* search results returned via a systematic search are considered for inclusion in the review (against a set of pre-defined inclusion criteria). Giving equal consideration to all returned results ensures that no potentially relevant information is omitted from the review.
- Assessing the risk that studies included in a systematic review are biased is a key feature of systematic review methodology (see Chapter 1) – which aims to consider the potential impacts that this bias might have on the results of the review.
- In contrast to narrative reviews – which may have a broader focus or may not be driven by a narrowly focused research question e.g. instead describing the "state of the science" within a field– systematic reviews address specific, closed-framed research questions. They employ narrative or statistical methods (as appropriate) for deriving an overall answer to this research-question.
- Many of the steps of the systematic review process are conducted in duplicate by at least two independent reviewers, including the literature screening, data extraction and risk of bias assessment processes. This ensures the rigour of the review process and helps to protect the review against human error.

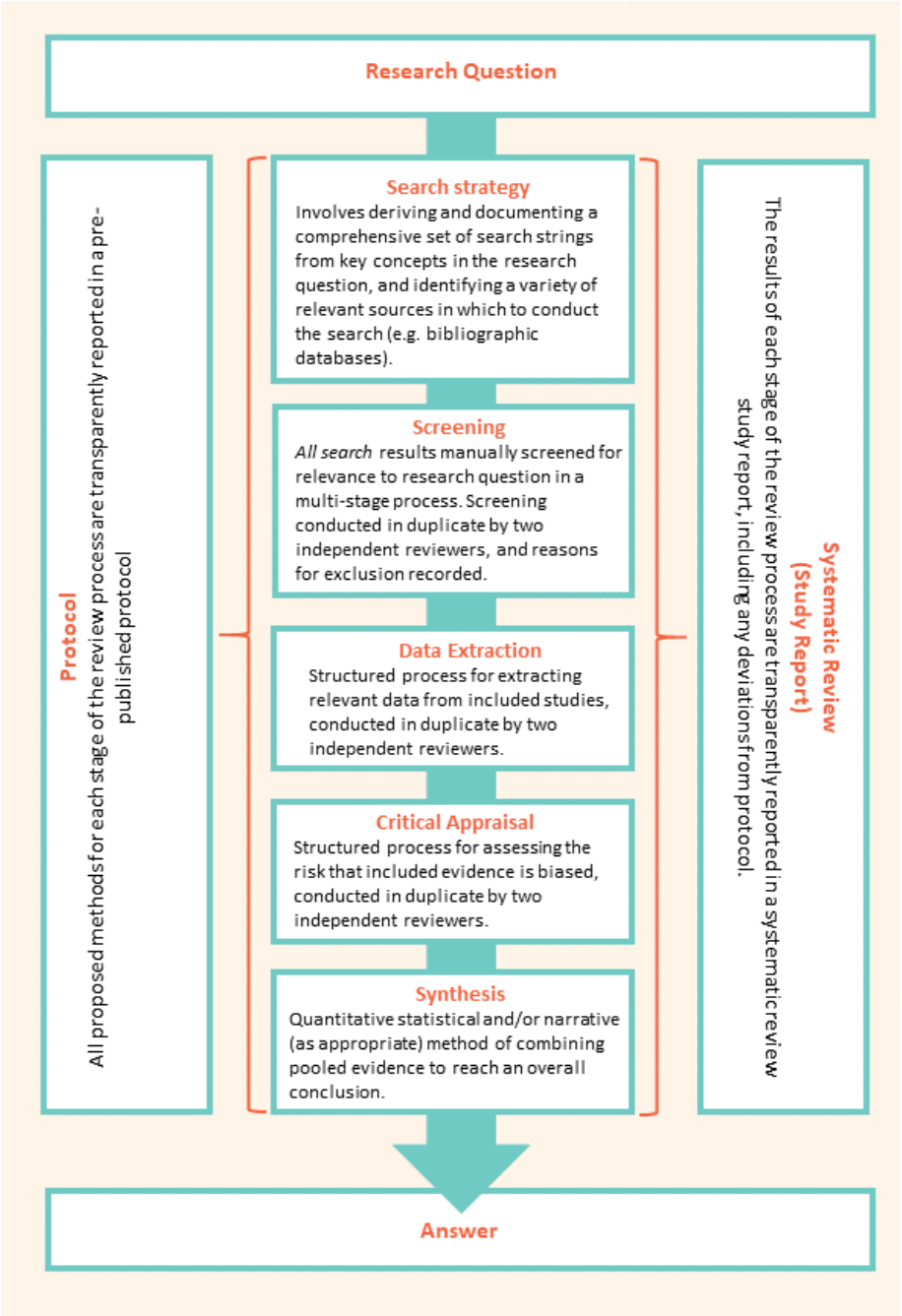

Figure 1: Brief outline of the methodological stages involved in conducting a systematic review. All proposed methods are clearly defined in a pre-published protocol, holding reviewers accountable to their methods and creating an opportunity for peer-review and stake-holder input. Several stages of the review process are conducted in duplicate by two independent reviewers – minimizing the influence of human error and ensuring consistent understanding of the research methods and objectives.

The origins of systematic review methodology lie in the field of medicine, where it was introduced as a tool for clinical decision-making. It arose out of the evidence-based medicine (EBM) movement in the early 1990s (Chalmers et al., 2002; Hooker, 1997), which sought to shift the paradigm of clinical-decision making away from its reliance on expert opinion, experience and intuition, and toward the more holistic consideration of best current available evidence (Guyatt et al., 1992). As well as increasing the robustness and precision of clinical decisions, the introduction of systematic review sought to increase transparency and accountability within a field plagued by discrepancy and bias (Goldacre, 2013). It has since become firmly established in the field, with organisations such as the Cochrane Collaboration (The Cochrane Collaboration, 2019) leading the production of accessible high-quality systematic reviews, methodological guidance and standards (Lefebvre, 1994).

The similarity between the issues faced in clinical decision-making and chemicals policy settings, and the demonstrable success of systematic review in overcoming these issues in the field of medicine have led to increasing interest in the application of systematic review to chemicals policy contexts such as chemical risk assessment (Whaley et al., 2016). Calls for a move toward evidence-based toxicology (EBT) (Hoffmann & Hartung, 2006) have seen systematic reviews on environmental health topics begin to emerge (Whaley & Halsall, 2016) along with networks and collaborative workgroups (e.g. (Johns Hopkins Bloomberg School of Public Health, 2019; NTP-OHAT, 2019; The Endocrine Disruption Exchange, 2019b; UCSF Program on Reproductive Health and the Environment, 2019)) dedicated to establishing the methodology within the field. While regulatory agencies such as ECHA are yet to incorporate systematic review in their assessment frameworks, the US EPA have recently taken up the methodology for TSCA risk evaluations (EPA, 2018).

Although increasing interest in systematic review is indicative of progress within the EBT movement, evidence-based methods are still relatively novel to the field and several barriers

to their effective adaptation and widespread uptake remain. In clinical settings, evidence synthesised in a systematic review is typically derived from studies of similar design (i.e. randomised controlled trials) in answer to a well-defined research question (e.g. *Is x an effective treatment for y in population z?*). However, evidence from a heterogenous range of study designs must be integrated when addressing research questions regarding chemical risk assessment. The focus of those questions is also more complex to define for chemical risk assessment where no measure of a single, consistent outcome is necessarily indicative or sufficiently informative of toxicity in the same way as the prevalence of a single, consistent outcome might be indicative of an effective clinical treatment.

Despite methodological guidance for addressing these and other challenges specific to environmental health systematic reviews (Hoffmann et al., 2017; NTP, 2015; Woodruff & Sutton, 2014), examples of questionable methodological conduct continue to emerge (Whaley & Halsall, 2016). Similarly, despite the growing presence and application of systematic review in chemical risk assessment contexts, ECHA have yet to follow the US EPA in adopting the methodology. This indicates a need for further research into the successful implementation of evidence-based methodologies in environmental health, as well as the need to disseminate such research to stakeholders working within a chemical risk assessment capacity – including regulatory bodies such as ECHA and the US EPA. It is this need which motivates the research discussed in this thesis.

Aims and structure of this thesis

This thesis aims to explore the adaptation and application of evidence-based methods for chemical risk assessment and risk management decision-making within chemicals policy and wider environmental health. This overarching aim is met through four key objectives. These objectives, and the chapters in which they are addressed, are briefly summarised below.

Finally, this thesis concludes by outlining the future work required to successfully implement evidence-based approaches to chemical risk assessment at scale.

Objective 1: Understand the challenges associated with implementing systematic review in chemical risk assessment and wider environmental health.

Chapter 1 of this thesis focuses on a significant challenge for systematic review: the use of quantitative systems for assessing risk of bias in included studies. Assessing the risk that studies included in a systematic review are biased is a key step of the systematic review process (Fig. 1) and determines the degree to which the conclusions of a systematic review can be trusted. The Cochrane Collaboration advises against the use of quantitative, scoring-based systems for assessing risk of bias in included studies, and offers an alternative, qualitative "domain-based" approach (Higgins, 2011). Despite this advice, quantitative scoring-systems have become a prevalent issue in the field of medicine.

As with the Cochrane Collaboration, there is an understanding of these issues among the workgroups dedicated to establishing systematic review in environmental health. Guidance published by The National Toxicology Program's Office for Health Assessment and Translation (NTP-OHAT) (OHAT, 2015) and the Navigation Guide (Woodruff & Sutton, 2014) both advise against the use of quantitative scoring systems and instead offer guidance for making qualitative, domain-based risk of bias assessments. However, as in the field of medicine, the allure of quantitative scoring systems for assessing risk of bias threaten the robustness and transparency of systematic review practice in environmental health – especially while the methodology and its associated best-practice are still relatively novel to the field. This can be evidenced by the fact that, despite methodological guidance advocating otherwise, the systematic review methodology adopted by the US EPA for TSCA risk evaluations uses a numeric scoring system for risk of bias assessment (EPA, 2018). Similarly, the critical appraisal process associated with chemical risk assessments conducted under REACH adopt the Klimisch

criteria (Klimisch et al., 1997) – a quality scale in which studies are rated using numeric judgements. Chapter 1 addresses the issue of quantitatively assessing risk of bias of included studies. It aims to improve understanding of systematic review methodology (and its motivating rationale) within environmental health – learning from the challenges encountered in the field of medicine and warning against their introduction to chemical risk assessment.

However, there are several additional challenges associated with the implementation of systematic review which are more specific to chemical risk assessment. These challenges stem from the higher degree of heterogeneity present in studies relevant to chemical risk assessment, as well as the breadth of research questions which must be assessed in chemical risk assessment. Additionally, in striving for robustness, systematic review makes significant demands on time and resources. Such demands are at odds with the increasingly strained availability of resources in chemicals policy (Pool & Rusch, 2014). These challenges are discussed further in Chapter 2, before the proposition of a methodological solution.

Objective 2: Seek methodological solutions which facilitate the uptake of systematic review and other evidence-based approaches in chemical risk assessment.

Chapter 2 identifies systematic evidence mapping as a promising methodological solution for overcoming many of the barriers associated with pursing evidence-based approaches to chemical risk assessment. A systematic evidence map (SEM) is a queryable database of references, data and meta-data which provides a use with computational access to the wider evidence-base. SEMs share much of their methodology with systematic review (see Table 2 of Chapter 2), but do not synthesise an overall conclusion and are not motivated by a single specific research question (Clapton et al., 2009; James et al., 2016). Instead, the purpose of a SEM is to characterise the evidence base more broadly – such that trends in the type, availability and outcomes of research can be investigated by end-users. This facilitates the

rapid identification of issues of emerging concern and allows resources to be more efficiently targeted (e.g. by focusing primary research efforts on evidence-gaps and secondary research efforts on evidence-clusters). Additionally, the breadth of SEMs allows a single mapping exercise to meet the needs of several, varied end-users – maximising the return on resource investment for these evidence-products.

To some degree, the IUCLID database which houses chemical risk assessment data for REACH already emulates the output of a systematic evidence map – although the data it houses remain unsystematically curated, and inaccessible for broader query by varied stakeholders. Introducing systematic evidence mapping methodology offers a potential resolution of these issues.

Objective 3: Characterise methodological solutions in the context of environmental health and toxicology, identifying specific adaptations required for chemical risk assessment.

Systematic evidence mapping conceptualises the barriers associated with implementing evidence-based methods as a problem of data management and access, approaching the issue with the transparency and robustness associated with systematic review. Although novel to chemical risk assessment, the methodology has been successfully applied in the social and wider environmental sciences (Clapton et al., 2009; James et al., 2016). Chapter 3 studies this successful application for lessons applicable to adapting the methodology for environmental health.

By conceptualising current evidence mapping practice through the lens of environmental health, Chapter 3 identifies key methodological considerations of relevance to chemical risk assessment applications. Most notably, this chapter demonstrates how the rigid, tabular data

structures favoured in current evidence mapping practice are ill-suited to housing environmental health data. This is owed to the complexity of environmental health data – which is not only highly heterogeneous, but also highly connected. Looking beyond traditionally employed databasing solutions, Chapter 3 identifies knowledge graphs as the future of evidence-mapping in environmental health. The flexibility of the graph data model, and its ability to preserve complex connections increases transparency and access to the evidence-base and is readily compatible with increasing research efforts in machine-learning and automation within the field.

Objective 4: Explore the practical application of these methodological solutions to environmental health research problems, identifying remaining challenges and clarifying the direction of future work.

Interest in systematic evidence mapping is beginning to accelerate in the field of environmental health (e.g. (Beverly, 2019; NTP-OHAT, 2019; The Endocrine Disruption Exchange, 2019a)), with the first protocol for an environmental health SEM recently published in the Environment International journal ((Pelch et al., 2019), Appendix). However, the demand for computational expertise in databasing and data modelling threaten accessibility of the methodology for the wider research community, and perpetuate the production of manually produced, low-resolution evidence maps. Thus, to sustain interest in developing the methodology to its full potential, Chapter 4 illustrates a proof-of-concept case study using the graph data model for mapping environmental health data.

In this chapter, a scoping review on the use of National Health and Nutrition Examination Survey (NHANES) datasets is expanded using an evidence-mapping approach and graph-based data model. This chapter aims to illustrate the greater return of the graph data model for evidence mapping by comparing this methodology to that of the corresponding scoping

review. It serves to raise the profile of graph-based data modelling in environmental health, and seeks to clarify a direction for future work in this field.

The flexibility of the graph data model, and its ability to maintain the complex relationships connecting datasets, could offer much to evidence mapping at scales akin to the REACH IUCLID database – facilitating high resolution queries and more pro-active/predictive chemical risk assessment.

Chapter 1: Scales, Scoring and Subjectivity: Assessing risk of bias when conducting systematic reviews of the environmental health literature

This chapter was submitted for publication in the journal *Toxicological Sciences* in January 2020. The revision of this manuscript is currently underway*.

The candidate's contribution was structuring own and co-authors' ideas and observations as a manuscript; writing the manuscript for supervisor review; co-ordinating co-author feedback; submission of manuscript.

Candidate: ..... Date: 30/01/2020
Ms. Taylor A. M. Wolffe

Supervisor: ..... Date: 30/01/2020
Prof. Crispin J. Halsall

*In response to reviewers' comments this chapter is being revised to better contextualise the role of systematic review within chemical risk assessment and toxicology – improving access for readers unfamiliar with the methodology.

Conducting systematic reviews in environmental health and toxicology requires that the methodological quality of primary studies included in reviews is assessed in a consistent, robust and transparent manner. However, the considerable variation in design and conduct among primary studies from a diverse range of fields, all of which may be eligible for inclusion in a single environmental health systematic review (Rooney et al., 2016), makes this a significant challenge. Studies of different design are prone to different specific systematic errors (biases) (Rooney et al., 2016). While many of these biases are well described – the specific impacts that these biases have on the overall results of a study (i.e. the direction and relative magnitude of the systematic errors they introduce) are understudied in environmental health and toxicology – and further complicated by study designs in which isolating the effects of a single source of systematic error is difficult (e.g. epidemiological studies) and/or by the empirically inaccessible nature of the "true" result of an effect under investigation in a study. However, within the field of medicine, meta-epidemiological studies have assessed the relative impacts that certain biases have on the results of a study. Several of these biases may be directly applicable to toxicology studies (Rooney et al., 2014). For example, failure to randomly assign study participants to intervention or control arms of a clinical trial can be likened to failure to randomise animals to exposure or control groups of an in-vivo toxicology study. Similarly, environmental epidemiology studies note the opposing directions that certain biases can operate e.g. differential misclassification bias can skew results away from null, whereas the healthy worker effect can skew results towards null (McMichael, 1976; Rothman & Greenland, 1998). More explicit mention of these toxicological and epidemiologically relevant sources of bias are being incorporated into the revision of this manuscript – such that

relatable examples improve the accessibility of the manuscript for environmental health practitioners.

As systematic review is still relatively novel in toxicology and environmental health, there are relatively few examples of tools developed for assessing the risk of bias of primary studies included in environmental health systematic reviews. Those that are available (such as the Navigation Guide (Woodruff et al., 2011; Woodruff & Sutton, 2014) and OHAT's Risk of Bias Rating Tool for Human and Animal Studies (OHAT, 2015) discussed in the current version of the manuscript) will not cover bias domains and signalling questions applicable or suitable to *all* study designs within environmental health. Therefore, conducting systematic reviews on environmental health topics may necessitate the development of new appraisal tools which guide reviewers through the assessment of biases specific to certain fields or study designs e.g. the RoB-SPEO tool (Pega et al., 2020) – which was very recently developed for the assessment of biases specific to studies estimating the prevalence of exposure to occupational risk factors.

Qualitative, domain-based approaches to assessing risk of bias provide a best practice framework with sufficient flexibility for adaptation to specific fields or novel study designs. Ensuring that qualitative, domain-based approaches are adopted in such scenarios – and that quantitative scoring approaches to risk of bias assessment are avoided, is a key aim of this manuscript. As well as highlighting existing tools that adopt the best-practice of qualitatively assessing risk of bias in a domain-based fashion, the revision of this manuscript further draws on examples of numerical appraisal tools which have traditionally been employed, or are currently being employed, in other areas of toxicology and chemical risk assessment e.g. the Klimisch Criteria (a numerical judgement system for assessing reliability, relevance, and adequancy of data to be included in a chemical hazard or risk assessment (Klimisch et al.,

1997)), the Newcastle-Ottawa Scale (a numerical scoring system for assessing the quality of primary nonrandomised studies included in a systematic review, originally developed for medicine but popular in environmental epidemiology (Wells et al., 2014)) and the recently developed numerical assessment tool developed by the EPA for application of systematic review to TSCA risk assessments (EPA, 2018). This is to better evidence the persistent threat of scoring-based practice in the field.

Finally, the illustrative scoring models in the current version of the manuscript (designed to illustrate the fundamental flaws of quantitative approaches to risk of bias assessment) are being further developed in the revision. Additional models which account for scenarios in which numerical scoring systems attempt to weight assessment criteria, and/or account for the direction of bias, are being incorporated into the revision and contrasted against the use of qualitative judgements in terms of the subjectivity required for assignment and interpretation.

Scales, Scoring and Subjectivity: Assessing risk of bias when conducting systematic reviews of the environmental health literature

Taylor A. M. Wolffe[1,2], Crispin Halsall*[1], Paul Whaley[1]

Affiliations:

[1]Lancaster Environment Centre, Lancaster University, Lancaster, LA1 4YQ, UK

[2]Yordas Group, Lancaster Environment Centre, Lancaster University, Lancaster, LA1 4YQ, UK

Correspondence:

*c.halsall@lancaster.ac.uk

Tel: UK +(0)1524 594330

Lancaster Environment Centre, Lancaster University, Lancaster LA1 4YQ, UK.

Key words:

Systematic review, risk of bias

Abstract

Systematic review is gaining popularity in environmental health as a robust and objective means of pursuing more evidence-based approaches to decision-making. A key part of the systematic review workflow is the critical appraisal step, in which the risk that evidence collated from primary studies is biased, is assessed by members of the review team. There is a wide range of tools available to help reviewers conduct this quality appraisal step, with quantitative scales (which produce an overall summary score) being particularly popular. However, published methodological guidance for conducting critical appraisal in environmental health systematic reviews advocate for a qualitative, structured and domain-based approach, eschewing the use of quantitative scales. In this commentary, we explore why this is the case – presenting a theoretical, visual exploration of how quantitative scales and summary scores fail to appropriately represent magnitude of bias.

Introduction

Environmental health (EH) encompasses a diverse range of disciplines producing a significant volume of heterogeneous but highly interwoven data, spanning evidence from human epidemiology studies to *in vitro* experiments. Considering data from all such avenues provides a fuller understanding of the effects that environmental exposures can have on human health. This is vital for evaluating and informing risk-management and regulatory decision-making. However, the growing volume and scope of environmental health data presents a challenge for its translation into regulatory outcomes. This has led to a growing interest in the application of evidence-based approaches to environmental health (e.g. (EPA, 2018; The National Academies of Sciences, 2017; World Health Organization, 2019).

Evidence-based approaches advocate for the identification and use of all relevant, pre-existing evidence for evaluating environmental health risks and mitigation strategies. They seek to increase the precision of risk-management decisions and to reduce the bias associated with analysing cherry-picked and non-representative subsets of an evidence base.

Among evidence-based approaches to evaluating environmental health risks, systematic review (SR) offers an objective, robust and transparent methodological framework for pursuing evidence-based approaches to decision-making, describing an extensive and comprehensive process for synthesising or integrating evidence in answer to a specific research question. Originally developed in the clinical and social sciences (Chalmers et al., 2002; Lau et al., 2013), SR methodology is now being adapted for the context of environmental health (Hoffmann et al., 2017; Whaley et al., 2016). Several examples of detailed methodological guidance for conducting environmental health SRs have been published, including the Navigation Guide (Woodruff & Sutton, 2014), the National Toxicology Program's Office of Health Assessment and Translation (NTP-OHAT) (NTP, 2015), the Texas Commission on Environmental Quality (Schaefer & Myers, 2017), and the SYRINA framework (Vandenberg et al., 2016), among others.

The above SR frameworks detail the formulation of well-focused research questions and the process of devising, documenting and conducting the steps required to answer such questions using existing evidence. Each step can be broadly categorized as belonging to one of three phases of the systematic review process: identifying evidence of potential relevance to addressing the research question; appraising this evidence; and synthesising or integrating evidence using quantitative and/or narrative techniques (Fig. 1).

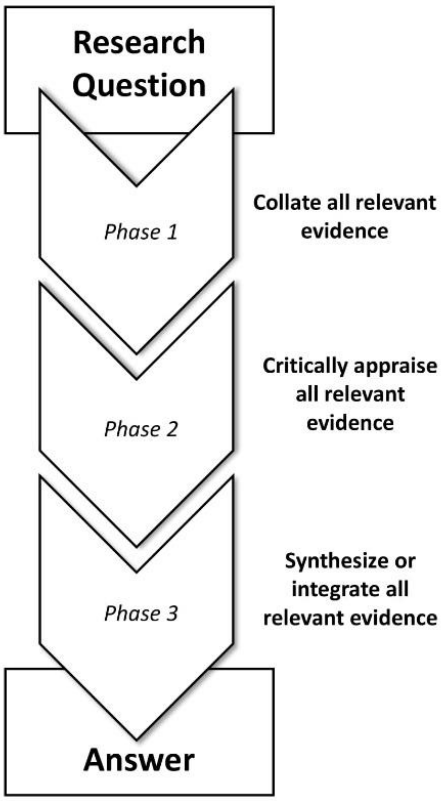

Figure 1: Key phases of the systematic review process.

Fundamental to the SR workflow is the critical appraisal step, in which the strengths and limitations of included studies are characterised. This step is necessary for allowing the final results of a SR to be contextualised in terms of the overall quality of the evidence base. In conducting a SR, it is not only important to synthesise a pooled result through combining multiple studies, but also to determine how trustworthy that pooled result is (Morgan et al., 2016).

Although the importance of the critical appraisal step in systematic reviews is well established (Juni et al., 2001; Lundh & Gøtzsche, 2008; Moja et al., 2005), there appears to be less consensus on *how* this step should be conducted. This is evidenced by the variety of tools designed for this purpose (see reviews by Deeks et al., 2003; Krauth et al., 2013; Samuel et al., 2016), which range from quantitative approaches that promote assessing primary study

quality with summary scores and numerical scales (e.g. Wells et al., 2014) to more qualitative approaches that promote assessing primary study quality on a narrative scale.

The critical appraisal steps of the SR frameworks published by both the Navigation Guide and NTP-OHAT target internal validity, which is assessed via a qualitative, domain-based analysis of risk of bias, and eschews quantitative scoring (NTP, 2015; OHAT, 2015; Woodruff & Sutton, 2014). This follows the recommended approach of Cochrane (Cochrane Community, 2019; Higgins & Green, 2011). In this commentary, we explain why this type of approach should be considered sound practice in EH SRs. We highlight the significance to SR of what these approaches assess (internal validity via risk of bias), how they assess it (using a qualitative scale evaluated on a domain-by-domain basis) and why they assess it in this way (why they do not advocate for a quantitative approach).

Why critical appraisal of studies included in a SR should target internal validity

The concept of "quality" when it comes to research is ambiguous, covering a variety of concepts of differing breadth and subjectivity. Characteristics of a study which are regularly identified in critical appraisal tools as contributing to being of "high quality" include: relevance to solving a research problem (Downs and Black 1998); how comprehensively the methods and results of a study have been reported and how easy the report is to understand (reporting and transparency) (Jadad et al. 1996); how likely the study is to suffer the impacts of random error (precision) (De Vet et al. 1997); how likely it is that the results of a study will be subject to systematic error or bias (Higgins et al., 2011); whether a study is sufficiently sensitive to detect the effect of interest (Cooper et al., 2016); and whether the study design conforms with a recognised international standard (Klimisch et al., 1997), among others.

However, in conducting a SR it is propensity for systematic error which should be the target of critical appraisal at the level of the individual study (Higgins et al., 2011). Systematic errors are reproducible inaccuracies, capable of introducing a consistent bias to the results of a primary study, resulting in either an over- or under-estimate of the true value of the effect under investigation. As systematic reviews are concerned with synthesising the results of primary studies, it is important that these systematic errors are not unwittingly carried forward to the synthesis step of the SR, where they would introduce bias to the overall conclusions of the review. Since the trustworthiness of this summary result is a direct function of the trustworthiness of the results of the individual included studies, it follows that it is the propensity of the design and conduct of each included study to introduce systematic error (i.e. bias) which must be targeted during critical appraisal. The extent to which the methods employed in a study are sufficient to prevent bias is equivalent to the extent to which a study is "internally valid" (Hartling et al., 2009).

Quality constructs other than internal validity will be relevant in critical appraisal contexts outside the systematic review of an environmental health risk. For example, reporting quality is a key construct for assessment during peer-review of scientific manuscripts to ensure transparent, comprehensive and concise reporting of methods and findings to the prospective reader. However, when conducting a systematic review with the objective of elucidating relationships between environmental exposures and subsequent health effects, the construct which matters is the one which directly affects that determination, i.e. potential for systematic error or bias.

Why critical appraisal should not use a quantitative scale and/or summary score to describe risk of bias

Having argued that assessing bias should be the primary focus of a systematic review's critical appraisal step, we now demonstrate why scoring systems should not be used to conduct this task. We do this in three steps: firstly, we show how scores do not reliably correlate with magnitude of systematic error, such that a high score can be consistent with a low degree of bias and vice versa; secondly, we argue that although calibration of a scoring system would address this issue, it is very unlikely to be practically achievable; thirdly, we argue that scales and checklists discourage the deeper level of subjective engagement with the quality of included evidence which SRs need to properly contextualise their results.

1. Scores do not reliably correlate with magnitude of systematic error

In this section, we show how scores do not reliably correlate with magnitude of systematic error, such that a high score can be consistent with a low degree of bias and vice versa. To do this, we present two models of bias, which we refer to as the "Simple Model" and the "Revised Model". The Simple Model exposes false assumptions made by linear scoring models in describing risk of bias. Correcting for these assumptions in the Revised Model then demonstrates how summary scores fail to scale with magnitude and direction of systematic error.

The Simple Model: At their simplest, quality scales operate by awarding individual points to a study for conforming with each item on a list of *n* criteria, with 0 out of *n* being the worst, and *n* out of *n* being the best possible quality scores. Table 1 depicts a simple model scale based on nine criteria, A-I. The specific methodological standards underpinning A-I are arbitrary for the purposes of the model, but can be considered to represent study design features which would safeguard a study from the introduction of bias. These features might involve e.g.

ensuring that comparator or control groups are otherwise treated identically to exposure groups, ensuring that outcome assessors are blinded to the exposure status of participants in epidemiology studies (or animals in toxicology studies), that all potential confounders in epidemiology studies have been identified and appropriately accounted for, etc. (OHAT, 2015; Rooney et al., 2014). One point is awarded for every such criterion a study fulfils, giving a discrete scoring range of 0 to 9 out of a possible 9. Table 1 illustrates how a study might score 4 out of 9 by complying with criteria C, F, G and H. There are multiple further possible ways of obtaining a score of 4/9, so long as a study complies with any four of the nine criteria.

Criterion	**Study complies with criterion?**
A	No
B	No
C	Yes
D	No
E	No
F	Yes
G	Yes
H	Yes
I	No
Total Score	4

Table 1: A simple quality scale comprised of nine quality criteria, (A to I), which each represent a different aspect of quality within a study (e.g. use of controls, blinding, etc.). One point is awarded for every criterion fulfilled, and points are subsequently summed to produce an overall score (e.g. 4 out of 9).

Since a score of 4/9 has the same meaning regardless of the specific criteria met in order to achieve the score, it follows that each criterion from A to I contributed equal weight to the overall quality score. In evaluating the bias of primary studies during a systematic review, simple scales such as this are therefore assuming that the magnitude of bias introduced to the study count equal for every criterion that is *not* met. In other words, the simple model assumes that each unmet criterion introduces the same degree of systematic error. This is represented in Table 2, where failure to comply with any criterion is assumed to introduce 10 units of systematic error. Table 2 shows how three different studies with three different sets of limitations, and therefore different sets of unmet criteria, are scored. In the Simple Model the sum total systematic error is equal for studies with the same score regardless of which unmet criteria introduced the error (Studies A and B, Table 2). As score increases the number of unmet criteria decreases, resulting in a proportional decrease in total units of systematic error (Study C, Table 2). The linear relationship between score and magnitude of error assumed by the Simple Model is shown in Figure 2.

		Study A		Study B		Study C	
Criterion	**Magnitude of bias introduced for non-compliance with criterion**	Study complies with criterion?	Magnitude of bias introduced	Study complies with criterion?	Magnitude of bias introduced	Study complies with criterion?	Magnitude of bias introduced
A	10	✗	10	✓	0	✗	10
B	10	✗	10	✓	0	✓	0
C	10	✓	0	✗	10	✓	0
D	10	✗	10	✓	0	✓	0
E	10	✗	10	✓	0	✓	0
F	10	✓	0	✗	10	✓	0
G	10	✓	0	✗	10	✗	10
H	10	✓	0	✗	10	✗	10
I	10	✗	10	✗	10	✓	0
Total	-	**4**	50	**4**	50	**6**	30

Table 2: The assumption that the magnitude of bias introduced by failing to meet a criterion is equal for all criteria allows a summary score to mean the same thing in all scenarios (e.g. Study A compared to Study B above), and allows the score to accurately scale studies; those with a higher score suffer a smaller magnitude of bias (i.e. Study C compared to Study A or B).

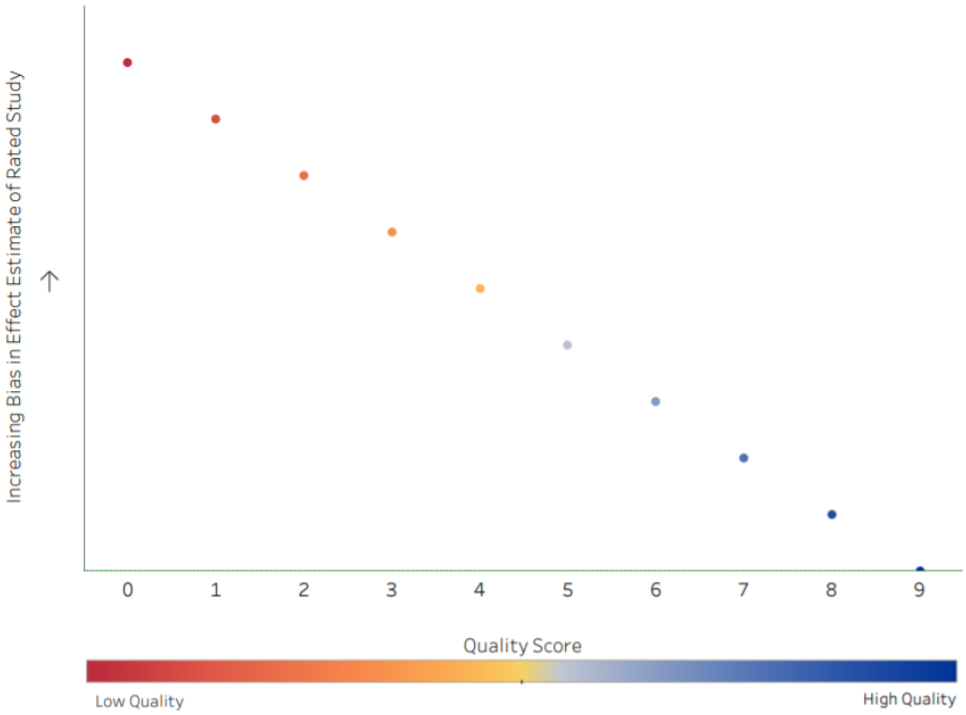

Figure 2: The Simple Model of scoring study quality assumes a linear, proportional relationship between magnitude of error and study score. As study score increases, bias decreases.

The Revised Model: A key problem with the simple model is its underlying assumption that all criteria count equal in terms of magnitude of systematic error is false. There is good empirical evidence that different limitations in study design and conduct introduce different degrees of systematic error in comparable studies, and also that the same limitations will introduce different degrees of systematic error in different research contexts (Cochrane Methods Group, 2017). For example, evidence from preclinical trials of treatments for glioma (brain tumour) show that failure to randomise animals to intervention and control arms introduces a larger bias than failure to blind study personnel (Macleod et al., 2015). The assumption that an equal magnitude of bias is introduced by different study limitations cannot therefore be sustained, and the Simple Model must be relinquished.

A more realistic picture of how scores actually represent bias can be developed by revising the assumption that all criteria count equal in terms of introducing systematic error. To do this, each unmet criterion is now assumed to introduce a different magnitude of systematic error (Table 3). We also relinquish the assumption that biases act in a single direction. For example, in epidemiology it is recognised that while recall bias resulting in differential misclassification can bias the apparent effect of an environmental exposure away from null, the healthy worker effect can bias the apparent effect towards null (e.g. McMichael, 1976; Rothman & Greenland, 1998). This variation in direction of bias is represented in Table 3 by some criteria introducing positive systematic error, while others introduce negative systematic error.

Under this more realistic model of bias, it becomes evident that different ways of achieving the same score will introduce different degrees of systematic error. This is illustrated in Table 3, using the same examples of scoring 4/9 as presented in Table 2. Not only does a score of 4/9 no longer mean the same thing in every context, but the variable and bidirectional nature of bias means that a higher score does not necessarily account for a lower overall sum magnitude of systematic error. This can be seen in the magnitude of bias for a higher score of 6/9 (Study C in Table 3) being greater than a lower score 4/9 (Study A in Table 3). It follows that if quality criteria do not have equal value, then a higher quality score does not correlate with a lower risk of bias – in the Revised Model, a “better” study can be giving a worse result.

		Study A		Study B		Study C	
Criterion	Magnitude and direction of bias introduced from non-compliance with quality criterion	Study complies with criterion?	Magnitude of bias introduced	Study complies with criterion?	Magnitude of bias introduced	Study complies with criterion?	Magnitude of bias introduced
A	+10	✗	+10	✓	0	✗	+10
B	-5	✗	-5	✓	0	✓	0
C	+30	✓	0	✗	+30	✓	0
D	+40	✗	+40	✓	0	✓	0
E	-25	✗	-25	✓	0	✓	0
F	-10	✓	0	✗	-10	✓	0
G	+15	✓	0	✗	+15	✗	+15
H	+10	✓	0	✗	+10	✗	+10
I	-5	✗	-5	✗	-5	✓	0
Total	-	**4**	+15	**4**	+40	**6**	+35

Table 3: Removing the assumption that the magnitude of bias introduced by failing to meet a criterion is equal for all criteria breaks the ability of the same score to represent the same magnitude of bias (e.g. Study A compared to Study B above). Additionally, removing the assumption that bias is unidirectional breaks the ability of a score to scale with quality, as it is no longer true that a higher score necessarily accounts for a lower sum magnitude of bias compared to a lower score (e.g. Study C compared to Study A).

The full range of values for the Revised Model, based on bias values presented in Table 3, is shown in Figure 3. This is calculated on the basis of there being one way of scoring 9/9 or 0/9, nine ways of scoring 8/9 or 1/9, thirty-six ways of scoring 7/9 or 2/9, and so forth. Since scoring follows a probability distribution, there are 512 possible scoring combinations in total. Figure 2 shows how each score (apart from 0/9 or 9/9) is consistent with multiple different sum total introductions of systematic error: a score seemingly indicative of high study quality (e.g. 7/9 or 8/9) can equate to an equal or larger degree of systematic error than scores indicative of lower quality (such as 3/9 or 4/9). The scores are collapsing a wide range of potential for systematic error into a single summary figure, therefore obscuring rather than revealing the effect of methodological shortcomings on the extent to which a study's results are likely to be biased.

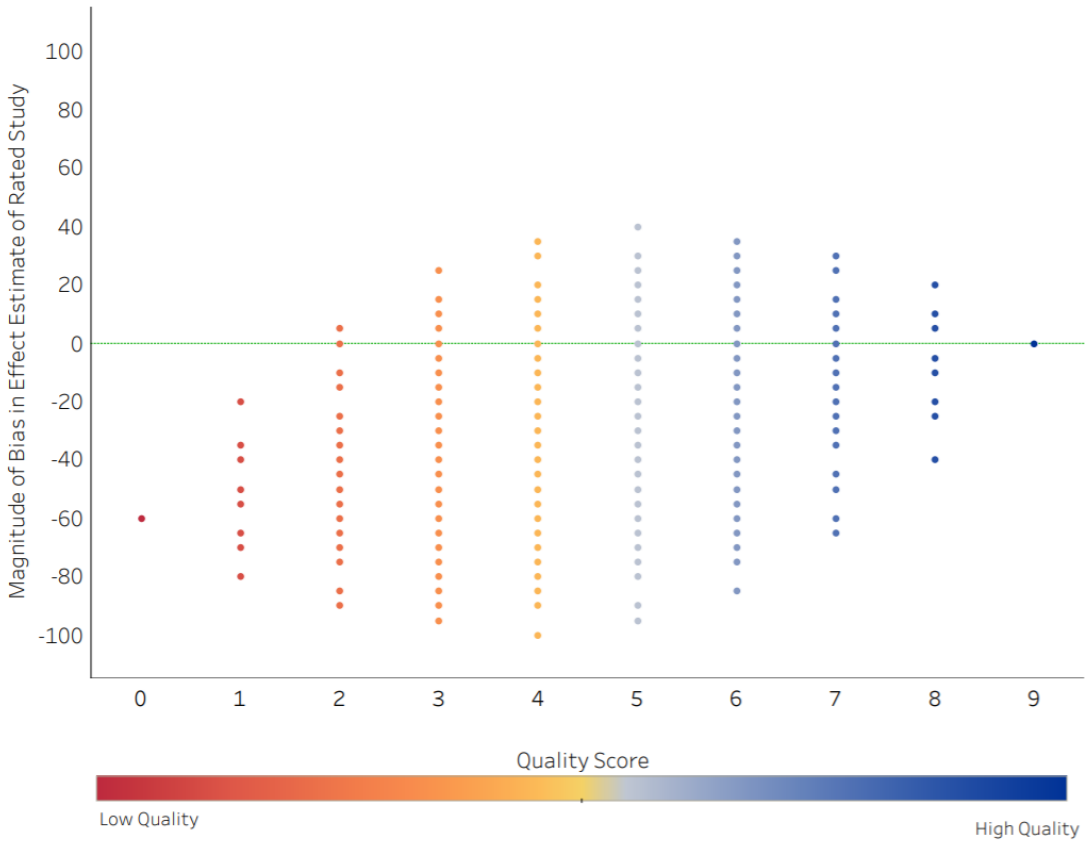

Figure 3: Revising the assumptions underpinning the Simple Model. Expanding on Table 3, the total magnitude of bias associated with every possible way of achieving a score on the

artificially generated scale is displayed, illustrating the range of meaning a single score might serve to mask.

2. Scales cannot readily be calibrated

In our illustrative evaluation of how summary scores cannot realistically represent bias, we have assumed that all the studies used as examples in Tables 2 and 3 are of the same type, designed to assess results for a shared research context, and therefore similar in overall study design. This is not representative of real-world environmental health systematic reviews, which will synthesise or integrate studies of varied design to obtain a summary result. Empirical evidence shows that differences in context have a large impact on how limitations in study design can bias the effect estimate of a primary study (Balk et al., 2002; Berkman et al., 2014), such that magnitude of bias introduced by a particular limitation in design or conduct will vary depending on the type and design of the study in which it is found (Cochrane Methods Group, 2017). For example, evidence from patient blinding techniques in clinical trials has shown that when patients are not blinded to treatment, intervention efficacy is exaggerated but the degree of bias varies according to study design (Hróbjartsson et al., 2014); and in studies in which investigators are not blinded, whether the outcome is objective (such as mortality) or subjective (such as patient reported pain levels) affects the magnitude of bias which failure to blind introduces (Wood et al., 2008).

While scales could in theory weight their scores according to whether e.g. blinding was occurring in the context of a subjective or objective outcome, in practice researchers would end up in a situation where, in order to accurately represent systematic error, each scale would have to be adapted for each individual study design. Even if this did not arguably defeat the purpose of a scale, which is to be readily applied by the user as a measure of quality in

multiple situations, such calibration would require knowledge of exactly how much bias a limitation in design and conduct of a study will introduce to its results.

Acquiring information for calibration either requires a near-perfect version of a study to have been conducted against which the study under appraisal can be compared, or it requires meta-epidemiological research designed to determine the relative effect a methodological limitation has on the findings of a group of otherwise sufficiently comparable studies. Near-perfect studies, such as sufficiently-similar randomised trials against which an observational study can be compared are almost always either unethical, or simply impractical. Even where this gold standard is available, where an exceptional observational study might provide a reasonable benchmark for comparison, there is little guarantee that any given observed effect represents the true effect because the true effect is empirically inaccessible (Groenwold and Rovers, 2010); (Jadad and Enkin, 2008). When it comes to meta-epidemiology, there are rarely enough studies to power precise analysis of the effect of e.g. failure to blind in various research contexts (Giraudeau et al., 2016) - let alone produce enough data to permit a scoring scale to be calibrated for each study design likely to be included in a systematic review .

Scale calibration is therefore beyond the practical reach of most research teams working in most systematic review contexts (Balk et al., 2002; Berkman et al., 2014). This makes the process of appraising studies for bias in the majority of circumstances a qualitative, subjective process.

3. Scores mask the subjectivity of appraising studies for risk of bias

Despite theoretical (Greenland & O'Rourke, 2001) and empirical (Jüni et al., 1999) arguments against numerical approaches to critical appraisal, and over a decade of official guidance arguing against their use (Higgins & Green, 2009), scales and scoring systems continue to be popular (Beronius & Hanberg, 2017; EPA, 2018; Wells et al., 2014; Schneider et al., 2009). In addition to ease of use, the persistence of scales may in part be due to a perceived need to

resist subjectivity when conducting assessments which are supposed to be objective and scientifically robust; such are the values of systematic review. Using scales and checklists may offer an impression of objectivity by producing a fixed numerical summary which appears to involve minimal subjective interpretation.

However, in attempting to limit the subjectivity of interpretation, simply scoring a list of quality constructs does not do enough to elicit further discussion or justification from the reviewer. This compromises the transparency that is fundamental to the systematic review process. Furthermore, because presenting a summary score in lieu of the explicit reasoning behind reviewers' judgements masks the subjective judgements involved in reaching the score, it arguably encourages the sort of inconsistent and subjective interpretation by users of the review that quality scales are seeking to avoid (Shamliyan et al., 2010).

Masking subjectivity in this manner also has the potential to stymie the progression of risk of bias assessment and related tools, especially in the fields of environmental health, toxicology and chemical risk, where systematic review is gaining prominence. It does not encourage understanding and appreciation of factors important for reducing bias in different experimental designs. The attempted rigidity of the scales leaves little room for accommodating innovative study designs, as studies are only expected to rate favourably in a risk of bias assessment if they fulfil the criteria the makers of the scale dictate as valid (Groenwold and Rovers, 2010).

Where scales attempt to simplify and render objective the process of evaluating risk of bias, they instead create a system whereby not only are the scales used to assign points to a study based on a subjective process, but the overall summary score presented to the reader remains open to interpretation (Sanderson et al., 2007). This makes scales a poor choice for thorough, robust and transparent systematic reviews.

Domain-based assessment of risk of bias as a response to the shortcomings of scores and scales

Since magnitude of bias cannot normally be measured, and scores cannot reliably represent magnitude of bias, studies included in a SR have to be appraised for bias on a case-by-case basis. This is best conducted by targeting *risk* of bias, and managing subjectivity through a domain-based risk of bias assessment.

1. Targeting risk of bias

While the precise extent to which any given study is biased cannot be readily quantified, there is good empirical evidence from meta-epidemiological research (e.g. Bolvig et al., 2018; Crossley et al., 2008; Dechartres et al., 2016) that certain methodological features consistently introduce systematic error to the results of a study. These include failure to blind study personnel to experimental and control arms of a study, not controlling for important confounders in observational studies, etc.

It can be assumed that, if a study has methodological features which have been shown elsewhere to introduce bias, then the study is at least at risk of likewise being biased - even if that risk may only be characterised qualitatively (Higgins & Green, 2011). Since most meta-epidemiological evidence for risk of bias comes from healthcare research, it is arguably even more important to characterise risk qualitatively, pending more detailed information about how study design variables can introduce systematic error into environmental health study results. While the precise magnitude of bias cannot be known, it should still be possible to come to a meaningful judgement as to whether the probability and likely direction and magnitude of bias in the study is important enough that it should reduce confidence that the reported results of the study in question are true.

2. Manage subjectivity through domain-based assessment of risk of bias

The process of managing subjectivity can be promoted through the use of domain based approaches to assessing risk of bias, such as those utilised by the Navigation Guide (Woodruff & Sutton, 2011, 2014) and by OHAT's Risk of Bias Rating Tool for Human and Animal Studies (OHAT, 2015), both of which are adapted from the Cochrane Collaboration's Risk of Bias tool for non-randomised studies (Higgins & Green, 2009). As opposed to scales, which result in opaque, quantitative conclusions, domain-based systems consider each specified bias construct individually (O'Connor et al., 2015), eliciting consistent appraisal of key issues via a structured questionnaire format to reach a transparent, qualitative conclusion about risk of bias. Formatted as a framework for the kinds of bias that are likely to impact the effect estimate, reviewers are guided to consider appropriate factors and the relative significance of these factors within the context of individual studies. This promotes distinction between the concepts of "quality" and "risk of bias" so as to account for scenarios in which studies may have been conducted according to the highest quality standards, but may still suffer a significant risk of bias – such as scenarios in which blinding may have been impossible (Armijo-Olivo et al., 2012) – or may be poorly reported yet still at low risk of bias.

Unlike scoring systems, domain-based risk of bias tools acknowledge the need for reviewers' experiences and knowledge when considering which types of bias are likely to be significant given the context of the review. They account for subjectivity by managing it in a transparent manner, prompting reviewers to pre-specify how risk of bias will be handled in their protocol and requesting that reviewers provide justification and evidence for their judgements, allowing readers to decide whether they agree with the results (Higgins & Green, 2011; Rooney et al., 2014; Woodruff & Sutton, 2011) – an approach which provides the necessary balance between qualitative, subjective judgements and transparency about the context in which those judgements are being made. By not defining *a priori* the relative weight or importance of any one specific source of bias (within a domain) compared to another,

approaches following the same principles as the NTP-OHAT and Navigation Guide allow for the context-dependent nature of bias. Arriving at qualitative judgements promotes consideration of a study's limitations and its context as a whole, rather than first considering limitations and only then assessing whether a study has these limitations, as is promoted in numerical scales.

The major advantage of these qualitative, domain-based approaches is their potential to yield consistent appraisal of potential for systematic error, even when the "true" result of a given study is empirically inaccessible: they elicit from experts what they do know about study limitations, to a judgement consistent with what can realistically be inferred from those limitations given limited access to "true" results. To limit the extent to which expert judgement can reduce transparency and be itself a risk of bias in a systematic review, this subjectivity is managed by domain-based tools through their structured approach to critical appraisal.

Conclusion

Systematic review is still relatively novel in the field of environmental health. However, a growing appreciation of the importance of making evidence-based decisions which consider all available data continues to see increasing interest in the application of systematic review to environmental health contexts (Whaley et al., 2016). It is therefore vital to adopt methodology that eschews the use of scales and scoring systems for rating risk of bias at this early stage, to avoid repeating the mistakes and learning the same hard lessons experienced in the medical field.

Where scales have tried to simplify and objectify the evaluation of risk of bias, they have instead simply masked the subjectivity associated with its assessment. Combining this with their inability to represent the true nature of bias makes scales a poor choice for thorough, robust and transparent systematic review. Qualitative domain-based approaches, on the other hand, offer a process for acknowledging and managing subjectivity.

Acknowledging the role of subjectivity, rather than hiding it behind a summary score, is currently the most transparent means of tackling the inherent subjectivity associated with assessing primary study quality. In an area as complex as risk of bias assessment, there may even be significant value in some degree of subjectivity. So far as this is transparently justified, the freedom to be subjective may result in a well-considered appraisal of included studies, particularly where unique or emerging study types are concerned, leaving room for progressive debate, review and future improvement. This is likely to be particularly important in chemical risk assessment, a field ever pressured to adopt new and alternative toxicity testing procedures, the quality of which cannot be appropriately accounted for by rigid scales.

However too much freedom in assessing risk of bias may have the opposite effect, providing insufficient means of focusing the process or holding reviewers accountable for their judgements, creating confusion and inconsistency, and resulting in a backward slide to the well-documented challenges with narrative appraisals of quality of evidence (Chalmers et al., 2002). It would therefore seem that the most suitable means of assessing risk of bias must target a middle ground, neither masking subjectivity nor giving it free reign, but rather "managing" it.

Well-managed and open consideration of the limitations, not only of the primary studies, but of the methods used to assess them, has the potential to increase the reliability of chemical risk assessment conclusions. Assessing risk of bias in a domain-based manner should allow a wider variety of resources to inform risk assessment, increasing the precision of safety estimates, and reducing the research waste and costs associated with repeat testing.

Acknowledgements

TW's PhD is financially supported by the Centre for Global Eco-Innovation (a programme funded by the European Regional Development Fund) and Yordas Group, a global consultancy in the area of chemical safety, regulations and sustainability. PW reports personal fees from

Elsevier Ltd (Environment International), the Cancer Prevention and Education Society, the Evidence Based Toxicology Collaboration and Yordas Group, and grants from Lancaster University, which are outside the submitted work but relate to the development and promotion of systematic review and other evidence-based methods in environmental health research, delivering training around these methods, and providing editorial services.

References

Armijo-Olivo, S., Stiles, C. R., Hagen, N. A., Biondo, P. D., & Cummings, G. G. (2012). Assessment of study quality for systematic reviews: A comparison of the Cochrane Collaboration Risk of Bias Tool and the Effective Public Health Practice Project Quality Assessment Tool: Methodological research. *Journal of Evaluation in Clinical Practice*, *18*(1), 12–18. https://doi.org/10.1111/j.1365-2753.2010.01516.x

Balk, E., Bonis, P., Moskowitz, H., Schmid, C. H., Ioannidis, J. P. A., Wang, C., & Lau, J. (2002). Correlation of QualityMeasures With Estimates of Treatment Effect inMeta-analyses of Randomized Controlled Trials. *JAMA*, *287*(22), 2973–2982.

Berkman, N. D., Lina Santaguida, M. P., Viswanathan, M., & Morton, S. C. (2014). The Empirical Evidence of Bias in Trials Measuring Treatment Differences. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK253181/pdf/Bookshelf_NBK253181.pdf

Beronius, A., & Hanberg, A. (2017). *SciRAP workshop report Bridging the gap between academic research and Workshop report : " Bridging the gap between academic research and*. Retrieved from http://www.diva-portal.org/smash/get/diva2:1074866/FULLTEXT02.pdf

Bolvig, J., Juhl, C. B., Boutron, I., Tugwell, P., Ghogomu, E. A. T., Pardo, J. P., … Christensen, R. (2018). Some Cochrane risk-of-bias items are not important in osteoarthritis trials: a meta-epidemiological study based on Cochrane reviews. *Journal of Clinical*

Epidemiology, *95*, 128–136. https://doi.org/10.1016/j.jclinepi.2017.11.026

Chalmers, I., Hedges, L., & Cooper, H. (2002). A brief history of research synthesis. *Evaluation & the Health …*, *25*(1), 12–37. https://doi.org/10.1177/0163278702025001003

Cochrane Community. (2019). MECIR Manual: Assessing risk of bias in included studies (C52-C60). Retrieved September 20, 2001, from https://community.cochrane.org/mecir-manual/standards-conduct-new-cochrane-intervention-reviews-c1-c75/performing-review-c24-c75/assessing-risk-bias-included-studies-c52-c60

Cochrane Methods Group. (2017). Assessing Risk of Bias in Included Studies. Retrieved November 3, 2017, from http://methods.cochrane.org/bias/assessing-risk-bias-included-studies

Cooper, G. S., Lunn, R. M., Ågerstrand, M., Glenn, B. S., Kraft, A. D., Luke, A. M., & Ratcliffe, J. M. (2016). Study sensitivity: Evaluating the ability to detect effects in systematic reviews of chemical exposures. *Environment International*, *92–93*, 605–610. https://doi.org/10.1016/j.envint.2016.03.017

Crossley, N. A., Sena, E., Goehler, J., Horn, J., Van Der Worp, B., Bath, P. M. W., … Dirnagl, U. (2008). Empirical evidence of bias in the design of experimental stroke studies: A metaepidemiologic approach. *Stroke*, *39*(3), 929–934. https://doi.org/10.1161/STROKEAHA.107.498725

De Vet, H. C. W., De Bie, R. A., Van Der Heijden, G. J. M. G., Verhagen, A. P., Sijpkes, P., & Knipschild, P. G. (1997). Systematic reviews on the basis of methodological criteria. *Physiotherapy*, *83*(6), 284–289. https://doi.org/10.1016/S0031-9406(05)66175-5

Dechartres, A., Trinquart, L., Faber, T., & Ravaud, P. (2016). Empirical evaluation of which trial characteristics are associated with treatment effect estimates. *Journal of Clinical Epidemiology*, *77*, 24–37. https://doi.org/10.1016/j.jclinepi.2016.04.005

Deeks, J. J., Dinnes, J., D'Amico, R., Sowden, A. J., Sakarovitch, C., Song, F., … European Carotid Surgery Trial Collaborative Group. (2003). Evaluating non-randomised intervention studies. *Health Technology Assessment (Winchester, England)*, *7*(27). https://doi.org/96-26-99 [pii]

Downs, S. H., & Black, N. (1998). The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *Journal of Epidemiology and Community Health*, *52*(6), 377–384. https://doi.org/10.1136/jech.52.6.377

EPA. (2018). Application of Systematic Review in Tsca Risk Evaluations, (May), 1–247. Retrieved from https://www.epa.gov/sites/production/files/2018-06/documents/final_application_of_sr_in_tsca_05-31-18.pdf

Giraudeau, B., Higgins, J. P. T., Tavernier, E., & Trinquart, L. (2016). Sample size calculation for meta-epidemiological studies. *Statistics in Medicine*, *35*(2), 239–250. https://doi.org/10.1002/sim.6627

Greenland, S., & O'Rourke, K. (2001). On the bias produced by quality scores in meta-analysis, and a hierarchical view of proposed solutions. *Biostatistics (Oxford, England)*, *2*(4), 463–471. https://doi.org/10.1093/biostatistics/2.4.463

Groenwold, R. H. H., & Rovers, M. M. (2010). The Catch-22 of appraisals on the quality of observational studies. *Journal of Clinical Epidemiology*, *63*(10), 1059–1060. https://doi.org/10.1016/j.jclinepi.2010.04.013

Hartling, L., Ospina, M., Liang, Y., Dryden, D., Hooton, N., Krebs Seida, J., & Klassen, T. (2009). Risk of bias versus quality assessment of randomised controlled trials: cross sectional study. *BMJ (Clinical Research Ed.)*, *339*, 1–6. https://doi.org/10.1136/bmj.b4012

Higgins, J.P.T., Green, S. (editors). (2009). Cochrane Handbook for Systematic Reviews of

Interventions (Version 5.0.2).

Higgins, J.P.T., Altman, D.G., Gøtzsche, P.C., Jüni, P., Moher, D., Oxman, A.D., … JAC, S. (2011). The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. . *British Medical Journal*, *343*, 889–893. https://doi.org/doi: 10.1136/bmj.d5928

Higgins, J.P.T, Green, S. (editors). (2011). Cochrane Handbook for Systematic Reviews of Interventions (Version 5.1.0).

Hoffmann, S., de Vries, R. B. M., Stephens, M. L., Beck, N. B., Dirven, H. A. A. M., Fowle, J. R., … Tsaioun, K. (2017). A primer on systematic reviews in toxicology. *Archives of Toxicology*, *91*(7), 2551–2575. https://doi.org/10.1007/s00204-017-1980-3

Hróbjartsson, A., Emanuelsson, F., Thomsen, A. S. S., Hilden, J., & Brorson, S. (2014). Bias due to lack of patient blinding in clinical trials. A systematic review of trials randomizing patients to blind and nonblind sub-studies. *International Journal of Epidemiology*, *43*(4), 1272–1283. https://doi.org/10.1093/ije/dyu115

Jadad, A. R., & Enkin, M. W. (2008). *Randomized Controlled Trials: Questions, Answers, and Musings, Second Edition*. Retrieved from http://www.blackwellpublishing.com/content/BPL_Images/Content_store/Sample_chapter/9781405132664/9781405132664_4_003.pdf

Jadad, A. R., Moore, R. A., Carroll, D., Jenkinson, C., Reynolds, D. J. M., Gavaghan, D. J., & McQuay, H. J. (1996). Assessing the quality of reports of randomized clinical trials: Is blinding necessary? *Controlled Clinical Trials*, *17*(1), 1–12. https://doi.org/10.1016/0197-2456(95)00134-4

Juni, P., Altman, D. G., & Egger, M. (2001). Systematic reviews in health care: Assessing the quality of controlled clinical trials. *Bmj*, *323*(7303), 42–46. https://doi.org/10.1136/bmj.323.7303.42

Jüni, P., Witschi, A., Bloch, R., & Egger, M. (1999). The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA : The Journal of the American Medical Association*, *282*(11), 1054–1060. https://doi.org/10.1001/jama.282.11.1054

Klimisch, H.-J., Andreae, M., & Tillmann, U. (1997). A Systematic Approach for Evaluating the Quality of Experimental Toxicological and Ecotoxicological Data. *Regulatory Toxicology and Pharmacology*, *25*(1), 1–5. https://doi.org/10.1006/rtph.1996.1076

Krauth, D., Woodruff, T. J., & Bero, L. (2013). Instruments for assessing risk of bias and other methodological criteria of published animal studies: A systematic review. *Environmental Health Perspectives*, *121*(9), 985–992. https://doi.org/10.1289/ehp.1206389

Lau, J., Rothstein, H., & Stewart, G. (2013). *History & progress of meta-analysis (Handbook of Meta-analysis in Ecology and Evolution)*. Princeton University Press. Retrieved from https://books.google.co.uk/books?id=Zh3GNd9M1oUC&printsec=frontcover&source=gbs_ge_summary_r&cad=0

Lundh, A., & Gøtzsche, P. C. (2008). Recommendations by Cochrane Review Groups for assessment of the risk of bias in studies. *BMC Medical Research Methodology*, *8*, 22. https://doi.org/10.1186/1471-2288-8-22

Macleod, M. R., Lawson McLean, A., Kyriakopoulou, A., Serghiou, S., de Wilde, A., Sherratt, N., … Sena, E. S. (2015). Risk of Bias in Reports of In Vivo Research: A Focus for Improvement. *PLoS Biology*, *13*(10), 1–12. https://doi.org/10.1371/journal.pbio.1002273

McMichael, A. J. (1976). Standardized Mortality Ratios and the “Healthy Worker Effect”: Scratching Beneath the Surface. *Journal of Occupational and Environmental Medicine*, *18*(3). Retrieved from

https://journals.lww.com/joem/Fulltext/1976/03000/Standardized_Mortality_Ratios_and_the__Healthy.9.aspx

Moja, L. P., Telaro, E., D'Amico, R., Moschetti, I., Coe, L., & Liberati, A. (2005). Assessment of methodological quality of primary studies by systematic reviews: results of the metaquality cross sectional study. *Bmj*, *330*(7499), 1053. https://doi.org/10.1136/bmj.38414.515938.8F

Morgan, R. L., Thayer, K. A., Bero, L., Bruce, N., Falck-Ytter, Y., Ghersi, D., … Schünemann, H. J. (2016). GRADE: Assessing the quality of evidence in environmental and occupational health. *Environment International*, *92–93*, 611–616. https://doi.org/10.1016/j.envint.2016.01.004

NTP. (2015). Handbook for conducting a literature-based health assessment using OHAT approach for systemic review and evidence integration, 1–98. Retrieved from https://ntp.niehs.nih.gov/ntp/ohat/pubs/handbookjan2015_508.pdf

O'Connor, S. R., Tully, M. A., Ryan, B., Bradley, J. M., Baxter, G. D., & McDonough, S. M. (2015). Failure of a numerical quality assessment scale to identify potential risk of bias in a systematic review: a comparison study. *BMC Research Notes*, *8*, 224. https://doi.org/10.1186/s13104-015-1181-1

OHAT. (2015). OHAT Risk of Bias Rating Tool for Human and Animal Studies, (January), 1–37. Retrieved from https://ntp.niehs.nih.gov/ntp/ohat/pubs/riskofbiastool_508.pdf

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), 711–718. https://doi.org/10.1289/ehp.1307972

Rothman, K. J., & Greenland, S. (1998). *Modern Epidemiology* (2nd Edition). Lippincott-

Raven.

Samuel, G. O., Hoffmann, S., Wright, R. A., Lalu, M. M., Patlewicz, G., Becker, R. A., … Stephens, M. L. (2016). Guidance on assessing the methodological and reporting quality of toxicologically relevant studies: A scoping review. *Environment International*, *92–93*, 630–646. https://doi.org/10.1016/j.envint.2016.03.010

Sanderson, S., Tatt, I. D., & Higgins, J. P. T. (2007a). Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: A systematic review and annotated bibliography. *International Journal of Epidemiology*, *36*(3), 666–676. https://doi.org/10.1093/ije/dym018

Sanderson, S., Tatt, I. D., & Higgins, J. P. T. (2007b). Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: A systematic review and annotated bibliography. *International Journal of Epidemiology*, *36*(3), 666–676. https://doi.org/10.1093/ije/dym018

Schaefer, H. R., & Myers, J. L. (2017). Guidelines for performing systematic reviews in the development of toxicity factors. *Regulatory Toxicology and Pharmacology*, *91*, 124–141. https://doi.org/10.1016/j.yrtph.2017.10.008

Schneider, K., Schwarz, M., Burkholder, I., Kopp-Schneider, A., Edler, L., Kinsner-Ovaskainen, A., … Hoffmann, S. (2009). "ToxRTool", a new tool to assess the reliability of toxicological data. *Toxicology Letters*, *189*(2), 138–144. https://doi.org/10.1016/j.toxlet.2009.05.013

Shamliyan, T., Kane, R. L., & Dickinson, S. (2010). A systematic review of tools used to assess the quality of observational studies that examine incidence or prevalence and risk factors for diseases. *Journal of Clinical Epidemiology*, *63*(10), 1061–1070. https://doi.org/10.1016/j.jclinepi.2010.04.014

The National Academies of Sciences. (2017). *Application of Systematic Review Methods in an Overall Strategy for Evaluating Low-Dose Toxicity from Endocrine Active Chemicals*. Washington, DC: The National Academies Press. https://doi.org/10.17226/24758

Vandenberg, L. N., Ågerstrand, M., Beronius, A., Beausoleil, C., Bergman, Å., Bero, L. A., … Rudén, C. (2016). A proposed framework for the systematic review and integrated assessment (SYRINA) of endocrine disrupting chemicals. *Environmental Health: A Global Access Science Source*, *15*(1), 1–19. https://doi.org/10.1186/s12940-016-0156-6

Whaley, P., Halsall, C., Ågerstrand, M., Aiassa, E., Benford, D., Bilotta, G., … Taylor, D. (2016). Implementing systematic review techniques in chemical risk assessment: Challenges, opportunities and recommendations. *Environment International*, *92–93*, 556–564. https://doi.org/10.1016/j.envint.2015.11.002

Wells, G.A., Shea, B., O'Connell, D., Peterson, J., Welch, V., Losos, M., Tugwell, P. (2014). The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. Retrieved from http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp

Wood, L., Egger, M., Gluud, L. L., Schulz, K. F., Juni, P., Altman, D. G., … Sterne, J. A. C. (2008). Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *Bmj*, *336*(7644), 601–605. https://doi.org/10.1136/bmj.39465.451748.AD

Woodruff, T. J., & Sutton, P. (2011). An evidence-based medicine methodology to bridge the gap between clinical and environmental health sciences. *Health Affairs*, *30*(5), 931–937. https://doi.org/10.1377/hlthaff.2010.1219

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), 1007–1014. https://doi.org/10.1289/ehp.1307175

World Health Organization. (2019). Framework for use of Systematic Review methods in chemical risk assessment – authors meeting. Retrieved from http://who.int/ipcs/events/SRmeeting_US/en/

Chapter 2: Systematic evidence maps as a novel tool to support evidence-based decision-making in chemicals policy and risk management

This chapter was published in the journal *Environment International* in 2019, volume 130.

The candidate's contribution was structuring own and co-authors' ideas as a manuscript; devising the illustrative example; writing the manuscript for supervisor review; co-ordinating co-author feedback; submission of manuscript.

Candidate: ..... Date: 30/01/2020
Ms. Taylor A. M. Wolffe

Supervisor: ..... Date: 30/01/2020
Prof. Crispin J. Halsall

Environment International 130 (2019) 104871

Contents lists available at ScienceDirect

Environment International

journal homepage: www.elsevier.com/locate/envint

Systematic evidence maps as a novel tool to support evidence-based decision-making in chemicals policy and risk management

Taylor A.M. Wolffe[a,b,*], Paul Whaley[a,d], Crispin Halsall[a], Andrew A. Rooney[c], Vickie R. Walker[c]

[a] Lancaster Environment Centre, Lancaster University, Lancaster, UK
[b] Yordas Group, Lancaster Environment Centre, Lancaster University, Lancaster, UK
[c] Division of the National Toxicology Program, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC, USA
[d] Evidence-Based Toxicology Collaboration, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, USA

ARTICLE INFO

Handling Editor: Hanna Boogaard

Keywords:
Systematic review
Evidence mapping

ABSTRACT

Background: While systematic review (SR) methods are gaining traction as a method for providing a reliable summary of existing evidence for health risks posed by exposure to chemical substances, it is becoming clear that their value is restricted to a specific range of risk management scenarios - in particular, those which can be addressed with tightly focused questions and can accommodate the time and resource requirements of a systematic evidence synthesis.

Methods: The concept of a systematic evidence map (SEM) is defined and contrasted to the function and limitations of systematic review (SR) in the context of risk management decision-making. The potential for SEMs to facilitate evidence-based decision-making are explored using a hypothetical example in risk management priority-setting. The potential role of SEMs in reference to broader risk management workflows is characterised.

Results: SEMs are databases of systematically gathered research which characterise broad features of the evidence base. Although not intended to substitute for the evidence synthesis element of systematic reviews, SEMs provide a comprehensive, queryable summary of a large body of policy relevant research. They provide an evidence-based approach to characterising the extent of available evidence and support forward looking predictions or trendspotting in the chemical risk sciences. In particular, SEMs facilitate the identification of related bodies of decision critical chemical risk information which could be further analysed using SR methods, and highlight gaps in the evidence which could be addressed with additional primary studies to reduce uncertainties in decision-making.

Conclusions: SEMs have strong and growing potential as a high value tool in resource efficient use of existing research in chemical risk management. They can be used as a critical precursor to efficient deployment of high quality SR methods for characterising chemical health risks. Furthermore, SEMs have potential, at a large scale, to support the sort of evidence summarisation and surveillance methods which would greatly increase the resource efficiency, transparency and effectiveness of regulatory initiatives such as EU REACH and US TSCA.

1. Introduction

Systematic review is the epitome of the evidence-based approaches that have revolutionized clinical decision-making. The methodology was developed in response to medical practitioners' need to distill clear and reliable conclusions about the efficacy of clinical interventions from an evidence base seemingly full of contradiction, heterogeneity and bias (Chalmers et al., 2002; Garg et al., 2008; Higgins and Green, 2011). This need parallels that of chemicals policy; where conclusions regarding the safety of exposure to a chemical substance must be synthesised from a significantly more disparate evidence base (Whaley et al., 2016). Consequently, interest in the application of systematic review to regulatory decision-making contexts within chemicals policy and wider environmental health is growing. This is evidenced by the increasing number of systematic reviews published in the field (Whaley and Halsall, 2016), the establishment of collaborations and workgroups dedicated to development and dissemination of environmental health systematic review methodology (Morgan et al., 2016; NTP, 2015; Woodruff and Sutton, 2014), and the adoption and use of systematic review by regulatory bodies such as the United States Environmental Protection Agency (US EPA) (EPA, 2018; The National Academies of Sciences, 2017) and World Health Organization (Mandrioli et al., 2018).

* Corresponding author at: Lancaster Environment Centre, Lancaster University, Lancaster, UK.
E-mail address: t.wolffe@lancaster.ac.uk (T.A.M. Wolffe).

https://doi.org/10.1016/j.envint.2019.05.065
Received 28 March 2019; Received in revised form 10 May 2019; Accepted 24 May 2019

Growing interest in systematic review approaches is indicative of the evolutionary journey chemicals regulation follows as it attempts to reconcile past oversights with present day knowledge and mounting future challenges. A number of legacy chemicals released to market under past regulatory workflows persist on the market without risk assessment. Meanwhile, an overwhelming number of new chemicals are presented for assessment each year while awaiting release to market under modern regulatory workflows (European Comission, 2007; Pool and Rusch, 2014). This amounts to increasing strain on regulatory processes, which must operate without a proportionate increase in resource availability. While providing and/or gathering relevant data for new chemicals now forms a vital part of risk assessment, advances in analytical techniques and scientific understanding continue to broaden the scope of this data beyond the realms of traditional *in vivo* toxicity testing. Although vital for compiling a more complete understanding of a chemical's toxicity, the broad scope and increasing availability of such data presents challenges for decision-makers tasked with handling, appraising and interpreting this data for risk assessment. Failure to have a transparent structure for considering all relevant data appropriate to risk assessment (e.g. a stepwise approach for addressing *in vitro* data following evidence from *in vivo* studies or comprehensive assessment of all *in vitro* data) reduces stakeholder confidence and has the potential to bias regulatory decisions. Studies reporting results amenable to the observer bias of independent assessors, or to the vested interests of non-independent assessors, may be cherry picked from the wider evidence base. Even where all relevant studies are considered, the role that scientific judgement plays in the process of appraisal and interpretation of data can lead to conflicting conclusions between different regulatory bodies (Whaley et al., 2016). Transparency in identifying both the evidence and scientific judgement are critical to establishing trust in decision-making.

Systematic review offers a framework for piecing together this varied data in a transparent and resource efficient manner, such that a more complete picture of toxicity can inform regulatory decision-making. It details methodology for ensuring *all* such data is identified, gathered and considered – preventing cherry picking of studies that only provide part of the complete toxicity profile for a chemical, or that present biased or unrepresentative results. As well as reducing bias, all steps of the methodology are designed to maximise transparency. A well conducted and reported systematic review effectively outlines the research question, the approach taken to address the question, the evidence considered, and the scientific judgement applied to reaching conclusions. Thus, differences across reviews or regulatory bodies can be effectively identified and explained. Considering the results of all relevant studies makes maximum use of existing data and increases the precision of a systematic review's conclusions. This allows reliable decisions to be made without the commissioning of redundant and repetitive primary research, or conversely identifies specific knowledge gaps at which smart testing strategies can be focused.

Although the aim of systematic review (i.e. to transparently and robustly synthesise all available data in answer to a research question) aligns well with the needs of chemicals policy, conflicts between the practicalities associated with the methodology and those associated with regulatory frameworks hinder their wider uptake, and/or the production of reviews that are of sufficient quality to produce trustworthy results (Kelly et al., 2016; Marshall et al., 2018; Reynen et al., 2018). Key areas of conflict include the time and resource intensity of the systematic review process, the scope of the research questions addressed by the methodology, and the ease with which the output of a systematic review can be accessed, interpreted and updated. Further, the fluid and rapidly expanding nature of scientific research and the chemicals industry creates a constant and pressing need for evidence surveillance, such that regulators can keep apace of the growing body of scientific literature and update regulation accordingly. This challenge demands a responsive and living solution beyond the reach of current systematic review practice.

In this manuscript, we briefly outline systematic review methodology to illustrate its strengths and highlight the transferable barriers which have been suggested as preventing its wider uptake in other fields (Oliver and Dickson, 2016). We discuss how these difficulties may be addressed through the novel implementation of systematic evidence mapping in environmental health. Systematic evidence maps (SEMs) provide a broad and comprehensive overview of an evidence base (Haddaway, Bernes, Jonsson, & Hedlund, 2016; James et al., 2016). They facilitate the identification of trends which can be used to inform more efficient systematic review, or more targeted primary research. The methodology behind SEMs, and how this might be adapted to suit the demands and limitations of regulatory decision-making in chemicals policy is discussed, along with the advantages and future potential of SEMs as a fundamental tool for evidence-informed risk management and decision-making.

2. The application of systematic review methods in chemical risk management

The utility and advantages of systematic review methods for advancing chemical risk assessment have been extensively documented elsewhere (Aiassa et al., 2015; Hoffmann et al., 2017; Hooijmans et al., 2012; Rooney et al., 2014; Vandenberg et al., 2016; Whaley et al., 2016; Woodruff and Sutton, 2014). Systematic review provides a transparent and reproducible approach to summarising and critically assessing existing evidence on potential health risks associated with exposure to a chemical substance. These transparent methods serve to document the basis of scientific judgments, minimising the potential for bias and error presented by more traditional narrative approaches in which opinion is not clearly distinguished from evidence.

The key features of a systematic review (Table 1) are:

- a clearly specified research objective - usually captured in a Population-Exposure-Comparator-Outcome (PECO) statement
- a comprehensive search strategy
- screening of the search results - for evidence relevant to addressing the research objective
- extraction of data from included studies - using a prespecified data extraction framework
- critical appraisal of included studies - according to a prespecified set of quality criteria, usually targeting risk of bias
- synthesis of findings from the included studies - using suitable quantitative statistical methods and otherwise qualitative methods as appropriate
- characterisation of confidence in the evidence for the results of the synthesis - according to a prespecified set of criteria
- statement of conclusions - including an assessment of limitations in design and conduct of the review itself.

Specific methodological decisions concerning each of these key features, from definition of the PECO statement to the chosen synthesis approach, are specified in a pre-published protocol.

However, with the methodology's pursuit of rigor and comprehensiveness comes a significant demand for time and resources. Evidence from medical systematic reviews indicates it takes on average approximately 70 weeks to progress a systematic review from protocol registration in the PROSPERO registry (National Institute for Health Research, 2018) to publication of the final systematic review (Borah et al., 2017). Variance around this average is wide (from 6 to 186 weeks), but the significance of person-hours and planning time prior to protocol registration is not considered in these estimates. More recent analysis of environmental science systematic reviews estimates an average of 164 (full time equivalent) person-days required for completion of systematic reviews (Haddaway and Westgate, 2018). However, in the absence of comparable evidence in the field of chemical risk assessment, these figures agree with anecdotal reports of the

Table 1
The key features of systematic reviews and their primary advantages. PECO = Population-Exposure-Comparator-Outcome.

Systematic review step	Primary advantages
Pre-published protocol	Reduces risk that expectation bias will influence reviewers' choice of methods and approaches for analysis mid-review; if formally published, external peer review can reduce risk of limitations in planned methods from compromising final results.
Statement of objectives	Provides a structured framework for the aims of the review (including specific statement of the research question and PECO criteria) against which appropriate review methods can be defined.
Comprehensive search	Reduces risk of only partial retrieval of the overall body of evidence that is relevant to answering the research question.
Screening against eligibility criteria (study inclusion)	Reduces risk of only partial retrieval of the overall body of evidence that is relevant to answering the research question, in particular the risk of selection bias when reviewers are deciding which evidence to include in the review.
Data extraction using appropriate extraction tools	Reduces risk of inconsistent or partial retrieval of data from studies included in the review, reducing risk of selective use of data from studies deemed relevant to answering the research question.
Critical appraisal of included studies	Encourages consistent assessment of validity of included studies according to factors internal to study design, reducing risk of expectation bias or other factors causing studies to be inappropriately weighted, and helping ensure that bias in the findings of the included studies is not transmitted through to the findings of the review.
Synthesis of included studies	Pooling or integration of sufficiently comparable studies increases the power of an analysis, whether quantitative or qualitative, allowing overall trends in results to be more reliably identified.
Characterisation of confidence in the evidence	Encourages consistent assessment of the validity of the results of the synthesis according to features which manifest at the level of body of evidence as a whole rather than the individual study. Outlining the scientific judgement applied in rating confidence is key to the transparency of subsequent conclusions.
Drawing conclusions/key review output	Qualitative and/or quantitative summary effect estimates help direct policy decisions based on permissible exposure levels and related controls; assessment of limitations in the review methods helps ensure that any residual potential biases in the review are made clear to the reader and can additionally be accounted for in uncertainty assessment and consequent risk management action.

average systematic review taking around 12 to 18 months to progress from inception to publication. A significant factor which contributes to the length of the systematic review process is the manual way in which each step of the methodology is conducted. All studies returned by a systematic search strategy are generally screened by human reviewers, in duplicate, one-by-one, before included studies undergo a similarly manual data extraction and critical appraisal step.

Systematic review management software has been developed (e.g. "HAWC: Health Assessment Workplace Collaborative,", 2013; Covidence, 2019; Evidence Partners, 2019; Science for Nature and People Partnership Evidence-Based Conservation working group, Conservation International, Datakind, 2018; Sciome, 2018; Thomas et al., 2010; CAMARADES-NC3Rs, 2019) to assist human reviewers with maintaining transparency in SRs and with organising the review process. Acknowledging the impedance caused by a review's manual workload, review management software is beginning to incorporate machine learning as a means of automating labour-intensive tasks (e.g. Evidence Partners, 2019; Science for Nature and People Partnership Evidence-Based Conservation working group, Conservation International, Datakind, 2018; Sciome, 2018; CAMARADES-NC3Rs, 2019). Automation has the potential to result in significantly reduced workloads and subsequent demands for time and resources (Mara-eves et al., 2015). Pending further advances, the time and resource demands of systematic review are at conflict with the intense time/resource pressure under which regulatory processes must operate (Innvaer et al., 2002; Oliver and Dickson, 2016).

Also at conflict with the demands of regulatory decision-making is the narrow scope of systematic reviews, which are designed to address a specific and clearly defined objective or research question. To ensure a manageable, relevant and focused review, suitable research questions are typically closed framed, such that the review can synthesise a single, coherent answer. These closed-framed questions are well suited to the decision-making contexts of medicine (the field from which systematic reviews originate), but may be difficult to apply to chemical risk assessment. The web of interlinked endpoints, potential variation in sensitive populations, uncharacterised low dose effects, and unknown behaviour of a chemical in the environment or in contact with other chemicals can mean that the decision-critical information which can be supplied by a tightly focused research question is often not readily apparent in chemical risk assessment contexts. Even where such a question can be devised, and the answer reached through systematic review, the specificity of the research problem and its resolution are likely to comprise only part of the much broader range of unaddressed decisions and information requirements faced by risk managers.

3. Systematic evidence maps for chemical risk management

In light of the time and resource intensity of current systematic review practice, identifying the most informative research questions is important for maximising the value and efficiency of systematic reviews in regulatory decision-making. Investing resources in systematic review as a means of addressing specific research questions is inefficient if there is a lack of data available for answering those questions. Devising specific research questions therefore becomes a reactive process, rather than a proactive one. This is at odds with the goals of chemicals policy, which aims to predict and prevent harm as a result of exposure to chemical substances.

Decision-makers therefore need to monitor and understand the evidence base as a whole – such that emerging trends or issues of potential concern can be identified and investigated in a timely manner. Identifying trends in the evidence base, including evidence clusters and evidence gaps, facilitates the formulation of proactive research questions by relevant stakeholders. Reviewers need not rely on environmental health outcomes becoming infamous or epidemic as an indicator of sufficient evidence for an efficient and valuable synthesis. Instead, trends in the availability of evidence ensure prevention of synthesis attempts for which there is insufficient data (or for which syntheses already exist) and promote the targeting of primary research efforts at evidence gaps. This kind of evidence surveillance has traditionally been the domain of scoping reviews. These reviews are often narrowly focused precursors to systematic reviews. Thus a specific systematic review question has already begun to be framed, and the literature scoped for sufficient data to address/focus it – rather than vice versa (e.g. Bolden et al., 2017). Scoping reviews also typically present their findings in tabular format. This compromises the accessibility of the evidence they scope, and makes them ill-suited for applications beyond determining whether there is sufficient literature to merit a systematic review (Grant and Booth, 2009).

Instead, the introduction of systematic evidence mapping, a methodology recently adapted from the social sciences (Clapton et al., 2009) for environmental management (James et al., 2016), has the potential to facilitate evidence surveillance in a transparent and reproducible manner, providing a broader understanding of the extant evidence base through interactive outputs.

Table 2
A comparison of systematic review and systematic evidence mapping methodology and their respective roles in risk management decision-making (adapted from James et al., 2016). SR = systematic review, SEM = systematic evidence map, RM = risk management, TDI = tolerable daily intake.

Step	Conduct of step in SRs related to assessing chemical health risks	Conduct of step in SEMs related to assessing chemical health risks	SR vs SEM for responding to risk management needs
Pre-published protocol	Define all methods in advance of conduct of review	Same	Provides transparency; reduces bias; opportunity for peer review and stakeholder engagement. Applies to both SRs and SEMs.
Statement of objectives	Question concerns the effect of an exposure on health; or the effect of intervening to reduce exposure in terms of health benefit. Usually targets a single or few exposures and outcomes.	Question concerns the state of the evidence base for a topic. Usually open-ended and encompassing a range of multiple related exposures and outcomes.	SR: Focused, closed questions of SRs best service specific RM decisions such as characterising specific health risks/TDIs. SEM: Open questions of SEMs best service scenarios in which evidence should be surveyed and scoped, such as problem identification and priority-setting.
Comprehensive search	Search terms highly resolved and specified for most key elements of the objective statement, returning a moderate volume of evidence.	Wide ranging search strings of lower specificity based on topic rather than defining all key elements of the objective in the search.	SR: Narrow searches efficiently identify evidence related to exposure-outcome pairs. Maximum feasible number of sources searched to ensure collation of all relevant evidence for synthesis. SEM: Broader, topic-based SEM search allows evidence supportive of multiple decision scenarios to be identified. Flexible number of sources searched, or sources searched in a step-wise manner as appropriate to broader research objectives.
Screening against eligibility criteria (study inclusion)	Inclusion criteria specified in detail for all key elements of the objective.	Inclusion criteria defined in terms of topic rather than key elements of the objective.	SR: As for search, specific inclusion criteria ensure SRs efficiently service a specific research question. SEM: Broad objectives ensure inclusion of evidence relating to multiple decision scenarios.
Data extraction using tested extraction sheets	Complete extraction of meta-data and study findings.	Extraction of meta-data; optional extraction of study findings and other study characteristics depending on SEM objectives.	SR: Data extraction determined by objectives. SEM: Data extraction more flexible and can respond to needs of risk management process to develop fit-for-purpose maps of varying degrees of comprehensiveness.
Coding of extracted data using controlled vocabularies	Coding facilitates grouping of included studies for synthesis/integration according to review objectives. Coding is closely related to review objectives and data extraction process, whereby narrow research question and PECO statement inherently define specific code applicable to raw extracted data.	Coding facilitates broad comparison of heterogeneous data across an evidence base. Broad map objectives necessitate extensive coding process, whereby specific code must be defined in a step distinct from the formulation of end-users' specific research questions.	SR: Tight review objectives pre-specify applied code (e.g. considering ages 0–18 as 'Child' for reviews focusing on a population of 'Children'). Narrower range, or greater specificity of controlled vocabulary terms applicable per item of extracted data. SEM: Code pre-specified where possible, but addition of new terms (which could not be accounted for *a priori*) considered flexible. Any one item of extracted data may be coded by multiple and variably resolved terms. Openly accessible ontologies may be used for coding to promote consistency and interoperability.
Critical appraisal of included studies	Assessment of internal validity (risk of bias) conducted for all included studies.	Study validity assessment is optional and to some extent restricted if outcome is not a defined aspect of the SEM; study characteristics relevant to risk of bias assessment can be extracted.	SR: Describe the internal validity of the evidence base, which is an essential step of characterising confidence in the evidence. SEM: Flexible, critical appraisal step can be omitted; study methods are mapped or methodological quality assessed to goals, can be part of stepwise approach where quality only assessed for studies addressing key outcomes etc.
Synthesis of included studies	Quantitative synthesis where possible to produce characterisation of hazard from exposure; qualitative synthesis where pooling studies is not possible.	Reports of systematic maps can provide narrative synthesis of characteristics of the evidence key to a given decision-making context.	SR: Synthesis supports a specific type of decision context. SEM: Primary output is a more context-agnostic database which can be used by risk managers to support multiple decisions in the RM workflow; or to aid in a stepwise approach.
Characterisation of confidence in the evidence	Assessment of confidence or certainty in the results of the synthesis, according to characteristics of the evidence base taken as a whole.	SEMs do not synthesise included studies. SEMs help identify regions of evidence with characteristics indicative of being worth further, detailed analysis in support of a prospective decision.	SR: Provide detailed conclusions on certainty of evidence in hazard characterisation or to support risk assessments. SEM: Support a range of decisions, particularly decisions to focus research and review, e.g. indicating clusters where evidence may be strong enough to warrant SR (e.g. have a reasonable likelihood of changing a TDI), fill in gaps to reduce uncertainty and for surveillance.
Drawing conclusions/key review outputs	SRs primarily provide a summary effect estimate and surrounding uncertainty based on strength of the evidence and review methods.	SEMs primarily provide a searchable database of the characteristics of the evidence base, making the knowledge base locked away in manuscripts accessible to decision-makers.	SR: provide a qualitative and/or quantitative summary effect estimate in answer to a narrow and specific decision-making question. SEM: identify evidence gluts for synthesis. When combined with an understanding of RM needs, transparent criteria for prioritization of gluts for synthesis and gaps for commissioning primary research can be presented.

The methodological steps involved in constructing a systematic evidence map are similar to those involved in the initial stages of producing a systematic review (see Table 2, adapted from James et al., 2016) whereby a systematic search strategy is employed to collate evidence, which is subsequently screened for relevance before undergoing data extraction. The key difference between the methodologies comes in the form of their aims and subsequent outputs. Systematic reviews collate a relatively narrow subset of the evidence base to answer a specific research question. Conversely, SEMs do not attempt to answer a specific, closed-framed research question, and are instead guided by much broader research objectives. SEMs collate a sufficiently broad subset of evidence such that many different specific research questions might be formulated from, and addressed with, a single systematic evidence map. SEMs are concerned with characterising the evidence base within a given research area, such that the availability, type and features of the evidence can be clearly mapped and explored through data visualization.

To facilitate this exploration, the output of a SEM takes the form of a queryable database (Clapton et al., 2009; James et al., 2016) as opposed to the lengthy and technical documents which form the main output of a systematic review. The database format allows users to query the evidence base according to their research interests, providing functionality which is void from systematic review documents and their associated static data tables. This format addresses the inability of systematic evidence mappers to predict what the specific research interests of users might be by providing the option to search for, and select, the specific subsets of data relevant to a particular use case.

Whereas systematic reviews present users with select information from included studies (i.e. data relevant to addressing the research question), SEMs aim to extract a broader range of data from included studies and aim to maintain the native format of these data. In this sense, the search and screening process are the steps of SEM methodology most affected by its research objective or context, as the focus of data extraction remains broad regardless. This is in contrast to systematic review, where all steps are heavily influenced by its research question. The data extracted for inclusion in a SEM database can then be flexibly categorised, or "coded" to facilitate comparison of an otherwise heterogeneous evidence base.

Resolution of coding can be adapted to suit the needs of regulators. For example, coding the species under investigation in a study might use categories such as "Sprague-Dawley", "Rat", "Rodent" or "Mammal"; or may use all of these categories such that the data can be interrogated in successively deeper levels of detail. As well as facilitating variably resolved interrogation of the evidence base, coding plays a significant role in systematic mapping's amenability to updating. Use of universal, standardised ontologies for coding, such as the Unified Medical Language System (UMLS) (U.S. National Library of Medicine, 2016), offers a degree of consistency that future users can readily exploit when updating a map (Baker et al., 2018). These ontologies also offer interoperability between SEMs, creating the potential to expand and merge evidence maps – a feature likely to become increasingly attractive as the scope of evidence relevant to assessing toxicity grows along with our understanding of its interconnectedness.

In current practice it is common to present users with SEMs that house only coded information for simplicity and ease of access (e.g. Papathanasopoulou et al., 2016). However, this conflates data extraction with coding. Maintaining the native format of extracted data and applying coding on top of this therefore ensures maximum transparency in SEMs. This additionally promotes the ease with which a map can be updated as advancing scientific understanding calls for coding categories to be redefined. As with systematic reviews, the data extraction and coding steps of a SEM represent a manual workload. Presenting *only* coded data may offer a saving in the resource intensity of the process. However, in maintaining a transparent link between raw extracted data and the code used to categorise it, SEMs offer a gateway to automation – whereby controlled vocabulary ontologies can be used to train machine learning algorithms to automatically identify, extract and code data from the literature.

Pending such advances, the time required to conduct a fit for purpose systematic map in environmental health is uncharacterised. Evidence from the wider environmental sciences (Haddaway and Westgate, 2018) suggests that (on average) systematic maps take longer to complete than systematic reviews. This is due to the generally larger number of studies they manually collate, screen and extract data from. While maps might present a larger upfront cost in terms of time, their multipurpose nature has the potential to offer more long-term resource savings compared to exclusively conducting systematic reviews. This is because a single systematic evidence map may continue to be useful to several different aspects of the regulatory workflow (see Sections 4 and 5 below).

As the purpose of a SEM is to characterise the evidence base, there is no risk of allocating resources to the production of an inconclusive output, as is the case for "empty" systematic reviews (systematic reviews which ask research questions for which there is too little included evidence for them to reach a conclusion or be supportive of a decision). In fact, systematic evidence maps may reduce the resource strain associated with systematic reviews. A SEM's broad overview of the evidence base allows fast identification of topics for which there is sufficient data to warrant a full systematic review. The SEM itself, if conducted to sufficiently rigorous standards, can even replace the literature search and screening process of a systematic review. As SEMs present all available relevant evidence on a broader topic such as the "health effects of bisphenol-A" (obtained through a systematic but less specific search strategy), filtering this information according to the PECO statement of a systematic review may act in an equivalent manner to approaching the literature with a more focused search strategy in the first instance. The pre-screened nature of this subset is likely to reduce the number of false positive results, facilitating faster syntheses.

As advances in machine learning facilitate more highly resolved data extraction processes, future SEMs may even store enough detail for them to form the basis of meta-analytical syntheses. If all data contained within study reports is extracted and indexed within a SEM, there would be no data required specifically for syntheses which could not be found in the SEM. This would allow SEMs to form the dataset on which meta-analytical and predictive toxicological models are based, the results of which may additionally be incorporated into the SEM itself – facilitating more transparent, resource-efficient and easily updated syntheses.

4. Exploring the evidence base with SEMs

Systematic evidence mapping facilitates identification of trends which are informative for many risk management scenarios. To illustrate the flexibility and potential utility of SEMs' trendspotting capacity, this section highlights the type of data visualization and exploration possible through querying subsets of information in a SEM database. Specifically, "priority setting" (National Academy of Sciences, 1983; Pool and Rusch, 2014), the process by which regulators identify the most pressing chemical substances for assessment and regulation (e.g. from a pool of unassessed legacy chemicals) is presented as context for the exploration of a hypothetical SEM.

Several factors are relevant to prioritizing individual chemicals for assessment, broadly ranging from recorded levels of exposure to evidence for toxicity. Underlying these broad considerations are several more specific factors such as the bio-accessibility of the chemical, the relevance of its toxicity evidence for predicting health risks in human populations etc. In order to make the most efficient use of resources and the systematic review process, decision-makers require access to a means of comparing these features to justify prioritization of a particular chemical for review/risk assessment.

This is the role of a SEM, which may be constructed with the aim of

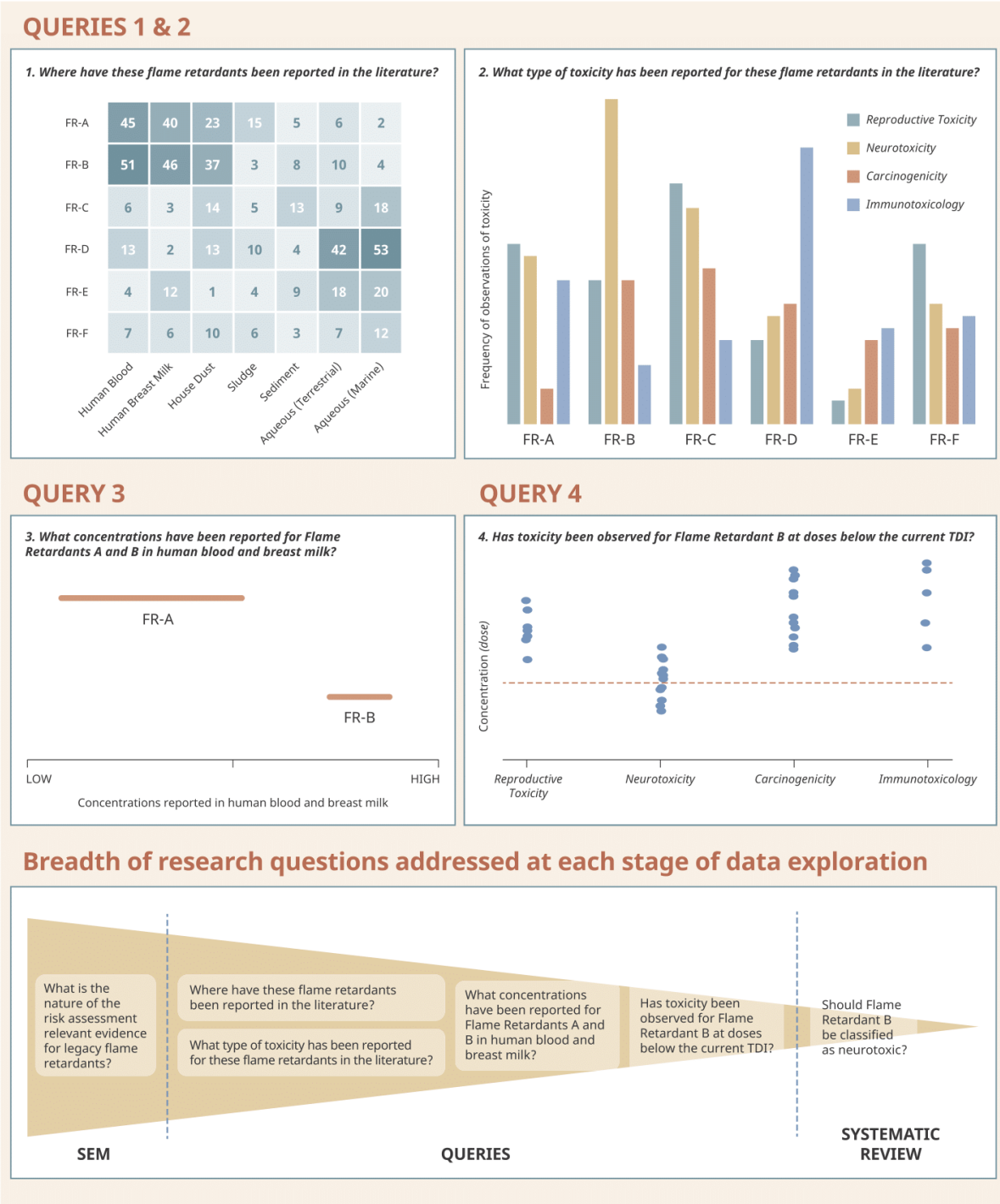

Fig. 1. The process of identifying trends and exploring the evidence landscape involves querying the SEM database and visualizing the results of the query. Queries may start by asking broader questions which consider a wider range and volume of data (e.g. Queries 1 and 2). Users may then further explore any trends of interest discovered in the results of these broad queries by running narrower queries which consider a more specific subset of data (e.g. Queries 3 and 4). Data displayed in this Figure have been artificially generated to illustrate a hypothetical use case for SEMs. FR = flame retardant, TDI = tolerable daily intake, SEM = systematic evidence map.

identifying and characterising the risk assessment relevant evidence for a broader group of legacy chemicals, e.g. flame retardants. Once data has been extracted and coded from the literature, the SEM can be explored with a succession of queries of increasingly narrow focus, each considering a narrower subset of the evidence base than the last, such that a research question appropriate for more detailed synthesis is resolved at the end of a process which begins with a very broad research objective. This is illustrated in Fig. 1 using the hypothetical context of priority setting with a group of arbitrary chemicals, in this case flame retardants (FRs) A–F.

Queries 1 and 2 depicted in Fig. 1 explore the frequency with which the literature observes a flame retardant in a coded location category (e.g. human blood, human breast milk, house dust, etc.) and the frequency with which the literature observes an association between a flame retardant and a coded toxicity category (e.g. reproductive toxicity, neurotoxicity etc.). The heatmap visualizing the results of Query 1 shows a comparatively large number of observations of FRs A and B in location categories directly relevant to human populations (i.e. human blood and breast milk). Query 2 clarifies whether these observations require further attention by indicating what kind of toxicity information is available for each flame retardant. The bar chart visualization indicates comparable numbers of observations for most of the flame retardants and types of toxicity but a comparatively large number of observations that associate FR B with neurotoxicity.

Based on (hypothetical) existing evidence, Queries 1 and 2 indicate flame retardants A and B as potential candidates for full assessment. Resolving which to prioritize involves accessing more study-specific information through a series of queries which consider a successively narrow subset of the evidence base. Despite availability of toxicity data, observing flame retardants in human relevant locations might not be concerning if the concentrations observed are negligible. Thus Query 3 examines the range of concentrations reported in the literature for FRs A and B in human blood and breast milk. Visualization of Query 3 indicates a wider range of lower concentrations reported for FR A, compared to a narrower range of higher concentrations for FR B. Query 4 then examines the relevance of these concentrations against the current estimated tolerable daily intake (TDI) for FR B, indicating several observations of toxicity below the current TDI and supporting prioritization of FR B for assessment. Further, the relatively large volume of observations of neurotoxicity may indicate sufficient data available to conduct a systematic review on FR B's relationship with neurotoxicity.

However, it is important to distinguish the results of SEM queries from synthesis. SEMs only present what has been studied in the literature – they cannot present what has not been studied, and do not always assess the risk of bias of the findings they report. Thus, while a high number of observations of flame retardants A and B in human relevant locations is a valid trend to explore further, it does not necessarily mean that there are fewer of the other flame retardants present in human relevant locations, but rather that there may simply be fewer of these flame retardants studied at all. Identification of such evidence gaps is equally valid for focusing primary research. For example, the relatively high number of observations of reproductive toxicity for FR F, but comparatively low number of observations of this flame retardant in any exposure locations might warrant re-analysis of samples or new exposure studies to verify whether exposure to this substance is of concern.

The SEM is also sufficiently flexible that different trends can be investigated, and different research questions formulated, based on the priorities of regulators. For example, the number of observations in the literature which found FR D in aquatic environments might spur further investigation into the ecotoxicity of this compound. A single SEM exercise therefore makes efficient use of resources in its potential to meet the varied needs of several end users.

5. The role of SEMs in wider risk management workflows

In addition to priority setting, SEMs have the potential to fill several roles within wider workflows.

5.1. Data gathering

Although evidence synthesis methodology can be considered costly in terms of time and resources, this cost can be dwarfed by the equivalent resource demands associated with conducting primary research relevant to assessing the hazards associated with exposure to a chemical, as illustrated with more established examples in the field of medicine (Glasziou et al., 2006). In an effort to manage these demands, reduce the production of research waste, and comply with principles such as the three Rs (European Chemicals Agency, 2018a, 2018b; National Centre for the Replacement Refinement and Reduction of Animals in Research, 2018), a key first step in many regulatory workflows is the identification and gathering of all pre-existing evidence relevant to a specific risk management decision. This can be illustrated in regulatory frameworks such as the European Union's REACH (Registration, Evaluation, Authorisation and Restriction of Chemicals) initiative, which requires registrants to make an attempt to identify all available, pre-existing evidence on the hazards associated with the chemical substance under registration (European Chemicals Agency, 2018a, 2018b). Similarly, REACH imposes a “one substance, one registration” policy, whereby all parties with an interest in registration of a substance must share data, minimising repeat testing. Although promoted in guidance documents (European Chemicals Agency, 2016), a lack of a sufficiently robust methodology for finding, collating, housing and reporting these data leads to poor transparency, and therefore does not remove the potential for cherry picking of key studies which may not be representative of the evidence base as a whole.

SEMs have the potential to provide this much needed transparency. The nature of a SEM's output being a collection of relevant search results, and specific information coded from those results, introduces a greater level of accountability for registrants. Studies are identified by registrants as “key”, “supporting” etc. based on the perceived relevance, adequacy and reliability of the evidence they provide for a specific endpoint, assessed using “sound scientific judgement” (European Chemicals Agency, 2011). These assignments are aided by application of the Klimisch criteria (Klimisch et al., 1997) – a rating methodology criticised for its lack of transparency and failure to consider non-industry sources of evidence (Ingre-Khans et al., 2019). This poor transparency hinders the appraisal of registrants' choices (e.g. of key study), and the degree to which those choices can be considered representative of the wider evidence base. Using SEM methodology alleviates this issue by requiring registrants to clearly document the efforts of their search and screening process, constructing a database of the pool of evidence considered in their evaluations. Additionally, applying code to the specific extracted study features which influence a decision to assign a study as “key”, “supporting”, “weight-of-evidence” etc. serves to document the basis for these decisions in a structured and queryable way. As registrants submit SEMs at the level of single substances, these efforts can be merged to build a SEM that spans all registered substances. This facilitates appraisal of registrants' choices of key study in the context of the wider evidence base. The ability to explore trends in the features influencing assignment of key studies may even assist in refining and improving the registration process – as emerging issues or shortcomings can be quickly evidenced.

5.2. Problem formulation

Beyond offering improvements in transparency during the data gathering phase, SEMs may be of particular value to the problem formulation stage of regulatory decision-making. Problem formulation is a prerequisite to conducting a chemical risk assessment, identifying an

issue of regulatory relevance around which the assessment will be focused (Solomon et al., 2016). These issues can be subtle and difficult to identify at a sufficiently early stage in the field of environmental health, putting the problem formulation process at risk of focusing on issues of lower severity or significance. In implementing a SEM with a broad (lower resolution) coding process, but with a key focus on the hierarchy of coded data and the nature in which this data is related, trends in the evidence base can be effectively and efficiently identified. This allows risk assessors to use these broad, coded parameters to reliably identify problems in need of further assessment, either through secondary syntheses (if the SEM presents a sufficiently large evidence cluster) or primary research (if the SEM indicates an evidence gap).

5.3. Read-across

Identifying trends in the evidence base may also play a significant role in read-across applications. Read-across allows the toxicologically relevant properties of a chemical to be inferred by comparison with a structurally similar chemical of known toxicological behaviour (European Chemicals Agency, 2017a). Read-across aligns well with the need to make best use of existing evidence (van Leeuwen et al., 2009), and the storage of data in a related manner within a SEM could allow the identification of appropriate read-across scenarios. In filtering an evidence map by outcome features, exposures which behave in a similar manner can be identified and investigated further for chemical similarity and/or shared modes of action. This information can be used to group substances, such that data-rich members of the group can be used to make predictions about data-poor members, without pursuing further primary research (Vink et al., 2010). Conversely, filtering an evidence map by chemical group or structural similarity may allow identification of shared outcomes, of similar relevance to read-across applications.

5.4. Evidence surveillance

Once regulation is in place, it is vital that it is kept up to date. Such is the role of the ongoing, evidence surveillance phase of regulatory decision-making. Within REACH, registrants are required to update their registration dossiers “whenever new information is available” (European Chemicals Agency, 2017b), such that dossiers are living products. However, a report commissioned by the European Chemicals Agency (ECHA) found that 64% of REACH registration dossiers submitted to ECHA since 2008 have never been updated (Amec Foster Wheeler Environment and Infrastructure UK Limited, 2017). The report details several obstacles experienced by registrants faced with updating dossiers, including technical difficulties, issues of ownership or responsibility for updates among co- and lead registrants, the potentially labour-intensive nature of updating dossiers and a perception of REACH registration being the “end of a process”.

Openly accessible and easily updated SEMs may serve to address such obstacles. As the population of a SEM database does not require detailed analysis or complex interpretation of the raw data, SEMs could be amenable to automation. Technological advances in text-mining and artificial intelligence might assist the automatic screening, extraction and coding of new information as it is published, based on the data fields and coding ontologies used to populate the original SEM. Although some years away from implementation, application of SEM methodology in the interim will promote fast uptake of such technological advances.

6. Conclusion

Systematic evidence mapping presents a transparent and robust methodological framework with which to assess the evidence landscape at the level of individual chemical risk management and innovation, to regulatory decision-making in chemicals policy. The broad scope of SEMs lowers the barrier to evidence synthesis in chemical risk assessment through more efficient use of resources. Future developments in text mining and machine learning are likely to further reduce the resource intensity of the methodology, and of chemical risk assessment in general. These advances will enable the automatic production of highly resolved SEMs capable of synthesising evidence or feeding predictive models.

In the interim pursuit of a more evidence-based approach to chemicals policy, the resource strain associated with producing a SEM can be managed through adaptation of the methodology to present day limitations. Depending on the needs of the user and the constraints of their use case, SEM methodology is sufficiently flexible that it may be adapted (e.g. by searching fewer databases, extracting data based on only title/abstract etc.) without compromising the utility of the end product in the same way as the results of a synthesis might be adversely affected by modification of systematic review methodology. By working closely with stakeholders to define objectives, the scope of the SEM (i.e. bibliographic databases covered, types of studies included, etc.) can be adjusted as appropriate to objectives. For example, critical appraisal of studies may not be imperative to the aim of the SEM and may therefore be omitted or might be planned as part of a stepwise approach after the SEM identifies pockets of evidence of interest to stakeholders. Although designed to reduce the resource strain of SEM exercises, such flexible adaptation of the methodology does not compromise the fitness-for-purpose of SEMs as a means of identifying and comparing trends in the availability of evidence in a vast and heterogeneous information landscape.

Consequently, examples of research activities producing fit-for-purpose SEM outputs and/or developing aspects of SEM methodology specific to chemicals policy contexts are beginning to emerge (Beverly, 2019), with research institutes such as NTP-OHAT and The Endocrine Disruption Exchange (TEDX) conducting evidence mapping activities (NTP-OHAT, 2019; The Endocrine Disruption Exchange, 2019). A key consideration for these emerging efforts is the accessibility of SEMs' queryable output for non-technical audiences. To this end, researchers have made use of a variety of readily available and user-friendly tools (e.g. Datawrapper GmbH, 2019; IBM, 2019; QlikTech International AB, 2019; Tableau Software, 2019 etc.) to facilitate visualization of, and promote interaction with, the data collated in evidence surveillance exercises (e.g. Pelch et al., 2019; Walker et al., 2018). These tools may similarly serve to lower the barrier to accessing (as well as producing) SEMs, provided the underlying database is made available for more specialist users. Although future technological advances will have significant implications for the production and use of SEMs, these efforts indicate how SEM methodology can be effectively applied in present day, highlighting how SEMs can be adapted for engaging with a variety of stakeholders. More immediate establishment of (adapted) SEM infrastructure in current regulatory workflows will therefore not only lower resource barriers to evidence-based decision-making, but will ensure that technological advances in automation, and in SEM methodology itself, can be readily exploited by regulatory decision-makers in chemicals risk management.

Funding sources

TW's PhD is financially supported by the Centre for Global Eco-Innovation (a programme funded by the European Regional Development Fund) and Yordas Group, a global consultancy in the area of chemical safety, regulations and sustainability. PW's contribution to the manuscript was funded by the Evidence-Based Toxicology Collaboration at Johns Hopkins Bloomberg School of Public Health. AR and VW were supported by the National Institute of Environmental Health Sciences, Division of the National Toxicology Program. The authors declare they have no actual or potential competing financial interests.

Author contributions

TW, PW, CH, AR and VW established the principal ideas for the manuscript and developed an outline. TW wrote the first draft of the manuscript following further discussion of the concept with PW and CH. All authors reviewed and edited the manuscript and contributed to its development.

Acknowledgements

The authors would like to acknowledge the input of the two internal reviewers at the National Toxicology Program, Kristen Ryan and Scott Masten, who provided suggestions for improving the manuscript prior to submission. The authors would also like to thank Miriam Sturdee for help with re-formatting Fig. 1.

References

Aiassa, E., Higgins, J.P.T., Frampton, G.K., Greiner, M., Afonso, A., Amzal, B., Verloo, D., 2015. Applicability and feasibility of systematic review for performing evidence-based risk assessment in food and feed safety. Crit. Rev. Food Sci. Nutr. 55 (7), 1026–1034. https://doi.org/10.1080/10408398.2013.769933.

Amec Foster Wheeler Environment & Infrastructure UK Limited, 2017. A Study to Gather Insights on the Drivers, Barriers, Costs and Benefits for Updating REACH Registration and CLP Notification Dossiers.

Baker, N., Boobis, A., Burgoon, L., Carney, E., Currie, R., Fritsche, E., Daston, G., 2018. Building a developmental toxicity ontology. Birth Defects Res. 110 (6), 502–518. https://doi.org/10.1002/bdr2.1189.

Beverly, B., 2019. Abstract 3267: Potential Alternatives to Systematic Review: Evidence Maps and Scoping Reviews. Retrieved from. https://www.toxicology.org/events/am/AM2019/program-details.asp.

Bolden, A.L., Rochester, J.R., Schultz, K., Kwiatkowski, C.F., 2017. Polycyclic aromatic hydrocarbons and female reproductive health: a scoping review. Reprod. Toxicol. 73, 61–74. https://doi.org/10.1016/j.reprotox.2017.07.012.

Borah, R., Brown, A.W., Capers, P.L., Kaiser, K.A., 2017. Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry. BMJ Open 7 (2), 1–7. https://doi.org/10.1136/bmjopen-2016-012545.

CAMARADES-NC3Rs, 2019. Preclinical Systematic Review & Meta-analysis Facility (SyRF). Retrieved February 6, 2019, from. http://syrf.org.uk/.

Chalmers, I., Hedges, L., Cooper, H., 2002. A brief history of research synthesis. Eval. Health Prof. 25 (1), 12–37. https://doi.org/10.1177/0163278702025001003.

Clapton, J., Rutter, D., Sharif, N., 2009. SCIE Systematic Mapping Guidance April 2009 MG. Retrieved from. https://www.scie.org.uk/publications/researchresources/rr03.pdf.

Covidence, 2019. Covidence. Retrieved January 14, 2019, from. https://www.covidence.org/home.

Datawrapper GmbH, 2019. Datawrapper. Retrieved from. https://www.datawrapper.de/.

EPA, 2018. Application of Systematic Review in Tsca Risk Evaluations, 1–247. Retrieved from. https://www.epa.gov/sites/production/files/2018-06/documents/final_application_of_sr_in_tsca_05-31-18.pdf.

European Chemicals Agency, 2011. Guidance on Information Requirements and Chemical Safety Assessment Chapter R. 4: Evaluation of Available Information. Retrieved from. https://echa.europa.eu/documents/10162/13643/information_requirements_r4_en.pdf.

European Chemicals Agency, 2016. Guidance on registration. https://doi.org/ECHA-12-G-07-EN.

European Chemicals Agency, 2017a. Guidance on information requirements and chemical safety assessment: QSARs and grouping of chemicals. In: Guidance for the Implementation of REACH, pp. 134. https://doi.org/10.2823/43472.

European Chemicals Agency, 2017b. Study finds companies lack incentives for updating their REACH registrations. Retrieved November 8, 2018, from. https://echa.europa.eu/-/study-finds-companies-lack-incentives-for-updating-their-reach-registrations.

European Chemicals Agency, 2018a. Information requirements. Retrieved November 9, 2018, from. https://echa.europa.eu/regulations/reach/registration/information-requirements.

European Chemicals Agency, 2018b. Strategy for gathering your data. Retrieved November 9, 2018, from. https://echa.europa.eu/support/registration/strategy-for-gathering-your-data.

European Comission, 2007. REACH in Brief. https://doi.org/10.1067/j.cpsurg.2007.09.005.

Evidence Partners, 2019. DistillerSR. Retrieved January 14, 2019, from. https://www.evidencepartners.com/products/distillersr-systematic-review-software/.

Garg, A.X., Hackam, D., Tonelli, M., 2008. Systematic review and meta-analysis: when one study is just not enough. Clin. J. Am. Soc. Nephrol. 3 (1), 253–260. https://doi.org/10.2215/CJN.01430307.

Glasziou, P., Djulbegovic, B., Burls, A., 2006. Are systematic reviews more cost-effective than randomised trials? Lancet 367 (9528), 2057–2058. https://doi.org/10.1016/S0140-6736(06)68919-8.

Grant, M.J., Booth, A., 2009. A typology of reviews: an analysis of 14 review types and associated methodologies. Health Inf. Libr. J. 26 (2), 91–108. https://doi.org/10.1111/j.1471-1842.2009.00848.x.

Haddaway, N.R., Westgate, M.J., 2018. Predicting the time needed for environmental systematic reviews and systematic maps. Conserv. Biol. 0 (0), 1–10. https://doi.org/10.1111/cobi.13231.

Haddaway, N.R., Bernes, C., Jonsson, B.-G., Hedlund, K., 2016. The benefits of systematic mapping to evidence-based environmental management. Ambio 45 (5), 613–620. https://doi.org/10.1007/s13280-016-0773-x.

HAWC Project, 2013. HAWC: Health Assessment Workplace Collaborative. Retrieved from. https://hawcproject.org/.

Higgins, J.P.T., Green, S., 2011. Cochrane Handbook for Systematic Reviews of Interventions (editors) Retrieved from. http://handbook-5-1.cochrane.org/.

Hoffmann, S., de Vries, R.B.M., Stephens, M.L., Beck, N.B., Dirven, H.A.A.M., Fowle, J.R., Goodman, J.E., Hartung, T., Kimber, I., Lalu, M.M., Thayler, K., Whaley, P., Wikoff, D., Tsaioun, K., 2017. A primer on systematic reviews in toxicology. Arch. Toxicol. 91 (7), 2551–2575. https://doi.org/10.1007/s00204-017-1980-3.

Hooijmans, C.R., Rovers, M., de Vries, R.B., 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), 356–357. https://doi.org/10.1258/la.2012.012052.

IBM. (2019). Cognos Analytics. Retrieved from https://www.ibm.com/analytics/cognos-analytics

Ingre-Khans, E., Ågerstrand, M., Beronius, A., Rudén, C., 2019. Reliability and relevance evaluations of REACH data. Toxicol. Res. 8 (1), 46–56. https://doi.org/10.1039/c8tx00216a.

Innvaer, S., Vist, G., Trommald, M., Oxman, A., 2002. Review article health policy-makers' perceptions of their use of evidence: a systematic review. J. Health Serv. Res. Policy 7 (4), 239–244. https://doi.org/10.1258/135581902320432778.

James, K.L., Randall, N.P., Haddaway, N.R., 2016. A methodology for systematic mapping in environmental sciences. Environ. Evid. 5 (1), 7. https://doi.org/10.1186/s13750-016-0059-6.

Kelly, S.E., Moher, D., Clifford, T.J., 2016. Quality of conduct and reporting in rapid reviews: an exploration of compliance with PRISMA and AMSTAR guidelines. Syst. Rev. 5 (1), 1–19. https://doi.org/10.1186/s13643-016-0258-9.

Klimisch, H.-J., Andreae, M., Tillmann, U., 1997. A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data. Regul. Toxicol. Pharmacol. 25 (1), 1–5. https://doi.org/10.1006/rtph.1996.1076.

Mandrioli, D., Schlünssen, V., Ádám, B., Cohen, R. A., Colosio, C., Chen, W., … Scheepers, P. T. J. (2018). WHO/ILO work-related burden of disease and injury: protocol for systematic reviews of occupational exposure to dusts and/or fibres and of the effect of occupational exposure to dusts and/or fibres on pneumoconiosis. Environ. Int., 119(June), 174–185. doi:https://doi.org/10.1016/j.envint.2018.06.005.

Mara-eves, A.O., Thomas, J., Mcnaught, J., Miwa, M., Ananiadou, S., 2015. Using text mining for study identification in systematic reviews: a systematic review of current approaches. Syst. Rev. 1–22. https://doi.org/10.1186/2046-4053-4-5.

Marshall, I., Marshall, R., Wallace, B., Brassey, J., Thomas, J., 2018. Rapid reviews may produce different results to systematic reviews: a meta-epidemiological study. J. Clin. Epidemiol. https://doi.org/10.1016/j.jclinepi.2018.12.015.

Morgan, R.L., Thayer, K.A., Bero, L., Bruce, N., Falck-Ytter, Y., Ghersi, D., Schünemann, H.J., 2016. GRADE: assessing the quality of evidence in environmental and occupational health. Environ. Int. 92–93, 611–616. https://doi.org/10.1016/j.envint.2016.01.004.

National Academy of Sciences, 1983. Risk Assessment in the Federal Government: Managing the Process. National Academy Press, Washington, DC. https://doi.org/10.17226/366.

National Centre for the Replacement Refinement & Reduction of Animals in Research, 2018. The 3Rs. Retrieved November 8, 2018, from. https://www.nc3rs.org.uk/the-3rs.

National Institute for Health Research, 2018. PROSPERO - International Prospective Register of Systematic Reviews. Retrieved December 31, 2018, from. https://www.crd.york.ac.uk/prospero/.

NTP, 2015. Handbook for Conducting a Literature-based Health Assessment Using OHAT Approach for Systemic Review and Evidence Integration. pp. 1–98. Retrieved from. https://ntp.niehs.nih.gov/ntp/ohat/pubs/handbookjan2015_508.pdf.

NTP-OHAT, 2019. About the Office of Health Assessment and Translation. Retrieved from. http://ntp.niehs.nih.gov/.

Oliver, S., Dickson, K., 2016. Policy-relevant systematic reviews to strengthen health systems: models and mechanisms to support their production. Evid. Policy 12 (2), 235–259. https://doi.org/10.1332/174426415X14399963605641.

Papathanasopoulou, E., Queirós, A.M., Beaumont, N., Hooper, T., Nunes, J., 2016. What evidence exists on the local impacts of energy systems on marine ecosystem services: a systematic map. Environ. Evid. 5 (1), 1–12. https://doi.org/10.1186/s13750-016-0075-6.

Pelch, K.E., Bolden, A.L., Kwiatkowski, C.F., 2019. Environmental chemicals and autism: a scoping review of the human and animal research. Environ. Health Perspect. 127 (4), 46001. https://doi.org/10.1289/EHP4386.

Pool, R., Rusch, E., 2014. Identifying and Reducing Environmental Health Risks of Chemicals in Our Society: Workshop Summary. The National Academies Press, Washington, DC.

QlikTech International AB, 2019. Qlik Sense. Retrieved from. https://www.qlik.com/us/products/qlik-sense.

Reynen, E., Robson, R., Ivory, J., Hwee, J., Straus, S.E., Pham, B., Tricco, A.C., 2018. A retrospective comparison of systematic reviews with same-topic rapid reviews. J. Clin. Epidemiol. 96, 23–34. https://doi.org/10.1016/j.jclinepi.2017.12.001.

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. Environ. Health Perspect. 122 (7), 711–718. https://doi.org/10.1289/ehp.1307972.

Science for Nature and People Partnership Evidence-Based Conservation working group, Conservation International, Datakind, 2018. Colandr. Retrieved January 14, 2019, from. https://www.colandrapp.com/signin.

Sciome, 2018. SWIFT-Review. Retrieved January 14, 2019, from. https://www.sciome.com/swift-review/.

Solomon, K.R., Wilks, M.F., Bachman, A., Boobis, A., Moretto, A., Pastoor, T.P., Embry, M.R., 2016. Problem formulation for risk assessment of combined exposures to chemicals and other stressors in humans. Crit. Rev. Toxicol. 46 (10), 835–844. https://doi.org/10.1080/10408444.2016.1211617.

Tableau Software (2019). Tableau. Retrieved from https://www.tableau.com/

The Endocrine Disruption Exchange. (2019). TEDX Publications. Retrieved from https://endocrinedisruption.org/interactive-tools/publications/

The National Academies of Sciences, 2017. Application of Systematic Review Methods in an Overall Strategy for Evaluating Low-dose Toxicity From Endocrine Active Chemicals. Washington, DC: The. National Academies Press. https://doi.org/10.17226/24758.

Thomas, J., Brunton, J., Graziosi, S., 2010. EPPI-Reviewer 4.0: Software for Research Synthesis. London: Social Science Research Unit. Institute of Education, University of London. http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=3299#Research.

U.S. National Library of Medicine, 2016. Unified Medical Language System (UMLS). Retrieved December 31, 2018. In: from, . https://www.nlm.nih.gov/research/umls/.

van Leeuwen, K., Schultz, T.W., Henry, T., Diderich, B., Veith, G.D., 2009. Using chemical categories to fill data gaps in hazard assessment. SAR QSAR Environ. Res. 20 (3–4), 207–220. https://doi.org/10.1080/10629360902949179.

Vandenberg, L.N., Ågerstrand, M., Beronius, A., Beausoleil, C., Bergman, Å., Bero, L.A., ... Rudén, C., 2016. A proposed framework for the systematic review and integrated assessment (SYRINA) of endocrine disrupting chemicals. Envrion. Health 15 (1), 1–19. https://doi.org/10.1186/s12940-016-0156-6.

Vink, S.R., Mikkers, J., Bouwman, T., Marquart, H., Kroese, E.D., 2010. Use of read-across and tiered exposure assessment in risk assessment under REACH – a case study on a phase-in substance. Regul. Toxicol. Pharmacol. 58, 64–71. https://doi.org/10.1016/j.yrtph.2010.04.004.

Walker, V.R., Boyles, A.L., Pelch, K.E., Holmgren, S.D., Shapiro, A.J., Blystone, C.R., Rooney, A.A., 2018. Human and animal evidence of potential transgenerational inheritance of health effects: an evidence map and state-of-the-science evaluation. Environ. Int. 115 (December 2017), 48–69. https://doi.org/10.1016/j.envint.2017.12.032.

Whaley, P., Halsall, C., 2016. Assuring high-quality evidence reviews for chemical risk assessment: five lessons from guest editing the first environmental health journal special issue dedicated to systematic review. Environ. Int. 92–93, 553–555. https://doi.org/10.1016/j.envint.2016.04.016.

Whaley, P., Halsall, C., Ågerstrand, M., Aiassa, E., Benford, D., Bilotta, G., ... Taylor, D., 2016. Implementing systematic review techniques in chemical risk assessment: challenges, opportunities and recommendations. Environ. Int. 92–93, 556–564. https://doi.org/10.1016/j.envint.2015.11.002.

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. Environ. Health Perspect. 122 (10), 1007–1014. https://doi.org/10.1289/ehp.1307175.

Chapter 3: A survey of systematic evidence mapping practice and the case for knowledge graphs in environmental health & toxicology

This chapter was published in the journal *Toxicological Sciences* in February 2020, volume 175. **Electronic supplementary materials are available for this chapter.**

The candidate's contribution was structuring own and co-authors' ideas as a manuscript; establishing the assessment criteria for the survey; conducting the survey in duplicate with a co-author; writing the manuscript for supervisor review; co-ordinating co-author feedback; submission of manuscript.

Candidate: ..... Date: 30/01/2020
Ms. Taylor A. M. Wolffe

Supervisor: ..... Date: 30/01/2020
Prof. Crispin J. Halsall

OXFORD
SOT | Society of Toxicology
academic.oup.com/toxsci

TOXICOLOGICAL SCIENCES, 2020, 35–49

doi: 10.1093/toxsci/kfaa025
Advance Access Publication Date: February 25, 2020
Research Article

A Survey of Systematic Evidence Mapping Practice and the Case for Knowledge Graphs in Environmental Health and Toxicology

Taylor A.M. Wolffe,*,†,1 John Vidler,‡ Crispin Halsall,* Neil Hunt,† and Paul Whaley*,§

*Lancaster Environment Centre, Lancaster University, Lancaster LA1 4YQ, UK; †Yordas Group, Lancaster University, Lancaster LA1 4YQ, UK; ‡School of Computing and Communications, Lancaster University, Lancaster LA1 4WA, UK; and §Evidence-Based Toxicology Collaboration, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland 21205

[1]To whom correspondence should be addressed at E-mail: t.wolffe@lancaster.ac.uk.

ABSTRACT

Systematic evidence mapping offers a robust and transparent methodology for facilitating evidence-based approaches to decision-making in chemicals policy and wider environmental health (EH). Interest in the methodology is growing; however, its application in EH is still novel. To facilitate the production of effective systematic evidence maps for EH use cases, we survey the successful application of evidence mapping in other fields where the methodology is more established. Focusing on issues of "data storage technology," "data integrity," "data accessibility," and "transparency," we characterize current evidence mapping practice and critically review its potential value for EH contexts. We note that rigid, flat data tables and schema-first approaches dominate current mapping methods and highlight how this practice is ill-suited to the highly connected, heterogeneous, and complex nature of EH data. We propose this challenge is overcome by storing and structuring data as "knowledge graphs." Knowledge graphs offer a flexible, schemaless, and scalable model for systematically mapping the EH literature. Associated technologies, such as ontologies, are well-suited to the long-term goals of systematic mapping methodology in promoting resource-efficient access to the wider EH evidence base. Several graph storage implementations are readily available, with a variety of proven use cases in other fields. Thus, developing and adapting systematic evidence mapping for EH should utilize these graph-based resources to ensure the production of scalable, interoperable, and robust maps to aid decision-making processes in chemicals policy and wider EH.

Key words: systematic evidence map; knowledge graph; evidence synthesis.

Data relevant to assessing the human and ecological health risks associated with exposure to chemical substances are increasingly available to stakeholders (Barra Caracciolo *et al.*, 2013; Lewis *et al.*, 2016). This trend is owed to a variety of factors, including the advent of the Internet and increasingly sensitive analytical techniques (Lewis *et al.*, 2016), regulatory and economic changes (Lyndon, 1989; Pool and Rusch, 2014), demands for increased transparency (Ingre-Khans *et al.*, 2016), stricter regulatory data requirements (Commission of the European Communities, 2001; United States Environmental Protection Agency, 2016), reform of regulatory reliance on *in vivo* toxicity testing (ECHA, 2016), and a continually growing chemicals industry. The growing pool of available evidence has significant potential for informing regulatory and risk management decision making.

Box 1 Glossary of Terms

Database	An organized and structured collection of information (data) stored electronically within a computer system, which allows data to be accessed, manipulated, and updated.
Systematic evidence map (SEM)	A queryable database of systematically gathered evidence (eg, academic literature and industry reports). SEMs extract and structure data and/or metadata for exploration following a rigorous methodology which aims to minimize bias and maximize transparency.
Coding	The process of assigning controlled vocabulary labels or categories (referred to as "code") to data, which allows comparisons to be drawn despite the heterogeneity of the underlying dataset. For example, extracted data such as "mouse," "rat," and "guinea pig" might all be coded as "rodent" for broad comparison.
Query	A request for data from a database. By requesting data that meets a particular set of conditions, users can query a database for a subset of information of relevance to their specific research interests.
Schema	The organizational plan ("blueprint") for the structure of a database, detailing the entities stored in the database, the attributes associated with those entities, how those entities are related, what datatypes can be stored in the database, etc.
Schemaless	Refers to databases which do not have a fixed and predefined schema.
Schema, on-write	Refers to the application of a schema before data is stored (written) to the database.
Schema, on-read	Refers to the application of a schema after data has been written to the database, at the time the data is accessed (read).
Ontology	A shared and reusable conceptualization of a domain which applies a logically related controlled vocabulary to describe the domain concepts, their properties and relations.

Evidence-based approaches aim to minimize the bias associated with cherry-picking an unrepresentative subset of evidence for consideration in the decision-making process. They advocate for robust, transparent consideration of *all* relevant, available data and are the core of the evidence-based toxicology movement (Hoffmann and Hartung, 2006; Hoffmann *et al.*, 2017). However, locating, organizing, and evaluating all relevant data is challenging when the quantity of that data is very large and growing exponentially.

Systematic evidence mapping is 1 such evidence-based approach to drawing into consideration all data which are relevant to chemicals policy and risk management workflows (see Wolffe *et al.*, 2019). Systematic evidence maps (SEMs) are queryable databases of systematically gathered research (Box 1). They provide users with the computational access needed to organize, compare, analyze, and explore trends across a broader evidence base (Clapton *et al.*, 2009; James *et al.*, 2016) by:

- Collating data from different sources and storing it in a single location, such that users need only query a single database to satisfy their information requirements;
- Extracting unstructured data and storing it in a structured format, such that data can be programmatically accessed and analyzed;
- Categorizing extracted data using controlled vocabulary code, such that evidence can be broadly and meaningfully compared despite its inherent heterogeneity.

SEMs organize and characterize an evidence base such that it can be explored by a variety of end-users with varied specific research interests. The methodology was developed to address some of the limitations of systematic review and has found application in fields where formulating a single, narrowly focused review question is difficult or uninformative (Haddaway *et al.*, 2016; James *et al.*, 2016; Oliver and Dickson, 2016; Wolffe *et al.*, 2019). Similarly faced with this challenge is chemicals policy and the fields which it encompasses, ie, environmental health (EH) and toxicology. It is difficult to frame a single research question with a scope which is simultaneously narrow enough to elicit the synthesis of a coherent conclusion through systematic review, and also broad enough to address the varied information requirements of chemicals policy workflows. This means that potentially several syntheses over multiple systematic reviews are required to facilitate a single decision-making process in chemicals policy. However, the significant demand for time and resources associated with systematic reviews, and the unmatched resource availability of chemicals policy, necessitates a priority setting, or problem formulation process to ensure the most efficient use of systematic review. Thus, systematic evidence mapping provides a valuable first step in this prioritization process, where the identification of emerging trends across the wider evidence base ensures resources can be targeted most efficiently (see Wolffe *et al.* [2019] for further discussion of the applications of SEMs in chemicals policy).

These issues are likely to become increasingly pressing as the chemicals policy paradigm shifts toward more evidence-based approaches and methods such as systematic review gain prominence. For example, agencies such as the U.S. EPA (EPA, 2018), EFSA (European Food Safety Authority, 2010), and WHO (Mandrioli *et al.*, 2018; World Health Organization, 2019) have already begun to incorporate systematic review in their chemical risk assessment frameworks. Thus, ensuring that evidence synthesis efforts are targeting the most appropriate issues, and that the data collated for synthesis can be accessed for alternative applications, potentially across agencies, is increasingly important.

Interest in the application of SEM methodology for this context is beginning to emerge in the form of SEM exercises targeting chemicals policy issues (Martin *et al.*, 2018; Pelch *et al.*, 2019), various working groups expanding their evidence synthesis activities to include broader scoping and surveillance exercises (NTP-OHAT, 2019; Pelch *et al.*, 2019; The Endocrine Disruption Exchange, 2019; Walker *et al.*, 2018), and conference sessions discussing the potential benefits of SEMs for EH (Beverly, 2019). This emerging interest in SEM methodology, and its ability to facilitate evidence-based approaches, necessitates study of the factors key to its successful adaptation to EH contexts.

Therefore, we seek to understand how SEM databases are built and presented to end-users in fields where the practice is more mature. We hope that contextualizing this understanding within the needs of chemicals policy, risk management, and wider EH research will expedite the development of effective evidence mapping methods in this domain.

To achieve this, we examine the current state-of-the-art and common practices associated with constructing and presenting a SEM database in environmental management, a field with a strong history of systematic mapping publications and method development (Collaboration for Environmental Evidence, 2019c; Haddaway *et al.*, 2016, 2018a; James *et al.*, 2016). We discuss the implications of current practices for EH and highlight the challenges associated with using rigid data structures for storing the highly connected and heterogeneous data associated with the field. We outline the need for more flexible data structures in

Table 1. The Concepts Used to Guide Data Extraction and Subsequent Assessment and Discussion of the Outputs of CEE Systematic Mapping Exercises

Concept	Definition	Metadata Extracted
Data storage technology	How data extracted and collated during the systematic mapping exercise were stored for future exploration	Format in which the systematic map database is presented to users (eg, spreadsheet, relational database, in-text data table, and in-text figure).
Data integrity	How accurately the systematic map is able to represent the raw study data on which it is based	How the relationships between entities (or study attributes) which underpin the raw data are maintained in the systematic map.
Data accessibility	How easy it is for end-users to access the data relevant to their research interests, or the ability of the systematic map to return data relevant to an end-user's queries	The querying mechanisms recommended in the systematic map's study report (eg, filtering table columns and navigating interactive dashboards).
Transparency	The ability of end-users to verify how the systematic map represents the raw study data on which it is based, ie, whether the map maintains a link between raw extracted data and eg, controlled vocabulary code.	Whether the map maintains a link between raw extracted data and controlled vocabulary code (eg, map presents code-only, map presents raw data and code), and how this link is maintained.

EH SEMs and introduce the concept of "knowledge graphs" as an effective and intuitive model for the storage and querying of highly connected EH data. Finally, we discuss graph-based SEMs in the context of current, complementary efforts in the development of toxicological ontologies, outlining the future of systematic evidence mapping for regulatory decision making.

MATERIALS AND METHODS

Survey of published Collaboration for Environmental Evidence SEMs. We identified a dataset of exemplar SEMs for analysis: the complete set of SEMs of the Collaboration for Environmental Evidence (CEE). These maps were chosen because of CEE's role in pioneering the adaptation of systematic mapping methodology from the social sciences (Clapton *et al.*, 2009; James *et al.*, 2016). Through example (Collaboration for Environmental Evidence, 2019b), communication (Collaboration for Environmental Evidence, 2019a), published guidance (James *et al.*, 2016), and reporting standards (Haddaway *et al.*, 2018b), CEE advocate for systematic mapping and represent an ongoing case study for how the methodology can be developed as a policy and decision-making tool. Understanding how systematic map outputs serve this function, and what methodological adaptation is required to produce these outputs, is vital for successfully applying the methodology in EH. Thus, the outputs (ie, the queryable databases) of CEE's more firmly established systematic mapping practice were surveyed.

All CEE systematic maps completed before July 2019 were identified in the CEE Library (http://www.environmentalevidence.org/completed-reviews, last accessed July 2019). The study reports and the Supplementary information for these maps were downloaded and key metadata extracted, including title, authors, publication date, and map objectives (Supplementary Table 1). Metadata regarding the output of the systematic mapping exercises were then gathered and assessed in duplicate by T.A.M.W. and P.W. using a data extraction sheet which asked open-ended questions relating to 4 key themes of analysis: data storage technology; data integrity; data accessibility; and transparency (Table 1). These themes were developed in discussion among J.V., T.A.M.W., and P.W.

"Data storage technology" concerns the software used to construct the systematic map databases and their associated data storage formats.

"Data integrity" concerns the structures of the CEE maps. Although an important aspect of data integrity, appraising the data extraction efforts of mappers (ie, confirming that the data extracted, coded, and stored in the database are an accurate representation of their raw counterparts in the primary literature) was beyond the scope of this exercise. Rather than verifying the data, *how* that data are represented (regardless of *what* is represented) by the systematic map database output was assessed by focusing on the ability of the systematic map to maintain the relationships which underpin these data. For example, a mapper may have extracted data from a study which investigates outcomes in a population. Although the mapper may have extracted data such as "outcome x" and "population y"—the manner in which the database structures and organizes these data will determine whether end-users can decipher that "outcome x" is somehow related to "population y."

"Data accessibility" concerns the capacity for CEE's systematic maps to facilitate data exploration by end-users. Systematic maps are research products in their own right (Haddaway *et al.*, 2016). They should therefore present end-users with a means of programmatically accessing and querying the data they store, such that trends in potentially large datasets can be quickly identified with minimal manual effort. Accessibility is an important consideration when producing maps for an audience of varied technical skill, where ensuring that the map is accessible for nonspecialist users should not compromise the ability of more technical users to run complex queries. Therefore, the extent to which CEE systematic mapping exercises consider accessibility from the perspective of users was surveyed by extracting eg, details on the level of guidance provided to end-users wishing to query the systematic map database, and recording P.W. and T.A.M.W.'s experience of interacting with and querying the maps.

Finally, "transparency" concerns how systematic maps facilitated an end-user's ability to validate the extent to which the data presented in a map represents the data in the primary research. This was achieved by determining whether the map preserved a link between raw data and assigned controlled vocabulary labels/categories ("code" - see Box 1)).

T.A.M.W. and P.W. independently noted answers to the data extraction questions before discussing and agreeing on an aggregate, consensus view. This was to contribute to comprehensive coverage of potential discussion points in relation to each theme. These aggregate assessments are presented in Supplementary Tables 1–6 and are used to evidence the state-of-the-art in terms of producing queryable systematic map databases for exploration of the environmental management

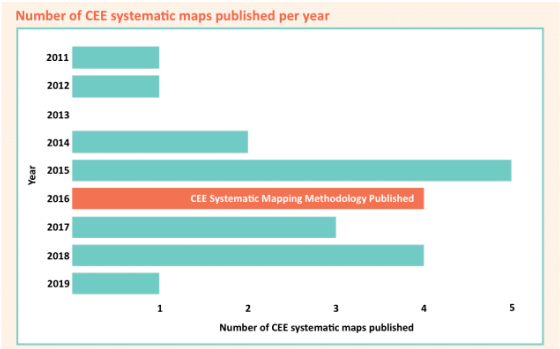

Figure 1. Publication history of CEE SEMs indicating the number of maps published per year. The year in which the CEE guidance on systematic mapping methods was published (2016) is marked on the corresponding bar (James *et al.*, 2016).

literature. Their contents are referenced throughout the Results and Discussion sections of this survey.

RESULTS

Twenty-one systematic maps covering a variety of topics were identified in the CEE library, published between October 2011 and January 2019 (Figure 1).

The aggregated, narrative assessments of each CEE systematic map can be found in Supplementary Tables 1–6. The extracted data and aggregated assessments for each CEE systematic map are organized as follows:

- Supplementary Table 1—Bibliographic information
- Supplementary Table 2—Data storage technology
- Supplementary Table 3—Data integrity
- Supplementary Table 4—Data accessibility
- Supplementary Table 5—Transparency
- Supplementary Table 6—Additional notes

Excluded Maps

Two systematic maps (Johnson *et al.*, 2011; Mcintosh *et al.*, 2018) are assessed in the Supplementary information but are excluded from further analysis, as neither provided a database output which could be analyzed using our framework. Mcintosh *et al.* (2018) yielded a null result and therefore provided no database; Johnson *et al.* (2011) predated CEE's Environmental Evidence journal and its definition of systematic mapping and, although it is included in the CEE library, presented only in-text tables without an accompanying database.

Data Storage Technology

Two different data storage technologies are used in the outputs of CEE systematic mapping projects: spreadsheets constructed in Microsoft Excel ($n = 14$); and relational databases constructed with the Microsoft Access relational database management system ($n = 5$.) One mapping exercise used both of these technologies to present its outputs in 2 different formats (Haddaway *et al.*, 2014). The 2 versions of Haddaway *et al.* (2014) appear to be identical except that the spreadsheet version includes the results of a critical appraisal process where the relational database version does not. As the spreadsheet version presents the more complete dataset, Haddaway *et al.* (2014) has been coded as a spreadsheet-based systematic map for the purposes of this survey (see Supplementary Table 2, discussed in the "Data Integrity" section). A brief description of each identified storage technology can be found in Table 2.

Data Integrity

A single, flat data table (2-dimensional array of rows and columns) was the output for the majority (84%) of CEE systematic maps (16 of 19 maps surveyed, ignoring any look-up tables housing controlled vocabulary code). 80% (4 out of 5) of the maps using the relational database storage technology were also structured as a single, flat data table.

Three maps presented more than 1 table. Two presented at least 2 tables in separate files which were not formally related to each other (Haddaway *et al.*, 2018a; Sola *et al.*, 2017), and 1 presented multiple tables which were related to each other in a 1:1 manner within a relational database. Systematic maps were considered to be stored in more than 1 table if there was limited overlap of the data fields housed in each table ie, if querying the map required accessing information from more than 1 table. Sola *et al.* (2017) is an example of this, providing the results of its quality appraisal process separately to the data it extracted and coded from the literature—thus any queries investigating critical appraisal in conjunction with another variable require the user to access information from both tables. This distinction was required because some maps, Haddaway *et al.* (2014) and Randall *et al.* (2015), presented their outputs in multiple tables, but the additional tables were simply subsets of the most complete table (ie, there was no data in the smaller tables not already present in the largest table).

Several studies included in the systematic maps contained multiple potential values for a particular attribute eg, if a single study had multiple populations and/or multiple outcomes.

Common strategies for maintaining relationships between such data in the tables of CEE maps included "expanding rows" ($n = 6$), "expanding columns" ($n = 2$), or a combination of both ($n = 5$) (see Figure 2). The remaining 6 maps either did not present/extract studies with multiple potential values per attribute ($n = 1$) or opted to house multiple values within a single cell of the table ($n = 5$, discussed further below).

"Expanding rows" refers to the practice of structuring a data table in *long form:* recording an entity over multiple rows. In long-form tables, a study investigating eg, 3 different outcomes might be recorded over 3 different rows. Although the data entered under the "outcome" data field might be unique in each of these 3 rows, the data for all other attributes will be repeated (Figure 3A).

In contrast, "expanding columns" describes the practice of structuring a data table in wide form; expanding what would be considered a single data field in long-form tables across several columns. Thus, all unique values associated with the data field can be recorded across a single row, eg, a study reporting 3 different outcomes might be recorded across a single row if the "outcome" attribute is split into 3 unique columns (eg, "outcome 1," "outcome 2," and "outcome 3") (Figure 3B).

The other strategy for presenting related data in a table was to record multiple values within a single cell for multiple data fields ($n = 11$), whereas 1 map presented multiple values per cell for only a single data field within the database (this distinction matters for reasons we discuss below). The practice of presenting multiple values in a single cell of the database was observed for most (5 of 6) of the maps which avoided expanding row/column structure, and similarly for most (5 of 6) of the maps adopting a long form, expanded row structure.

Table 2. Description of the Storage Technologies Used by CEE Systematic Maps

Storage Technology	Description
Spreadsheets	Spreadsheets are stand-alone applications which offer functionality for end-users wishing to explore and/or manipulate data (Zynda, 2013). A spreadsheet stores data in the cells of 2-dimensional arrays made up of rows and columns. By referencing the coordinates of cells in mathematical formulae, spreadsheet applications such as Microsoft Excel facilitate analysis, transformation, and visualization of tabular data. Although designed and optimized for quantitative data and accounting applications, spreadsheets are commonly used for storing and organizing data in a variety of research contexts, including systematic mapping exercises.
Relational databases	A relational database uses several formally described tables to organize data. Each table stores instances of an entity (across rows), described by a series of attributes (columns). In contrast to storing data in a single, flat data table, relational databases are able to preserve the connection between related entities. These connections are predefined and created through a system of referencing unique identifiers (primary/foreign keys) in corresponding tables. This allows users to enrich their queries with connected information, such that more complex questions can be asked of the evidence base (Elmasri and Navathe, 2013).

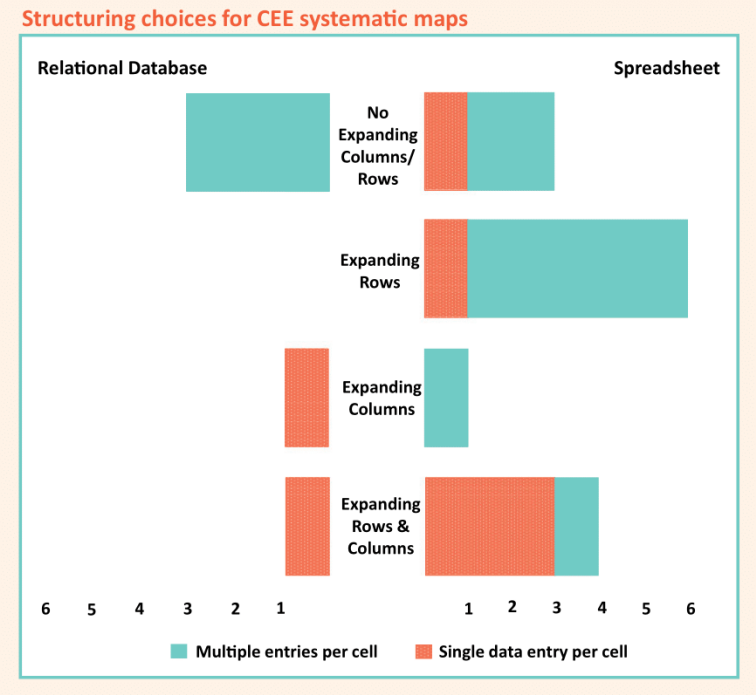

Figure 2. The number of CEE systematic maps that are structured with expanding rows and/or expanding columns as a means of preserving data relationships. Maps using the relational database storage technology are presented on the left, while maps using the spreadsheet storage technology are presented on the right. In addition, the numbers of systematic maps which store multiple values within a single cell of their data table/s are indicated by solid shading, whereas those that do not are indicated by patterned shading.

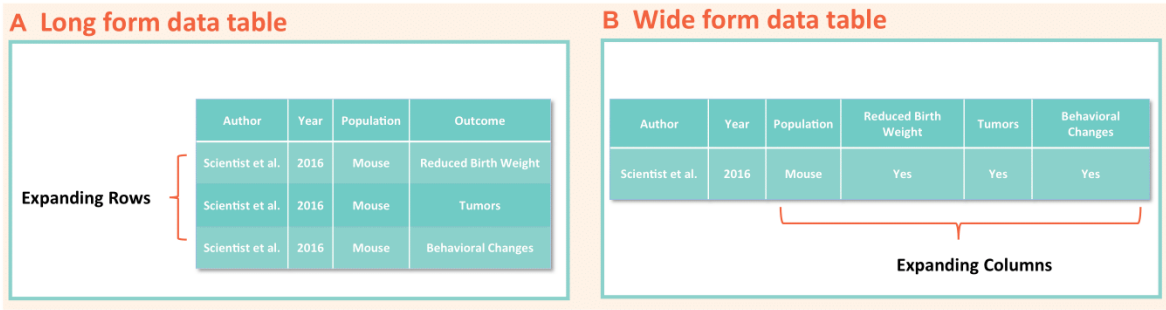

Figure 3. Illustrative example of how "expanding-rows" (A) and "expanding columns" (B) are used in long-form (A) and wide-form (B) tabular data structures, respectively.

Data Accessibility

Eighteen of 19 surveyed systematic maps presented users with static data visualizations within their study reports (eg, bar charts, tables, and heat maps) as a means of accessing trends within the evidence. Six systematic maps additionally provided users with an open-access interactive data visualization dashboard, such that users could choose trends for exploration within the map. Four of the 6 maps supplied comprehensive guidance and/or instruction for users wishing to interact with the visualization dashboard.

Far fewer mapping exercises provided any such comprehensive guidance for querying their database output, with only 2 of 19 maps providing a detailed help file for users wishing to query the database (Haddaway *et al.*, 2014; Randall and James, 2012). This was also seen in mapping exercises presenting guidance on interacting with their data visualization dashboards, none of which provided equivalent detailed guidance for querying the underlying database. Instead, 6 CEE systematic maps dedicated only brief discussion to querying within the text of their study reports, leaving 11 maps which offered no discernible guidance.

Where provided, the querying practices identified in user guidance/instruction were "filtering," "sorting"/"ordering," "searching," or some combination thereof (see Supplementary Table 4). Specific examples of queries which could be run against the database were rarely provided in such guidance, with only 2 of 19 maps providing an illustrative example of how a user's plain-text question is translated into querying the database (Haddaway *et al.*, 2014; Randall and James, 2012), and a further 2 of 19 making only brief mention of how a specific data field might be filtered (Cresswell *et al.*, 2018; Randall *et al.*, 2015). None of the maps reported the queries or querying processes used to generate visualizations or analyses. Two maps (Cheng *et al.*, 2019; McKinnon *et al.*, 2016) indicated that an additional data processing step had been conducted eg, using the statistical programming language R. Cheng *et al.* (2019) provided a link to the code used for this analysis, however the link was broken at the time this survey was conducted.

Transparency

Thirteen of 19 surveyed CEE systematic maps presented only the controlled vocabulary code which was used to classify the data of interest, not recording the raw data itself in the map. Six of 19 maps maintained a link between this code and the raw data/the coders' interpretation of the raw data. Approaches to this included using data fields which contain free-form text alongside the controlled vocabulary terms applied to categorize this free text (5 of 6 maps, Macura *et al.*, 2015), and providing the location of the raw data within the original study report represented as code in the systematic map (1 of 6 maps, Haddaway *et al.*, 2015).

Seventeen of 19 CEE mapping exercises provided a codebook. Codebooks were generally supplied separate to the systematic map database, in a different file and/or format ($n = 14$), although some incorporated codebooks into the database as either look-up tables ($n = 1$, Leisher *et al.*, 2016), or separate spreadsheets within the same workbook as the systematic map ($n = 2$, Bernes *et al.*, 2015, 2017).

Codebooks largely presented the controlled vocabulary terms used to code study attributes (12 of 17) but did not always provide this detail (5 of 17). For codebooks which did provide controlled vocabulary terms, a narrative description or discussion of the potential types of data which might be assigned certain codes was presented in only 2 of the codebooks.

Relationships between controlled vocabulary terms were generally omitted from codebooks and/or the systematic map databases themselves, except for 1 map which structured its code as a hierarchy of nested terms (Haddaway *et al.*, 2015).

DISCUSSION

CEE has been a driving force for the introduction of systematic mapping to the environmental sciences. Their maps act as case studies for adapting evidence-based methodologies to other fields. CEE's involvement of stakeholders in their systematic mapping approach has undoubtedly resulted in outputs of value to those stakeholders and their specific research contexts (Haddaway and Crowe, 2018). The following discussion does not critique the use of CEE's systematic maps for their intended purposes, but instead takes the perspective of EH applications to identify transferable aspects of current practice and remaining challenges.

Systematically Mapping the EH Evidence Base: General Considerations

EH data are complex, heterogeneous, and highly interconnected (Vinken *et al.*, 2014). Chemical risk assessment and risk management seek to understand the outcomes which result from these complex connections—synthesizing evidence of varied resolution and origin eg, considering in combination evidence from bio- and/or environmental monitoring, *in vitro*, *in vivo*, *in silico*, and/or epidemiological studies (Martin *et al.*, 2018; Rhomberg *et al.*, 2013; Vandenberg *et al.*, 2016).

The relationships which hold the disparate EH evidence base together are vital for building a more complete understanding of toxicity. These relationships underpin adverse outcome pathways (ie, how molecular initiating events lead to apical outcomes through a causal pathway of connected key events [Edwards *et al.*, 2015]), quantitative structure-activity models (ie, how the chemical structure of a substance can be quantitatively related to its physicochemical properties and biological activity [Schultz *et al.*, 2003]), read-across applications (ie, where predictions for data-poor substances are based on structurally related data-rich substances) and other key components of chemicals policy workflows. Such relationships are also vital for understanding the impacts of real-world exposures to mixtures of chemical substances (Sexton and Hattis, 2007).

The interconnectedness of the EH evidence base means that even if SEM methodology is used to explore just a subset of EH research, or to facilitate just 1 component of chemicals policy workflows—the data collated, extracted, and coded are likely to be of relevance to a myriad of alternative EH research interests and chemicals policy applications. Thus producing "multi-purpose," interoperable EH SEMs that can be queried according to a variety of specific use cases is the most resource-efficient means of implementing the methodology.

However, many of the complex relationships constituting the EH evidence base are unknown to individual users, who will only have cognitive access to part of the total knowledge space in a given domain. Thus, in addition to facilitating the identification of trends which are based on relationships already known to users, EH SEMs should also facilitate the identification of relationships which are unknown to users. This would enable a more highly resolved and customizable querying process which extends beyond the user's personal understanding of the domain, adding valuable connected contextual information with which to explore and interpret trends. It is this value, gained through accessing as well as exploring relationships—

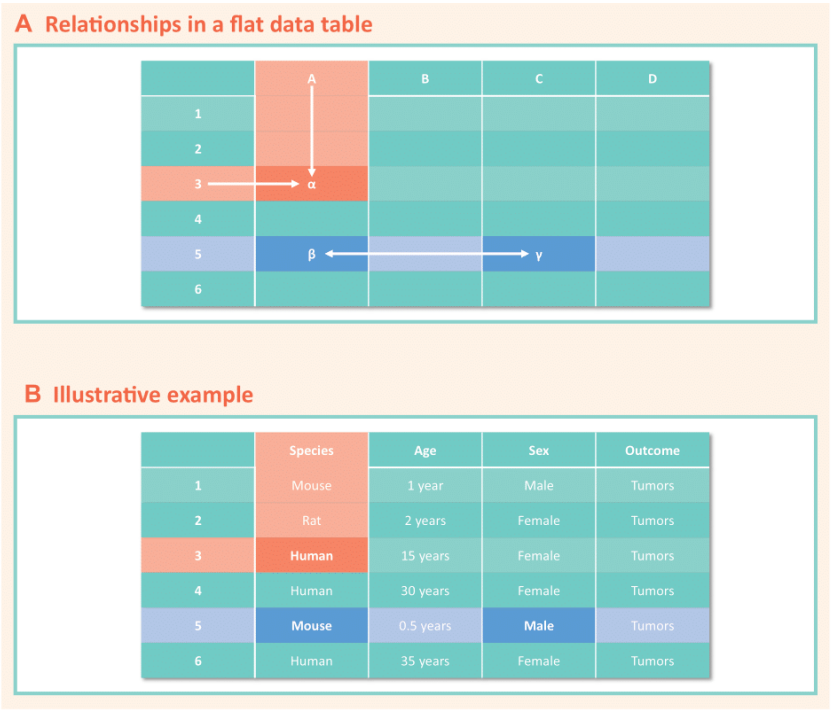

Figure 4. A, The relationship between attribute A and entity 3 is explicit in the formal structure of the array. However, the relationship between attribute A and attribute C is implicit and has to be inferred by the user from features external to the table eg, conventions around interpreting tabular data. The external conventions are not part of, or known to, the table and may not be known to the user. B, For example, a user may (in this case, correctly) infer that "sex" is a property of "species" and not "outcome," but this inference is made using external conventions and contextual understanding—the relationship is not in fact known to the table. All the table can assert is that each entity 1 through 6 has a relationship to properties of sex, age, species, and outcome, respectively.

along with the inherent complexity of those relationships—which makes the flat and rigid tabular data structures currently characterizing CEE systematic maps ill-suited to the task of systematically mapping EH data.

Limitations of Current Evidence Mapping Practice: Data Storage and Structure

Data storage is the fundamental component required for creating a systematic map database, underpinning many of the themes assessed in this survey. This discussion focuses on issues of data storage technology and its close relationship with data integrity.

Use of spreadsheets (and other flat data tables). The majority of CEE systematic maps are stored and structured as flat data tables, mostly as spreadsheets. Tables are a simple, familiar, and robust means of structuring data. However, maintaining relationships within a 2-dimensional array of rows and columns can be challenging. This is because the only explicit relationships in a 2-dimensional array (single table), are between the attributes (columns) and the entities (rows). Any relationships which exist between columns/attributes in a table can only be inferred by the user (Figure 4). We found making such inferences a challenge when surveying systematic maps of research outside of our own fields of expertise (see Supplementary Table 3). The prior knowledge required to successfully navigate data relationships within tabular maps limits their accessibility for less specialized users.

A variety of techniques were employed by CEE maps for maintaining the relationships between attributes, and for allowing attributes to record multiple values. Of particular note were the practices of expanding columns to produce wide-form tables, and of housing multiple values within a single cell. Although expanding columns and/or housing multiple data entries in single cells do not threaten data integrity when applied to only 1 single attribute (see Thorn *et al.*, 2016, Supplementary Table 3), a loss of referential integrity was noted for maps implementing this practice for multiple attributes.

Such loss is illustrated in Figure 5, whereby column expansion (Figure 5A), and similarly multivalued cells (Figure 5B), falsely assert data relationships unrepresentative of the raw extracted data. Loss of referential integrity is acknowledged by Neaves *et al.* (2015), where the authors highlight falsely asserted interattribute relationships as a limitation of their mapping exercise.

The alternative strategy used by CEE systematic mappers when structuring data as a flat table was row expansion. Although advantageous for maintaining referential integrity, these long-form data structures can be challenging to process. They can create confusion for end-users interpreting what the study "unit" (entity) which constitutes a new row in the data table is (see Supplementary Table 3). Users must also be cautious of duplicates when querying specific data fields within the table. Duplicating data can also increase the risk of data-entry errors for systematic mappers tasked with manually populating a long-form table, resulting in inconsistencies.

In summary, the spreadsheet storage technology is an unsuitable long-term solution for EH SEMs, with wide-form tables potentially compromising data integrity, and long-form tables being impractical and/or error-prone.

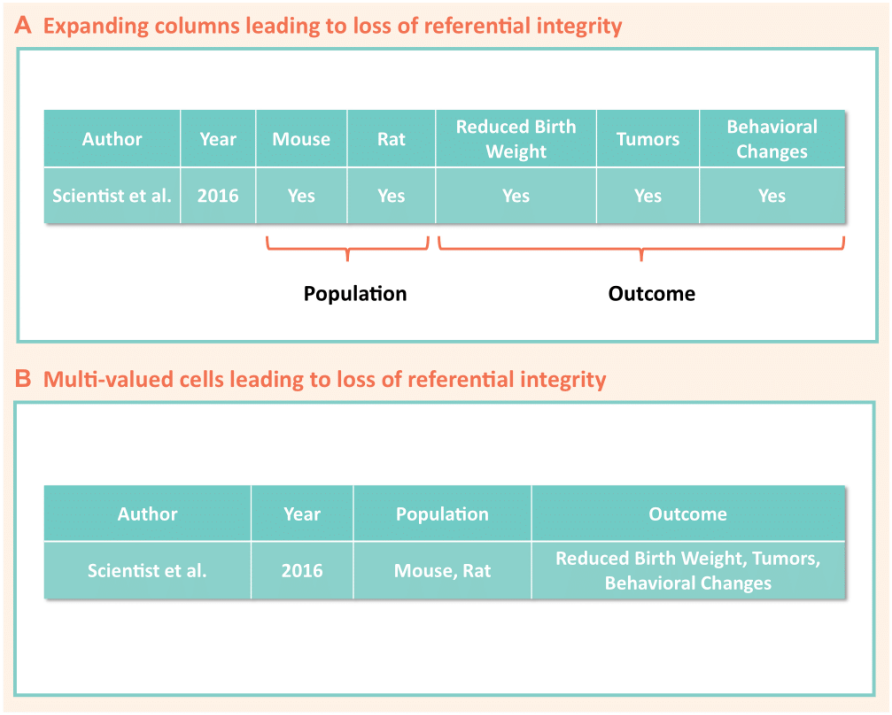

Figure 5. A, Loss of referential integrity resulting from the column expansion of more than 1 study attribute (data field). The recording of multiple populations and multiple outcomes on a single row compromises the ability of users to decipher *which* population was affected by *which* outcome. The table asserts that both populations (mice and rats) were affected by all 3 outcomes (reduced birth weight, tumors, and behavioral changes), respectively—which may not be truly representative of the raw data, compromising data integrity. B, This is similarly observed when multivalued cells are used for more than 1 study attribute.

Use of relational databases. Many of the discussed challenges associated with implementing systematic maps as flat data tables or spreadsheets are addressed by relational databases—the alternative storage technology identified in current systematic mapping practice (see Table 2). Relational databases divide entities into their own, referenceable tables—allowing links between related entities to be created and maintained. These links are coded into the database itself, and therefore do not rely on an end-user's implicit understanding of external conventions to correctly interpret.

The structure of a relational database is organized in an *on-write* schema, which is effectively a "blueprint" for the database (Karp, 1996); ie, the schema defines what constitutes an entity and therefore a data table, which attributes describe an entity, how an entity is related to other entities and therefore how data tables must reference others, all before data are stored. This necessitates a sound understanding of both the data to be stored in the database, and also the potential applications of the database. In fact, the optimization of end-users' capacity to query the database for a particular application is a key driver of schema design (Blaha *et al.*, 1988).

The "schema first, data later" (Liu and Gawlick, 2015) approach of relational databases requires a more detailed level of prior knowledge regarding the structure of the evidence and/or the applications of the database. This is problematic for EH SEMs for several reasons.

First, the potential applications of an EH SEM are varied. Even where a specific use case is known, an EH SEM should at least avoid restricting access to the evidence base for alternative uses. Second, SEM methodology advises against making decisions which are based on post hoc assessment of included studies (James *et al.*, 2016). However, without this assessment it is difficult to design a schema capable of housing all the entities and relationships likely to arise from the varied study designs and/or evidence streams collated through an EH SEM exercise. Even if this prior assessment were advocated by SEM methodology and did not lead to the introduction of bias or inconsistencies, there would likely be far too much data for mappers to feasibly consider in the design of an EH SEM's schema.

Third, SEMs are currently constructed by human mappers, who screen, assess, and extract data from 1 included study at a time. In this manner, mappers' understandings of the relationships between entities are limited to the level of the individual study. Thus, it can be difficult to design a schema able to appropriately account for relationships which occur at an interstudy level, compromising end-users' ability to query these relationships. For example, a one-to-many relationship between population and outcome entities may be appropriate at the level of the individual study, where a single population can be investigated for many outcomes. However, at the evidence-base level, a particular outcome may in fact have been reported by many studies, and therefore investigated in many different populations—making a many-to-many relationship between population and outcome, and a schema capable of representing this relationship, more appropriate. Alternatively consider the relationships between adverse outcomes along a causal pathway. Although a relationship between eg, Outcome A and Outcome C might become apparent at the evidence base level, mappers may only have access to relationships between eg, Outcome A and Outcome B, or Outcome B and Outcome C—which occur at the individual study level.

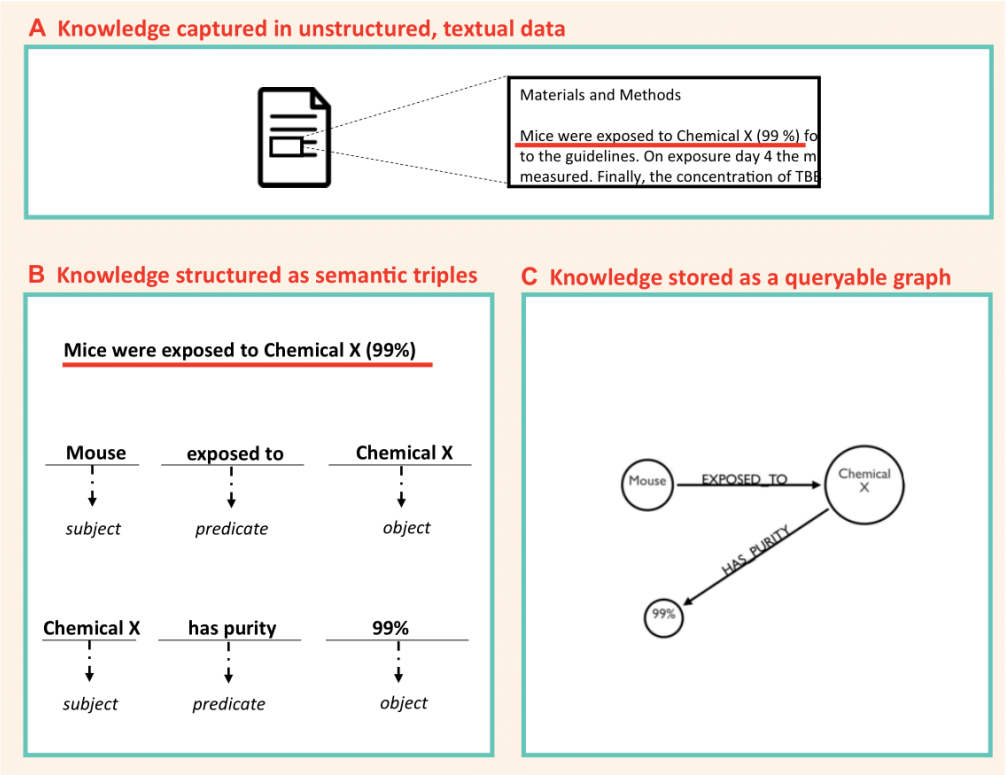

Figure 6. (A) Knowledge captured in unstructured, textual formats e.g. scientific articles, is distributed and programmatically inaccessible. (B) This knowledge can be structured in an intuitive and machine-readable way as a series of semantic subject-predicate-object triples – where entities are the subjects and/or objects and the relationships between entities are the predicates. (C) Entities can be stored as the nodes of a graph. The semantic value of the relationships between entities are preserved and stored as edges. The graph can continue to grow to produce a queryable representation of all knowledge on a topic (see Figure 7).

Finally, the growing volume and scope of EH data means that even if it were possible to devise a schema capable of accounting for all study designs that exist at present, new, and emerging study designs would soon out-date the schema, necessitating laborious, and potentially error-prone schema migration (Segaran *et al.*, 2009).

Avoiding these issues and attempting to balance the rigidity of a schema with the fluidity or heterogeneity of the data it organizes forces mappers to implement work-arounds (eg, compromising the resolution of SEMs), the likes of which might compromise the utility of SEMs for chemicals policy applications (see Supplementary File 1).

Overcoming the Limitations of Spreadsheets and RDBs: Knowledge Graphs for Mapping EH Evidence

Expanding and enriching the application of SEMs to varied EH research problems requires moving away from the rigidity of tabular data structures and their predefined relationships. Instead, SEMs in EH should utilize more flexible, *schemaless* data models and storage technologies. We believe this flexibility is offered by knowledge graphs and associated graph-based data storage technologies.

Knowledge graphs. The scientific knowledge codified in a study report can be readily formalized as a set of subject-predicate-object "triples." These triples can be stored as mathematical "graphs" (nodes and edges) where the nodes are the entities (subjects and objects) and the edges are the predicates, or relationships, between the subjects and the objects (see Figure 6). Because the graph is a direct representation of the semantic content of the studies being stored, it can be said to represent the knowledge captured in the study—hence "knowledge graph" (Ontotext, 2019b).

In graph database implementations, data are stored as nodes and relationships are stored as edges. Unlike the relational model, the graph model regards relationships as first-class entities, and keeps them alongside the values they connect. Rather than "artificially" creating relationships through cross referencing primary and foreign keys in data tables, graph databases natively *store* relationships, preserving their semantic value, and making them accessible to queries (Figure 6 and 7) (Robinson *et al.*, 2015). This is particularly valuable when the relationships underpinning data cannot be directly characterized *a priori*, or when the relationship between 2 pieces of information (nodes) can only be inferred through traversal of relationships which indirectly connect those nodes (Ontotext, 2019c) (eg, the inferred causal relationship between "Chemical X" and "Tumours" in Figure 7).

The graph model's flexibility and emphasis on relationships allows it to accommodate new developments in EH research. Data produced by studies of novel design can be incorporated among, and related to, preexisting data in the database without needing to update schema and subsequently migrate data (Robinson *et al.*, 2015). This is illustrated in Figure 7 which expands the amount of data populating the graph in Figure 6.

Knowledge graphs are already being exploited in other fields centered around the analysis of highly connected data (Ghrab

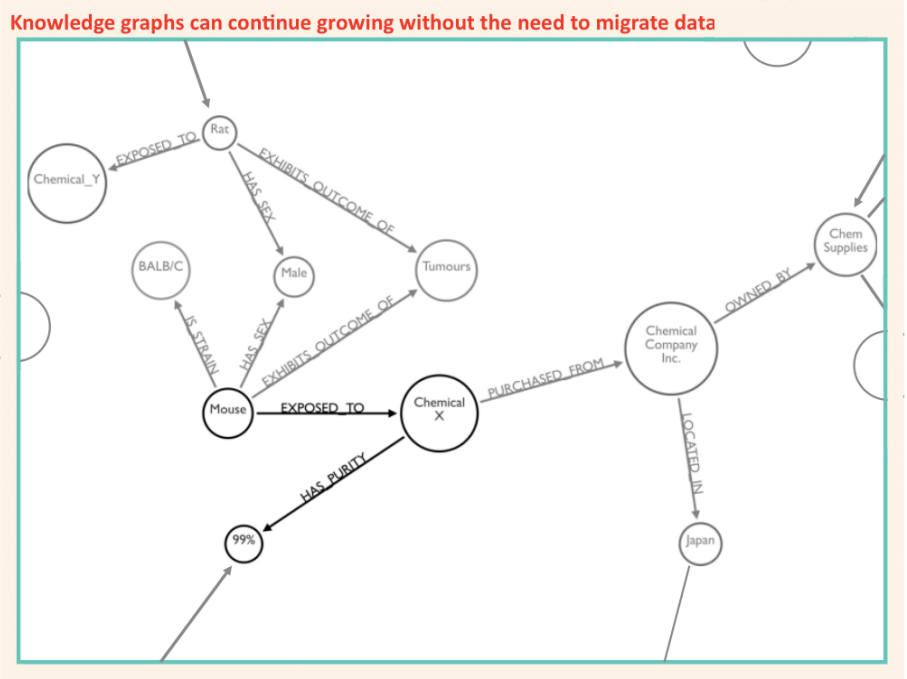

Figure 7. Storing relationships as first class entities allows knowledge graphs to continue to grow and expand without needing to revise schema and migrate data. This flexibility is particularly useful when relationships between entities cannot be characterised *a priori*.

et al., 2016). Notable use cases for graphs include: mapping complex networks of biological interaction (Aggarwal and Wang, 2010; Have and Jensen, 2013; Pavlopoulos *et al.*, 2011); representing chemical structures (Aggarwal and Wang, 2010); tracking communication and transaction chains for fraud detection (Castelltort and Laurent, 2016; Sadowski and Rathle, 2015); feeding recommendation engines for online retailers (Webber, 2018); facilitating highly customized outputs for social media platforms (Gupta *et al.*, 2013; Weaver and Tarjan, 2013); promoting a more proactive service from search engines (Singhal, 2012); and many more. The key commonality between these applications is the identification of trends or patterns of information that facilitate the generation of new knowledge that is actionable or of value to decision-making.

Schemaless data storage and data exploration. As relationships are *stored* as queryable, first-class entities—the schema which implicitly structures data begins to emerge naturally and can be discovered and exploited by knowledge finding applications *on-read* (Janković *et al.*, 2018; Kleppmann, 2017).

In CEE's current systematic mapping practice, trend exploration is predominantly reliant on filtering columns of a data table for specific values of interest. This requires that users are familiar with the structure of the database ie, they know which columns house values of interest, what those values of interest *are*, and that their interests align with the data model imposed by the tabular map. By comparison, graphs are amenable to some ambiguity in a user's query. Beyond the potential existence of an entity of interest, users do not require prior knowledge of the graph's structure, or the relationships connecting the entity of interest to others, to successfully gain an understanding of the graph space around that entity. This facilitates the building of data models which contextualize this understanding within a particular application.

In current systematic mapping practice, data models are closely tied to the data storage mechanism and its structure. Knowledge graphs do not fix data models *on*-write, separating data models from data storage—thus it is possible to apply multiple models to the same graph, optimizing access to the evidence base for a variety of interests and queries. Changes can also be readily incorporated into these data models without migrating the underlying data they access.

Ontologies. A key component of wider data modeling activities is the development of domain-specific ontologies. An ontology is an agreed upon and shared "conceptualization" of a domain (Dillon *et al.*, 2008), comprising a formal specification of terms used for describing knowledge and concepts within a domain and their relationships to each other, expressed through a standardized controlled vocabulary (Ashburner *et al.*, 2000; National Center for Biomedical Ontology, 2019). Developing domain-specific ontologies closely mirrors the coding step of systematic evidence mapping, which is designed to conceptualize the evidence base through organizing extracted data using a controlled vocabulary of terms.

In knowledge graph applications, ontologies are stored as data themselves (Noy and Klein, 2004)—forming an additional "layer" within the graph. Raw extracted data stored in the graph can be viewed as instances of an ontology's classes. By using data models to bind nodes of raw data to the nodes of a suitable ontology, users can navigate the evidence base through this ontology—but do not lose the ability to access the underlying raw data relevant to more highly resolved queries. Furthermore, maintaining a link between raw data and the controlled vocabulary code of a shared toxicological ontology serves to promote transparency, interoperability (Hardy *et al.*, 2012), and the development of training sets for machine-learning classifiers.

However, these concepts are underexplored in current evidence mapping practice where the majority of maps presented code in lieu of raw extracted data. This compromises transparency and limits users' ability to query data at variable resolution. In addition, coding vocabularies were rarely descriptive of the relationships that linked 1 term to another, with only 1 map organizing its code as a hierarchy of nested terms (Haddaway *et al.*, 2015). Where relationships between code were implied, this was generally stored in separate codebooks (ie, not as data within the database)—requiring users to consult a separate document for interpretation.

Other Lessons From Current Systematic Evidence Mapping Practice

Studying the key features of a systematic map database, ie, storage technology and the data structuring choices available for those technologies, highlights the need to pursue more flexible, schemaless approaches when adapting the methodology for EH. We have identified knowledge graphs as the technology capable of providing this flexibility. Although briefly covered in the above discussion, this survey identified additional aspects of current evidence mapping practice which are worthy of discussion.

Data accessibility, user-interfaces, and map documentation. A queryable database is the main, but not sole, output of mapping exercises. All CEE maps are accompanied by a study report which details methodology, presents key trends through data visualization, and/or describes further research needs. These accompanying reports can be thought of as documentation for their database products. In the context of software development, documentation is a formal written account of each stage of development and the effective use of the software for its intended application. It is an asynchronous means of communication between all involved stakeholders, including end-users and future developers, which transforms the tacit knowledge of developers into an explicit, exchangeable format (Ding *et al.*, 2014; Rus and Lindvall, 2002).

We found that, in general, the documentation of the maps was insufficient to make explicit the tacit knowledge of the map developers. This presented a barrier to successfully and efficiently querying the SEMs assessed in our survey. We observed that mappers' knowledge of their data model, database structure and intended uses for their database were generally underreported in accompanying SEM study reports. Discussion dedicated to instructing end-users on how they could or should interact with the database was particularly limited. This might compromise the ability of nonspecialist users to query SEMs for their own research interests. Similarly, trends visualized and analyzed in SEM study reports, which might serve as illustrative examples of how to interact with the SEM, were not accompanied by any documentation of the queries used to obtain the analyzed subset of evidence from the database—apart from 1 instance where the authors referred to code in GitHub, but the link was broken (Cheng *et al.*, 2019).

A more common practice for facilitating end-user access to trends in the evidence base was the development of interactive data visualization dashboards (Bernes *et al.*, 2015). Unlike their underlying databases, these dashboards were generally accompanied by documentation detailing how users could interact with the dashboard. This interaction was intuitive and required minimal technical expertise—with many dashboards adopting "point-and-click" functionality. However, interactive visualization dashboards should not be conflated with the systematic map database itself. These dashboards represent the visualized outputs of a set of predefined queries, where users can select which of the set to display. They can be thought of as user-interfaces which have been optimized for particular queries. However, users cannot devise and visualize customized queries through such dashboards. For this, access to the underlying database is required—reinforcing the need for its documentation.

Thus the role of high-quality software documentation in promoting transparency, growth, development and maintenance of SEMs as living evidence products should not be underestimated when adapting the methodology for EH.

Including database software capacity in evidence mapping teams. A final point of interest from this survey of current systematic mapping practice is that the multidimensionality of the relational database storage technology was not utilized in the CEE maps which employed the technology. This was evidenced by systematic maps which used a flat data structure even within a relational database software environment. Such maps included Neaves *et al.* (2015)—which presented a single, flat data table with expanded columns despite the authors' acknowledgment of the limitations of this structure and the capacity of the chosen storage technology to overcome them.

Reasons for implementing flat relational databases were unclear or unreported. However, facilitating the access of nonspecialist users to SEM outputs may have been a potential driver of this practice. Flat tables are associated with simple querying processes such as filtering columns, whereas relational databases require a more technically demanding process of constructing queries in structured query language (SQL). However, these concerns can readily be addressed by developing user-interfaces such as the visualization dashboards discussed above, and do not explain why inherently flat storage technologies, such as spreadsheets, were not used preferentially in such cases.

Thus, an alternative motivation for implementing flat relational databases might be a lack of familiarity with database storage technologies. This highlights a key challenge for adapting SEM methodology to EH, where subject specialists interested in mapping EH evidence may not have the necessary training to successfully implement graph-based storage. This underscores the value of comprehensive documentation—where the technical construction and querying of emerging maps might serve as training opportunities for others interested in the methodology. It also indicates the importance of developing these skills within mapping teams—where recruiting databasing specialists to SEM teams might be considered as important as recruiting statisticians to systematic review teams.

CONCLUSION

Systematic evidence mapping is an emerging methodology in EH. It offers a resource-efficient means of gaining valuable insights from a vast and rapidly growing evidence base. Its overarching aims, of organizing data and providing computational access to research, should facilitate evidence-based approaches to chemical risk assessment and risk management decision-making.

The methodology has been applied in the wider environmental sciences by the CEE. Characterizing the state-of-the-art of CEE systematic mapping practices offers valuable lessons for adapting the methodology for EH.

In particular, the rigid data structures which dominate current practice are ill-suited to the complex, heterogeneous and

highly connected data constituting the EH, and toxicology evidence bases. Flat data structures and those which are closely linked to predefined, on-write schema are optimized for a narrow range of specific use cases, which fits poorly with the much broader range of uses associated with chemicals policy workflows.

Successful adaptation of SEM methodology for EH would be accelerated by adopting flexible, schemaless database technologies in place of rigid, schema-first approaches. We have argued that knowledge graphs are 1 technological solution, which potentially provide an intuitive and scalable means of representing all of the connected, complex knowledge on a topic. Converse to the flat or relational databases favored by current practice, knowledge graphs store relationships between data as first-class entities, preserving their semantic value and making them accessible to queries. This ability to explore data through relationships or "patterns of information" does not require that users are familiar with a predefined data model or schema. This vastly expands the exploratory use cases of SEMs and even facilitates the discovery of new, previously uncharacterized relationships.

There are several readily available commercial and open-source graph database implementations (ArangoDB, 2019; Neo4j, 2019; Ontotext, 2019a; Stardog, 2019), and a variety of knowledge graph applications which demonstrate the power and utility of the graph data model and its inferencing capacity. Such resources are valuable for investigating the storage and exploration of SEMs as knowledge graphs and help to lower the entry barrier associated with familiarizing and training mappers in the use of a technology novel to the field.

SUPPLEMENTARY DATA

Supplementary data are available at *Toxicological Sciences* online.

ACKNOWLEDGMENTS

The authors would like to thank Mike Wolfe at Yordas Group for input helpful to the revision of this article.

FUNDING

T.A.M.W.'s PhD is financially supported by the Centre for Global Eco-innovation (a programe funded by the European Regional Development Fund) and Yordas Group, a global consultancy in the area of chemical safety, regulations, and sustainability. P.W.'s contribution to the manuscript was funded by the Evidence-Based Toxicology Collaboration at Johns Hopkins Bloomberg School of Public Health.

DECLARATION OF CONFLICTING INTERESTS

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

AUTHOR CONTRIBUTIONS

J.V. introduced the concept of graph databases, after which T.A.M.W., J.V., P.W., and C.H. established the principal ideas for the manuscript and developed an outline. T.A.M.W. wrote the first draft of the manuscript. T.A.M.W. and P.W. conducted the survey of CEE systematic maps. J.V. offered technical expertise and edited the discussion accordingly. N.H. contributed to the revision of the manuscript, offering regulatory and chemical risk assessment expertise. All authors reviewed and edited the manuscript and contributed to its development.

REFERENCES

Aggarwal, C. C., and Wang, H. (2010). *Managing and Mining Graph Data*, Vol. **40**. Springer, New York. https://doi.org/10.1007/978-1-4419-6045-0.

ArangoDB. (2019). *Graphs and ArangoDB*. Available at: https://www.arangodb.com/arangodb-training-center/graphs/. Accessed October 2019.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., *et al.* (2000). The Gene Ontology Consortium, Michael Ashburner1, Catherine A. Ball3, Judith A. Blake4, David Botstein3, Heather Butler1, J. Michael Cherry3, Allan P. Davis4, Kara Dolinski3, Selina S. Dwight3, Janan T. Eppig4, Midori A. Harris3, David P. Hill4, Laurie Is. *Nat. Genet.* **25**, 25–29.

Barra Caracciolo, A., de Donato, G., Finizio, A., Grenni, P., Santoro, S., and Petrangeli, A. B. (2013). A new online database on chemicals in accordance with REACH regulation. *Hum. Ecol. Risk Assess.* **19**, 1682–1699.

Bernes, C., Bullock, J. M., Jakobsson, S., Rundlöf, M., Verheyen, K., and Lindborg, R. (2017). How are biodiversity and dispersal of species affected by the management of roadsides? A systematic map. *Environ. Evid.* **6**, 1–16.

Bernes, C., Jonsson, B. U., Junninen, K., Lõhmus, A., Macdonald, E., Müller, J., and Sandström, J. (2015). What is the impact of active management on biodiversity in boreal and temperate forests set aside for conservation or restoration? A systematic map. *Environ. Evid.* **4**, 1–22.

Beverly, B. (2019). *Abstract 3267: Potential Alternatives to Systematic Review: Evidence Maps and Scoping Reviews*. Available at: https://www.toxicology.org/events/am/AM2019/program-details.asp. Accessed October 2019.

Blaha, M. R., Premerlani, W. J., and Rumbaligh, J. E. (1988). Relational database design using an object-oriented methodology. *Comput. Pract.* **31**, 414–427.

Castelltort, A., and Laurent, A. (2016). Rogue behavior detection in NoSQL graph databases. *J. Innov. Digit. Ecosyst.* **3**, 70–82.

Cheng, S. H., Macleod, K., Ahlroth, S., Onder, S., Perge, E., Shyamsundar, P., Rana, P., Garside, R., Kristjanson, P., McKinnon, M. C., *et al.* (2019). A systematic map of evidence on the contribution of forests to poverty alleviation. *Environ. Evid.* **8**, 1–22.

Clapton, J., Rutter, D., and Sharif, N. (2009). *SCIE Systematic Mapping Guidance April 2009*. Available at: https://www.scie.org.uk/publications/researchresources/rr03.pdf. Accessed October 2019.

Collaboration for Environmental Evidence. (2019a). *CEE Meetings*. Available at: https://www.environmentalevidence.org/meetings. Accessed October 2019.

Collaboration for Environmental Evidence. (2019b). *Completed Reviews*. Available at: http://www.environmentalevidence.org/completed-reviews. Accessed July 2019.

Collaboration for Environmental Evidence. (2019c). *Environmental Evidence: Systematic Map Submission Guidelines*. Available at: https://environmentalevidencejournal.biomedcentral.com/

submission-guidelines/preparing-your-manuscript/systematic-map. Accessed October 2019.

Commission of the European Communities. (2001). *White Paper: Strategy for a Future Chemicals Policy*, Vol. **13**. doi: 10.1007/BF03038641.

Cresswell, C. J., Wilcox, A., Randall, N. P., and Cunningham, H. M. (2018). What specific plant traits support ecosystem services such as pollination, bio-control and water quality protection in temperate climates? A systematic map. *Environ. Evid.* **7**, 1–13.

Dillon, T., Chang, E., Hadzic, M., and Wongthongtham, P. (2008). Differentiating conceptual modelling from data modelling, knowledge modelling and ontology modelling and a notation for ontology modelling. In *Conferences in Research and Practice in Information Technology Series*, 79.

Ding, W., Liang, P., Tang, A., and Van Vliet, H. (2014). Knowledge-based approaches in software documentation: A systematic literature review. *Inf. Softw. Technol.* **56**, 545–567.

ECHA. (2016). *Practical Guide How to Use Alternatives to Animal Testing to Fulfil Your Information Requirements for REACH Registration*.

Edwards, S. W., Tan, Y.-M., Villeneuve, D. L., Meek, M. E., and McQueen, C. A. (2015). Adverse outcome pathways—Organizing toxicological information to improve decision making. *J. Pharmacol. Exp. Ther.* **356**, 170–181.

Elmasri, R., and Navathe, S. B. (2013). The Relational Data Model and Relational Database Constraints. *Fundamentals of Database Systems*. Pearson Education, UK.

EPA. (2018). *Application of Systematic Review in TSCA Risk Evaluations*, pp. 1–247. Available at: https://www.epa.gov/sites/production/files/2018-06/documents/final_application_of_sr_in_tsca_05-31-18.pdf.

European Food Safety Authority. (2010). Application of systematic review methodology to food and feed safety assessments to support decision making[1]. EFSA Guidance for those carrying out systematic reviews. *EFSA J.* **8**, 1637.

Ghrab, A., Romero, O., Skhiri, S., Vaisman, A., and Zimányi, E. (2016). *GRAD: On Graph Database Modeling*. Available at: http://arxiv.org/abs/1602.00503. Accessed October 2019.

Gupta, P., Goel, A., Lin, J., Sharma, A., Wang, D, and Zadeh, R. (2013). *WTF: The Who to Follow Service at Twitter*, *WWW '13: Proceedings of the 22nd international conference on World Wide Web*, pp. 505–514. 10.1145/2488388.2488433.

Haddaway, N. R., Bernes, C., Jonsson, B.-G., and Hedlund, K. (2016). The benefits of systematic mapping to evidence-based environmental management. *AMBIO* **45**, 613–620.

Haddaway, N. R., Brown, C., Eales, J., Eggers, S., Josefsson, J., Kronvang, B., Randall, N. P., and Uusi-Kämppä, J. (2018a). The multifunctional roles of vegetated strips around and within agricultural fields. *Environ. Evid.* **7**, 1–43.

Haddaway, N. R., and Crowe, S. (eds) (2018). *Stakeholder Engagement in Environmental Evidence Synthesis*. Available at: http://www.eviem.se/Documents/projekt/2018/SRbookAll.pdf. Accessed October 2019.

Haddaway, N. R., Hedlund, K., Jackson, L. E., Kätterer, T., Lugato, E., Thomsen, I. K., Jørgensen, H. B., and Söderström, B. (2015). What are the effects of agricultural management on soil organic carbon in boreo-temperate systems? *Environ. Evid.* **4**, 1–29.

Haddaway, N. R., Macura, B., Whaley, P., and Pullin, A. S. (2018b). ROSES RepOrting standards for Systematic Evidence Syntheses: Pro forma, flow—Diagram and descriptive summary of the plan and conduct of environmental systematic reviews and systematic maps. *Environ. Evid.* **7**, 4–11.

Haddaway, N. R., Styles, D., and Pullin, A. S. (2014). Evidence on the environmental impacts of farm land abandonment in high altitude/mountain regions: A systematic map. *Environ. Evid.* **3**, 17–19.

Hardy, B., Apic, G., Carthew, P., Clark, D., Cook, D., Dix, I., Escher, S., Hastings, J., Heard, D. J., Jeliazkova, N., *et al.* (2012). Toxicology ontology perspectives. *ALTEX* **29**, 139–156.

Have, C.T., and Jensen, L. J. (2013). Databases and ontologies are graph databases ready for bioinformatics? *Bioinformatics* **29**, 3107–3108.

Weaver, J and Tarjan, P. (2013). Facebook linked data via the Graph API. *Semant. Web.* **4**, 245–250.

Hoffmann, S., de Vries, R. B. M., Stephens, M. L., Beck, N. B., Dirven, H. A. A. M., Fowle, J. R., Goodman, J. E., Hartung, T., Kimber, I., Lalu, M. M., *et al.* (2017). A primer on systematic reviews in toxicology. *Arch. Toxicol.* **91**, 2551–2575.

Hoffmann, S., and Hartung, T. (2006). Toward an evidence-based toxicology. *Hum. Exp. Toxicol.* **25**, 497–513.

Ingre-Khans, E., Ågerstrand, M., Beronius, A., and Rudén, C. (2016). Transparency of chemical risk assessment data under REACH. *Environ. Sci.: Process. Impacts* **18**, 1508–1518.

James, K. L., Randall, N. P., and Haddaway, N. R. (2016). A methodology for systematic mapping in environmental sciences. *Environ. Evid.* **5**, 7.

Janković, S., Mladenović, S., Mladenović, D., Vesković, S., and Glavić, D. (2018). Schema on read modeling approach as a basis of big data analytics integration in EIS. *Enterp. Inf. Syst.* **12**, 1180–1201.

Johnson, V., Fitzpatrick, I., Floyd, R., Simms, A. (2011). What is the evidence that scarcity and shocks in freshwater resources cause conflict instead of promoting collaboration? CEE review 10-010. Collaboration for Environmental Evidence, Bangor, UK.

Karp, P. D. (1996). Database links are a foundation for interoperability. *Trends Biotechnol.* **14**, 273–279.

Kleppmann, M. (2017). *Designing Data-intensive Applications*. O'Reilly Media, Inc, UK. Available at: https://www.oreilly.com/library/view/designing-data-intensive-applications/9781491903063/.

Leisher, C., Temsah, G., Booker, F., Day, M., Samberg, L., Prosnitz, D., Agarwal, B., Matthews, E., Roe, D., Russell, D., *et al.* (2016). Does the gender composition of forest and fishery management groups affect resource governance and conservation outcomes? A systematic map. *Environ. Evid.* **5**, 1–10.

Lewis, K. A., Tzilivakis, J., Warner, D. J., and Green, A. (2016). An international database for pesticide risk assessments and management. *Hum. Ecol. Risk Assess.* **22**, 1050–1064.

Liu, Z. H., and Gawlick, D. (2015). *Management of Flexible Schema Data in RDBMSs—Opportunities and Limitations for NoSQL in CIDR 2015*.

Lyndon, M. (1989). Information economics and chemical toxicity: Designing laws to produce and use data. *Mich. Law Rev.* **87**, 1795–1861.

Macura, B., Secco, L., and Pullin, A. S. (2015). What evidence exists on the impact of governance type on the conservation effectiveness of forest protected areas? Knowledge base and evidence gaps. *Environ. Evid.* **4**, 24.

Mandrioli, D., Schlünssen, V., Ádám, B., Cohen, R. A., Colosio, C., Chen, W., Fischer, A., Godderis, L., Göen, T., Ivanov, I. D., *et al.* (2018). WHO/ILO work-related burden of disease and injury: Protocol for systematic reviews of occupational exposure to dusts and/or fibres and of the effect of occupational exposure

to dusts and/or fibres on pneumoconiosis. *Environ. Int.* **119**, 174–185.

Martin, O. V., Geueke, B., Groh, K. J., Chevrier, J., Fini, J.-B., Houlihan, J., Kassotis, C., Myers, P., Nagel, S. C., Pelch, K. E., *et al.* (2018). Protocol for a systematic map of the evidence of migrating and extractable chemicals from food contact articles. 10.5281/zenodo.2525277.

Martin, P., Bladier, C., Meek, B., Bruyere, O., Feinblatt, E., Touvier, M., Watier, L., and Makowski, D. (2018). Weight of evidence for hazard identification: A critical review of the literature. *Environ. Health Perspect.* **126**, 076001–076015.

Mcintosh, E. J., Chapman, S., Kearney, S. G., Williams, B., Althor, G., Thorn, J. P. R., Pressey, R. L., McKinnon, M. C., and Grenyer, R. (2018). Absence of evidence for the conservation outcomes of systematicconservation planning around the globe: a systematic map. *Environ. Evid.* **7**, 22.

McKinnon, M. C., Cheng, S. H., Dupre, S., Edmond, J., Garside, R., Glew, L., Holland, M. B., Levine, E., Masuda, Y. J., Miller, D. C., *et al.* (2016). What are the effects of nature conservation on human well-being? A systematic map of empirical evidence from developing countries. *Environ. Evid.* **5**, 1–25.

National Center for Biomedical Ontology. (2019). *BioPortal*. Available at: https://bioportal.bioontology.org/. Accessed October 2019.

Neaves, L. E., Eales, J., Whitlock, R., Hollingsworth, P. M., Burke, T., and Pullin, A. S. (2015). The fitness consequences of inbreeding in natural populations and their implications for species conservation—A systematic map. *Environ. Evid.* **4**, 1–17.

Neo4j. (2019). *Neo4j*. Available at: https://neo4j.com/. Accessed October 2019.

Noy, N. F., and Klein, M. (2004). Ontology evolution: Not the same as schema evolution. *Knowl. Inf. Syst.* **6**, 428–440.

NTP-OHAT. (2019). *About the Office of Health Assessment and Translation*. Available at: http://ntp.niehs.nih.gov/. Accessed October 2019.

Oliver, S., and Dickson, K. (2016). Policy-relevant systematic reviews to strengthen health systems: Models and mechanisms to support their production. *Evid. Policy* **12**, 235–259.

Ontotext. (2019a). Ontotext *GraphDB*. Available at: https://www.ontotext.com/products/graphdb/. Accessed October 2019.

Ontotext. (2019b). *What is a Knowledge Graph?* Available at: https://www.ontotext.com/knowledgehub/fundamentals/what-is-a-knowledge-graph/. Accessed October 2019.

Ontotext. (2019c). *What is Inference?* Available at: https://www.ontotext.com/knowledgehub/fundamentals/what-is-inference/. Accessed October 2019.

Pavlopoulos, G. A., Secrier, M., Moschopoulos, C. N., Soldatos, T. G., Kossida, S., Aerts, J., Aerts, J., Schneider, R., and Bagos, P. G. (2011). Using graph theory to analyze biological networks. BioData Min. **4,** 1–27.

Pelch, K. E., Bolden, A. L., and Kwiatkowski, C. F. (2019). Environmental chemicals and autism: A scoping review of the human and animal research. *Environ. Health Perspect.* **127**, 046001.

Pelch, K. E., Reade, A., Wolffe, T. A. M., and Kwiatkowski, C. F. (2019). PFAS health effects database: Protocol for a systematic evidence map. *Environ. Int.* **130**, 104851.

Pool, R., and Rusch, E. (2014). *Identifying and Reducing Environmental Health Risks of Chemicals in Our Society: Workshop Summary*. The National Academies Press, Washington, DC.

Randall, N. P., Donnison, L. M., Lewis, P. J., and James, K. L. (2015). How effective are on-farm mitigation measures for delivering an improved water environment? A systematic map. *Environ. Evid.* **4**, 18.

Randall, N. P., and James, K. L. (2012). The effectiveness of integrated farm management, organic farming and agri-environment schemes for conserving biodiversity in temperate Europe—A systematic map. *Environ. Evid.* **1**, 4.

Rhomberg, L. R., Goodman, J. E., Bailey, L. A., Prueitt, R. L., Beck, N. B., Bevan, C., Honeycutt, M., Kaminski, N. E., Paoli, G., Pottenger, L. H., *et al.* (2013). A survey of frameworks for best practices in weight-of-evidence analyses. *Crit. Rev. Toxicol.* **43**, 753–784.

Robinson, I., Webber, J., and Eifrem, E. (2015). *Graph Databases*. O'Reilly Media, New York. doi: 10.1016/B978-0-12-407192-6.00003-0.

Rus, I., and Lindvall, M. (2002). Knowledge management in software engineering. *IEEE Softw.* **19**, 26–38.

Sadowski, G., and Rathle, P. (2015). Fraud Detection: Discovering Connections with Graph Databases. Neo Tehcnology. Available at: https://neo4j.com/use-cases/fraud-detection/. Accessed October 2019.

Schultz, T. W., Cronin, M. T. D., Walker, J. D., and Aptula, A. O. (2003). Quantitative structure-activity relationships (QSARS) in toxicology: A historical perspective. *J. Mol. Struct.: THEOCHEM* **622**, 1–22.

Segaran, T., Evans, C., and Taylor, J. (2009). *Programming the Semantic Web: Traditional Data-modeling Methods*. O'Reilly Media, USA.

Sexton, K., and Hattis, D. (2007). Assessing cumulative health risks from exposure to environmental mixtures—Three fundamental questions. *Environ. Health Perspect.* **115**, 825–832.

Singhal, A. (2012). *Introducing the Knowledge Graph: Things, Not Strings*. Available at: https://googleblog.blogspot.com/2012/05/introducing-knowledge-graph-things-not.html. Accessed October 2019.

Sola, P., Cerutti, P. O., Zhou, W., Gautier, D., Iiyama, M., Shure, J., Chenevoy, A., Yila, J., Dufe, V., Nasi, R., *et al.* (2017). The environmental, socioeconomic, and health impacts of woodfuel value chains in Sub-Saharan Africa: A systematic map. *Environ. Evid.* **6**, 1–16.

Stardog. (2019). *Stardog*. Available at: https://www.stardog.com/. Accessed October 2019.

The Endocrine Disruption Exchange. (2019). *TEDX Publications*. Available at: https://endocrinedisruption.org/interactive-tools/publications/. Accessed October 2019.

Thorn, J. P. R., Friedman, R., Benz, D., Willis, K. J., and Petrokofsky, G. (2016). What evidence exists for the effectiveness of on-farm conservation land management strategies for preserving ecosystem services in developing countries? A systematic map. *Environ. Evid.* **5**, 1–29.

United States Environmental Protection Agency. (2016). *The Frank R. Lautenberg Chemical Safety for the 21st Century Act*. Available at: https://www.epa.gov/assessing-and-managing-chemicals-under-tsca/frank-r-lautenberg-chemical-safety-21st-century-act/. Accessed October 2019.

Vandenberg, L. N., Ågerstrand, M., Beronius, A., Beausoleil, C., Bergman, Å., Bero, L. A., Bornehag, C.-G., Boyer, C. S., Cooper, G. S., Cotgreave, I., *et al.* (2016). A proposed framework for the systematic review and integrated assessment (SYRINA) of endocrine disrupting chemicals. *Environ. Health* **15**, 1–19.

Vinken, M., Whelan, M., and Rogiers, V. (2014). Adverse outcome pathways: Hype or hope? *Arch. Toxicol.* **88**, 1–2.

Walker, V. R., Boyles, A. L., Pelch, K. E., Holmgren, S. D., Shapiro, A. J., Blystone, C. R., Devito, M. J., Newbold, R. R., Blain, R., Hartman, P., *et al.* (2018). Human and animal evidence of potential transgenerational inheritance of health effects: An evidence map and state-of-the-science evaluation. *Environ. Int.* **115**, 48–69.

Webber, J. (2018). *Powering Real-time Recommendations With Graph Database Technology Powering Real-time*. Neo4j.

Wolffe, T. A. M., Whaley, P., Halsall, C., Rooney, A. A., and Walker, V. R. (2019). Systematic evidence maps as a novel tool to support evidence-based decision-making in chemicals policy and risk management. *Environ. Int.* **130**, 104871.

World Health Organization. (2019). *Framework for Use of Systematic Review Methods in Chemical Risk Assessment—Authors Meeting*. Available at: http://who.int/ipcs/events/SRmeeting_US/en/. Accessed October 2019.

Zynda, M. (2013). The first killer app: A history of spreadsheets. *Interactions* **20**, 68–72.

Supplemental File 1

A survey of systematic evidence mapping practice and the case for knowledge graphs in environmental health & toxicology

Taylor A. M. Wolffe[1,2,*], John Vidler[3], Crispin Halsall[1], Neil Hunt[2], Paul Whaley[1,4]

Affiliations:

[1]Lancaster Environment Centre, Lancaster University, Lancaster, LA1 4YQ, UK

[2]Yordas Group, Lancaster University, Lancaster, LA1 4YQ, UK

[3]School of Computing and Communications, Lancaster University, Lancaster, LA1 4WA, UK

[4]Evidence-Based Toxicology Collaboration, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, USA

Correspondence:

*t.wolffe@lancaster.ac.uk

Key words:

systematic evidence map, knowledge graph

Supplemental Discussion

Working around rigid schema and lowering resolution

Attempting to balance the rigidity of a schema with the fluidity or heterogeneity of the data it organizes might lead environmental health (EH) mappers to compromise the resolution of their data extraction and coding. Consider a set of heterogeneous *in-vitro* studies which meet the inclusion criteria of a systematic evidence map (SEM) exercise. Avoiding extracting specific study detail in favour of more broadly applicable study features (e.g. type of cell tested) allows those studies to occupy the same schema and avoids the need to update the schema, as all future encountered studies will likely contain the broad extracted study feature. Coding heterogeneous studies with a broad controlled vocabulary term can have a similar effect if this code is provided in lieu of raw data– e.g. broadly coding both a study investigating the contaminants in drinking water, and another investigating air quality as "environmental monitoring".

A preference for producing lower resolution maps was noted in the survey of CEE's current mapping practice, where only broad or even censored data were included in systematic map databases (e.g. Gumbo et al., 2018). The majority of maps also appeared to provide only the broader controlled vocabulary code in lieu of the raw extracted data to which the code was applied.

Beyond allowing heterogeneous or complex data to fit within a rigid structure, several additional motivators might contribute to the current preference for producing low resolution

systematic maps. Lower resolution maps represent a less severe demand for time and resources - making their outputs more achievable given the currently manual nature of systematic mapping. Mappers may also wish to prevent end-users from drawing inappropriate or premature conclusions based on data which has not been critically appraised by censoring specific results or data relationships. Alternatively, a lower resolution map may already be fit-for-purpose, and thus the most efficient and easily understood form of mapping exercise.

However, there are several drawbacks associated with the practices that result in low resolution maps. Transparency is reduced when end-users cannot access the raw data to which a code has been applied, and therefore cannot assess whether they agree with the application of that code. Data integrity is compromised when the relationships between broad coding categories are unrepresentative of the raw data, or when incorrect relationships are inferred between broad data fields. Accessibility of the data for end users wishing to query the systematic map is also limited to a narrower range of broad questions, restricting application of the map to the use-cases defined by the developer rather than meeting the potentially unanticipated needs of the user. Thus, while a SEM which facilitates identification of low-resolution trends might be an efficient research tool in other fields, the demands of chemicals policy for detail and contextual value limit the utility of these exercises for this application.

Finally, although low resolution maps might represent a smaller upfront cost in terms of time and resources – they may represent a less efficient approach in the long-term. Details omitted from a map which later become important or relevant for updated chemical risk assessment procedures means that data extraction efforts must be repeated. Similarly, although extracting and storing high-resolution semantic triples in a knowledge graph may incur higher demands on time/resources in the short term, the preservation of referential integrity in the graph means that the mapping exercise need not be repeated in order to facilitate a user's access to, and understanding of, the data.

References

Gumbo, D., Clendenning, J., Martius, C., Moombe, K., Grundy, I., Nasi, R., … Petrokofsky, G. (2018). How have carbon stocks in central and southern Africa ' s miombo woodlands changed over the last 50 years ? A systematic map of the evidence. *Environmental Evidence*, 1–19. https://doi.org/10.1186/s13750-018-0128-0

Chapter 4: A graph-based approach to mapping human exposure-outcome associations for chemical contaminants

This chapter is pending submission to the journal *Toxicological Sciences*. **Electronic supplementary materials are available for this chapter.**

The candidate's contribution was structuring own and co-authors' ideas as a manuscript; planning the analysis; screening and extracting data; populating a graph database with extracted data; querying graph database for trends; writing the manuscript for supervisor review.

Candidate: ..... Date: 30/01/2020
Ms. Taylor A. M. Wolffe

Supervisor: ..... Date: 30/01/2020
Prof. Crispin J. Halsall

A graph-based approach to mapping human exposure-outcome associations for chemical contaminants

Taylor A. M. Wolffe[1,2,*], Crispin Halsall[1], Paul Whaley[1,3]

Affiliations:

[1]Lancaster Environment Centre, Lancaster University, Lancaster, LA1 4YQ, UK

[2]Yordas Group, Lancaster University, Lancaster, LA1 4YQ, UK

[3]Evidence-Based Toxicology Collaboration, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, USA

Correspondence:

*t.wolffe@lancaster.ac.uk

Key words:

systematic evidence map, graph database, NHANES

1. Abstract

The National Health and Nutrition Examination Survey (NHANES) is a large, cross-sectional human biomonitoring program in the United States (US). Among the variables measured in the survey are biomarkers of exposure to hundreds of manufactured chemicals which are of interest to chemical risk assessment and chemicals policy applications. The NHANES datasets are publicly available and offer a unique opportunity for crowd-sourcing analysis efforts. This allows researchers with varied interests to uncover a broad range of toxicologically relevant associations between survey variables and ensures maximum return of the resources input into the survey. However, these analyses are typically published in the academic literature in unstructured formats. This makes it difficult to gain a broad overview of which associations have been studied, and whether there are any potential links between such associations which might inform future analyses. This is an issue which has traditionally been addressed through scoping review. Limitations in the outputs of scoping reviews make them difficult to update and compromise their broader utility for characterizing and exploring existing research. In this manuscript, we explore the future of such evidence-surveillance exercises by conducting a small-scale, graph-based systematic evidence mapping exercise, in which literature reporting exposure-outcome associations for the NHANES datasets are mapped. We highlight the efficacy of the graph data model for preserving data integrity of increasingly complex and highly resolved datasets – contrasting our approach to an equivalent scoping exercise. Finally, we outline the research and development required to conduct such graph-based exercises at scale.

2. Introduction

The National Health and Nutrition Examination Survey (NHANES) is a large, continuous, cross-sectional biomonitoring program in the United States. Every year, a representative sample of the US population are recruited to participate in questionnaires, interviews, physical examinations and/or biological sampling (Centers for Disease Control and Prevention, 2019a; Sobus et al., 2015). In addition to biomarkers of nutrition, health and communicable disease - NHANES measures biomarkers for hundreds of manufactured chemical exposures in samples of urine, whole blood, plasma or serum (Centers for Disease Control and Prevention, 2019b). These measures, along with questionnaire, interview and examination data, are made publicly available through the Centers for Disease Control and Prevention (CDC)'s website (https://www.cdc.gov/nchs/nhanes/index.htm).

This accessible resource of human exposure data is of significant value to regulatory decision-making in chemicals policy contexts, and has been described as a "a gold-mine of data for environmental health analyses" by the United States Environmental Protection Agency (US EPA) (EPA, 2003), with whom the CDC has a collaborative relationship. Analyses of the NHANES datasets are relevant to an array of health (Ahluwalia et al., 2016) and chemicals policy tasks, including: setting national reference levels for health-related variables and/or chemical exposures (e.g. CDC, 2001); monitoring trends in health-variables (e.g. disease prevalence) and/or chemical exposure (EPA, 2003); assessing the efficacy of policy interventions to control chemical exposure (e.g. through phase outs of toxic substances (Easthope & Valeriano, 2007)) or health-outcomes (e.g. through vaccination programmes (Markowitz et al., 2013); identifying disparities in the exposure/health variables associated with specific sub-populations (e.g. Kobrosly et al., 2012; Nelson et al., 2012; Tyrrell et al., 2013); and assessing risk-factors for health-outcomes.

The open access model of NHANES presents a unique opportunity for "crowd-sourcing" analyses. This means that researchers from across the globe are able to access and analyse NHANES data, finding trends and associations of potential interest to federal agencies and the wider research community. Thus, the potential returns of the survey are maximised - improving resource efficiency. Further, the open accessibility of NHANES promotes collaboration and progression within environmental health, where pooled expertise builds a more complete understanding of the data and the statistical methods required for its analysis. At minimum, the open accessibility of NHANES promotes transparency - whereby analyses can be verified through independent replication efforts (e.g. Brown et al., 2019).

However, analyses of NHANES datasets are not always as accessible as the raw data themselves - compromising the potential reach, impact and benefits of these crowd-sourced efforts. Difficulty accessing analyses may result in inadvertent but redundant duplication, which threatens the efficient allocation of resources to investigating associations of novel or emerging concern. For chemicals policy, a lack of accessibility further compromises consideration of such analyses in the risk assessment process - where associations are integrated with data from a range of heterogeneous evidence streams.

A central, searchable resource which catalogues research conducted using NHANES data would therefore maximise its value for stakeholders. Such a resource would facilitate the meta-research required for identifying trends across analyses, characterising research gaps on which to focus crowd-sourcing efforts. Several reviews have addressed this need for monitoring and understanding the research space around the analysis of NHANES data (e.g. (Bell & Edwards, 2015; Taboureau & Audouze, 2017). One such review of particular relevance to chemicals policy is that of Sobus et al. (2015) - which focused on analyses concerning chemical exposures. Broad, "scoping" reviews such as Sobus et al. (2015) characterise the

research landscape through a process of searching for, screening, extracting and categorizing evidence. While providing a valuable overview of research activity, the static and inaccessible outputs of such reviews (e.g. in-text data tables and visualisations) continue to limit their broader utility. In other words, users are unable to query collated evidence according to their specific research interests and must instead re-extract and re-structure data as appropriate to their queries. These are the shortcomings which systematic evidence mapping aims to overcome (James et al., 2016; Wolffe et al., 2019).

Systematic evidence mapping is an evidence-based methodology of growing interest in environmental health and toxicology (e.g. Beverly, 2019; Pelch et al., 2019), with wide potential application in regulatory workflows (Wolffe et al., 2019). It builds on the scoping review methods traditionally employed in evidence surveillance - with an emphasis on transparency, robustness and a broad, comprehensive coverage of the evidence landscape. The key output of a systematic evidence mapping exercise i.e. the systematic evidence map (SEM) itself, takes the form of a queryable database of references, extracted data and meta-data. This computationally accessible output can be readily updated without duplication of data-extraction effort.

In our previous work (Wolffe et al., 2020), we highlighted the utility of the flexible, schemaless graph data model for maintaining transparency and data integrity within SEMs. However, the application of graphs for evidence mapping in environmental health is still novel. To resolve a path toward graph-based approaches to evidence mapping in environmental health, we conduct an exploratory case-study in which we apply a graph-based approach for mapping exposure-outcome associations reported for the NHANES dataset - expanding on the outputs of Sobus et al.'s (2015) scoping exercise.

In addition to our findings regarding exposure-outcome associations - we highlight key advantages of the graph data model for evidence surveillance and systematic evidence mapping methodology and discuss the challenges and future work required to implement our approach at scale. Through this case-study, we aim to increase familiarity of the evidence synthesis community with the graph data model. We hope this bridging research will accelerate and unify efforts to make best use of existing data by better understanding the needs of evidence-surveillance and the computational tools required to meet those needs.

3. Methods

3.1 Aims

The primary aim of this mapping exercise is to conduct a methodological exploration of the graph-data model for systematic evidence mapping in a context relevant to chemicals regulation, and to compare this approach with traditional methods of evidence surveillance (e.g. scoping reviews). In using NHANES as a case study - the secondary aim of this mapping exercise is to explore which of the exposures and outcomes measured as part of NHANES have been investigated for association by the wider research community, and to identify the future research required to study these associations at scale.

3.2 Dataset

The complete set of 273 publications included by Sobus et al. (2015) in their scoping review on the use of NHANES data for chemical risk assessment were considered for inclusion in this mapping exercise. The search strategy and inclusion/exclusion criteria which generated this dataset can be found in the Methods and Supplemental Information of Sobus et al (2015).

Briefly, search strings covering the concepts of “NHANES”, the “United States”, and “biomonitoring” were combined with the “AND” boolean operator and filtered by publication dates which fell within the range 1999-2013. Results were screened by a single author (Sobus et al., 2015) at the level of title and abstract, and studies which exclusively reported use of “endogenous biomarkers (e.g. hormones, antibodies, inflammatory markers), tobacco-specific biomarkers (e.g. cotinine), dietary biomarkers (e.g. vitamins/nutrients, essential minerals), or biomarkers of phytoestrogens, isoflavonoids, or aflatoxin” excluded.

3.3 Inclusion Criteria

The full text of each of the 273 publications in the Sobus et al. (2015) dataset were screened for inclusion in this evidence mapping exercise by a single reviewer (TW). Only those which presented a statistical measure of association between a health outcome (i.e. a biological response or markers of biological response) and a chemical exposure were included.

In their scoping review, Sobus et al. (2015) categorised each of the included publications according to whether they reported a “health association” or “exposure assessment”. Any discrepancies regarding the inclusion status of publications in this mapping exercise compared to the category which these publications were assigned by Sobus et al. (2015) were documented and justified (see Table S1) i.e. indicating if a publication which was categorised as “exposure assessment” has been included, or conversely if a publication which was categorised as “health outcome” has been excluded in this mapping exercise.

3.4 Low Resolution Data Extraction

A simple data extraction workflow was developed to assist with the collation of exposure-outcome associations from included publications. Briefly, data extracted from included study

reports were divided into one of two categories: bibliographic information pertaining to the study report; or information pertaining to the studied associations. A relational database infrastructure was used to construct data extraction forms - whereby study reports were related to associations through a one-to-many relationship. This allowed the recording of multiple associations without manually repeating the bibliographic information in long-form, minimising extraction errors.

Data recorded in the "bibliographic information" component of the data extraction workflow were as follows:

- Reference ID (as assigned by Sobus et al., to facilitate comparison and validation)
- Title
- Authors
- Publication Year

Associations were defined as occurring between a chemical exposure and a health outcome - for which the results of a statistical measure of association were reported. Entities for which an association with chemical exposure was measured - but which represent exposures (including to other chemicals), non-health outcomes, covariates, adjustment factors or stratification variables (e.g. sex, age, smoking status etc.) were excluded. Likewise, associations in which neither entity was considered a chemical exposure (e.g. history of anaemia and head circumference) were excluded from extraction. Thus, data recorded in the "associations" component of the data extraction workflow were as follows:

- Chemical exposure
- Individual components of the chemical exposure, if applicable (e.g. ΣPFAS might comprise individual components of PFOA, PFOS and PFBS).
- Biological medium in which the exposure was measured

- Associated health outcome

Data were extracted by a single reviewer (TW) from the full-text pdf of each publication. As far as possible, data were extracted in a consistent manner by using consistent spellings and structures e.g. exposures reported as “PFOA”, or “perfluorooctanoic acid” were both extracted in the format “Perfluorooctanoic acid (PFOA)”.

3.5 Data Processing and Graph-based storage

A large, long-form data table (Table S2) was produced by querying the relational data extraction infrastructure. Chemical constituents were delimited as separate “repeating columns” (see Wolffe et al. 2020), and biological media concatenated with chemical exposure - so as to distinguish between e.g. Blood Cadmium and Urine Cadmium etc. Any exposures for which a biological medium was unreported were also extracted and represented as an independent exposure i.e. “Cadmium”, “Cadmium, blood” and “Cadmium, urine” were considered as different specific exposures. Any inconsistencies identified within each column (e.g. typos etc.) were manually amended.

A graph data model was devised for representing the relationships between entities in the long-form data table, and is presented in Figure 1. The data within the table was processed for storage using an iPython Jupyter Notebook (Project Jupyter, 2019) (see Supplementary File S1) according to the graph data model in Figure 1. The py2neo package (Small, 2019) was used to connect with a Neo4j graph database instance (Neo4j, 2019) and the graph populated with data as described in Supplementary File S1.

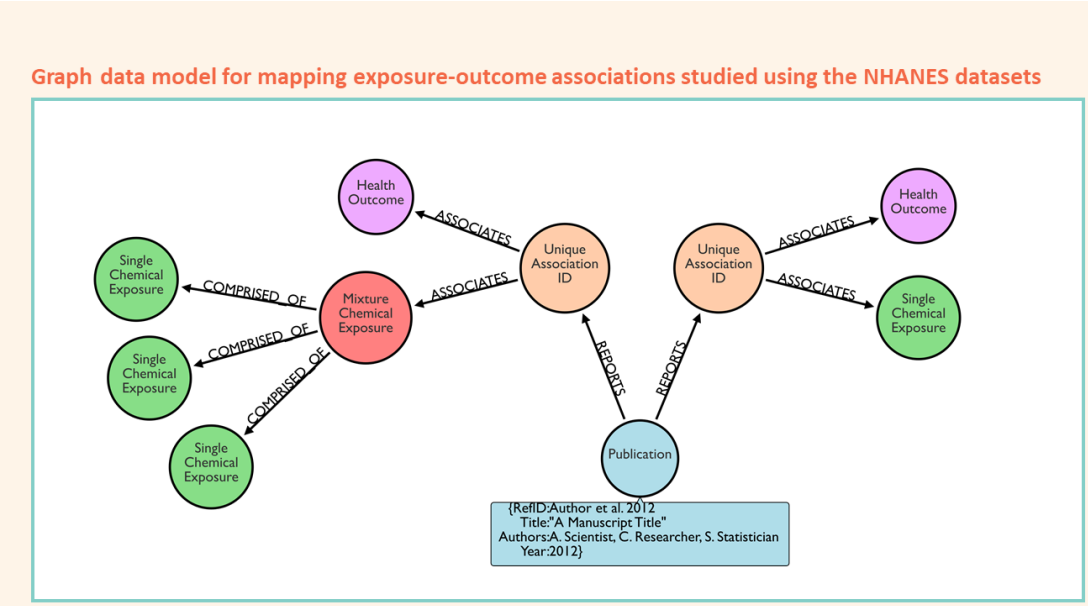

Figure 1: Graph data model describing the relationships between publications and the exposure-outcome associations they report. Neo4j's labelled property graph model was exploited to represent "RefID", "Title", "Authors" and "Year" as properties of Publication nodes.

3.6 Applying Controlled Vocabulary Code

10 of the 11 controlled vocabulary terms used to categorise studies by Sobus et al. were adopted in this mapping exercise i.e."BFRs" (brominated flame retardants), "Dioxins, furans, PCBs" (polychlorinated biphenyls), "environmental phenols", "metals/metalloids", "other", "PAHs" (polycyclic aromatic hydrocarbons), "pesticides", "PFCs" (perfluorinated compounds), "phthalates" and "VOCs" (volatile organic compounds) . As individual chemical exposures were extracted in this exercise, the "multi-group" controlled vocabulary term was omitted. Where possible - the use of controlled vocabulary by Sobus et al. was mapped directly (see Supplementary File S2, Table S3), except for the exposures which would solely have been categorised as "multi-group". These exposures were manually re-assigned a controlled vocabulary label (see Supplementary File S2, Table S4 and Table S5). Code was incorporated

into the graph as nodes and connected to single/mixed chemical exposure nodes through a "CODED_AS" relationship (see Fig 2A). Similarly, a controlled vocabulary of 18 terms was developed iteratively for the categorisation of health outcome nodes. These terms were manually assigned to extracted health outcomes and represented in the graph with a "CODED_AS" relationship (see Fig. 2B). Where appropriate, more than one controlled vocabulary term was assigned to a single health outcome (e.g. "Mortality, cancer" was labelled with both the terms "Mortality" and "Cancer").

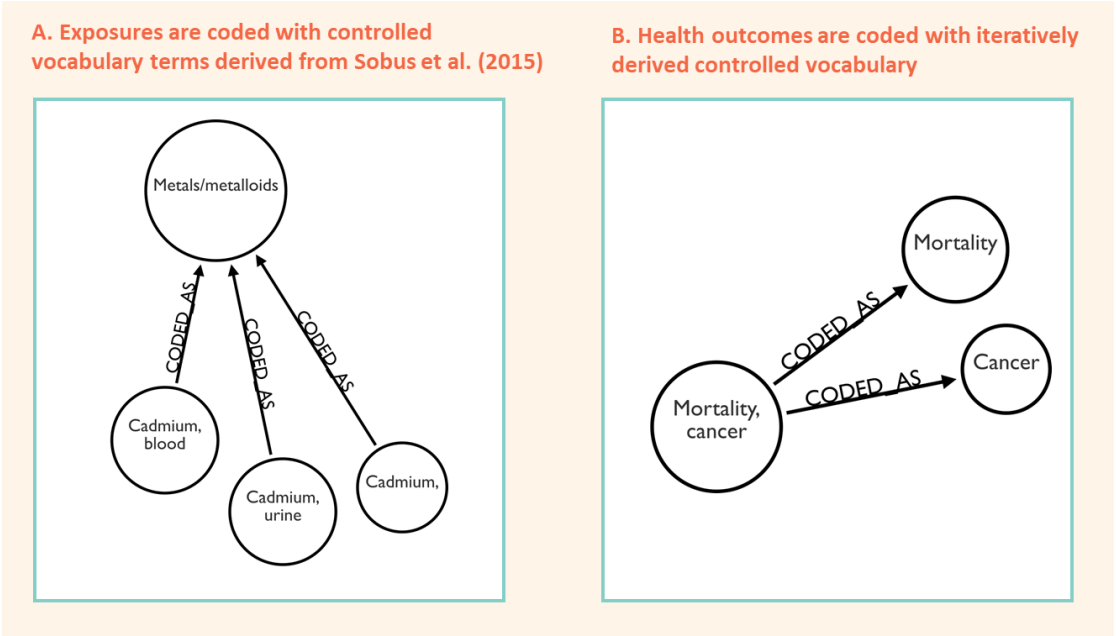

Figure 2: (A) Chemical exposure nodes are categorised by a "CODED_AS" relationship to nodes housing controlled vocabulary terms. (B) Similarly, health outcome nodes are categorised by a "CODED_AS" relationship to nodes housing controlled vocabulary terms. Due to the variation and complexity of health outcomes, multiple controlled vocabulary terms may be used to categorise a single outcome.

3.7 Exploring Associations

The evidence map was explored through a series of queries written in cypher (neo4j's graph query language) using the py2neo package (see Supplementary File S3). The results of these queries were typically processed as pandas dataframes and visualised using a variety of tools, including python data visualization packages (e.g. seaborne, matplotlib etc.) and Tableau.

4. Results

4.1 Included Publications and Number of Associations

In total, we extracted 1656 investigated associations from 132 included publications. These associations encompassed 326 different chemicals and 265 specific health outcomes. The number of associations reported within a single publication ranged between 1 and 150, with a median value of 4.

4.2 Exposures, Health Outcomes and Associations

4.2.1 Exposures

We found that the largest number of associations included in our map could be categorised as occurring between metals/metalloids and a health effect (see Fig. 3). A total of 86/132 included publications reported at least one association between a health outcome and a chemical exposure within the metals/metalloids group (Supplementary File S3). Blood lead and urinary cadmium were the two most frequently associated specific chemical exposures within this group (Supplementary File S3 & S4). The metals/metalloids exposure category had

over twice as many studied associations as the next most populous chemical group; Dioxins, Furans, PCBs - which were reported in only 9/132 included publications.

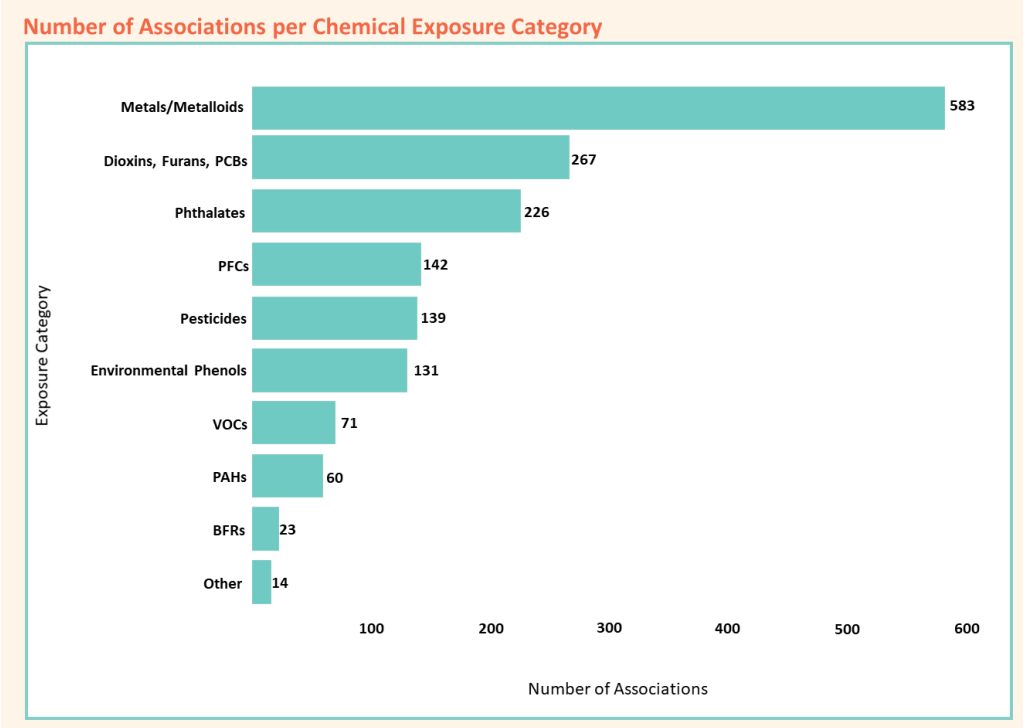

Figure 3: The number of associations (x axis) which investigated a chemical exposure within each of the exposure categories (y axis), across all included publications.

However, "Dioxins, Furans, PCBs" was also the most diverse exposure group, comprised of 110 distinct chemical exposures (see Supplemental File 3). Thus, this exposure group was characterised by a low frequency of associations for many individual chemical constituents. The full make-up of each exposure category, and the frequency with which each distinct chemical within an exposure category is associated with a health outcome, is visualised in Supplementary File S4. "Dioxins, Furans, PCBs" was also the category most frequently assigned to associations which investigated mixed chemical exposures, followed closely by phthalates (Supplementary File S3). The number of individual chemical constituents comprising a mixed chemical exposure for any category ranged from 2 e.g. for "PCB-196 & PCB-203, serum (Cave

et al. 2010)" to 28 e.g. for "Non dioxin-like polychlorinated biphenyls (PCBs), serum (Gallagher et al. 2013a)" (Supplementary File S3).

4.2.2 Health Outcomes

The most frequently associated health outcome category was the "Body Weight and Metabolism" group (Fig 4.), which incorporated 61 specific health outcomes (Supplementary File S3 & S5) and was reported by 34/132 included publications (Supplemental File 3). A full break-down of the specific health outcomes associated with each outcome category, and the frequency of associations for each specific outcome can be found in Supplementary File S5.

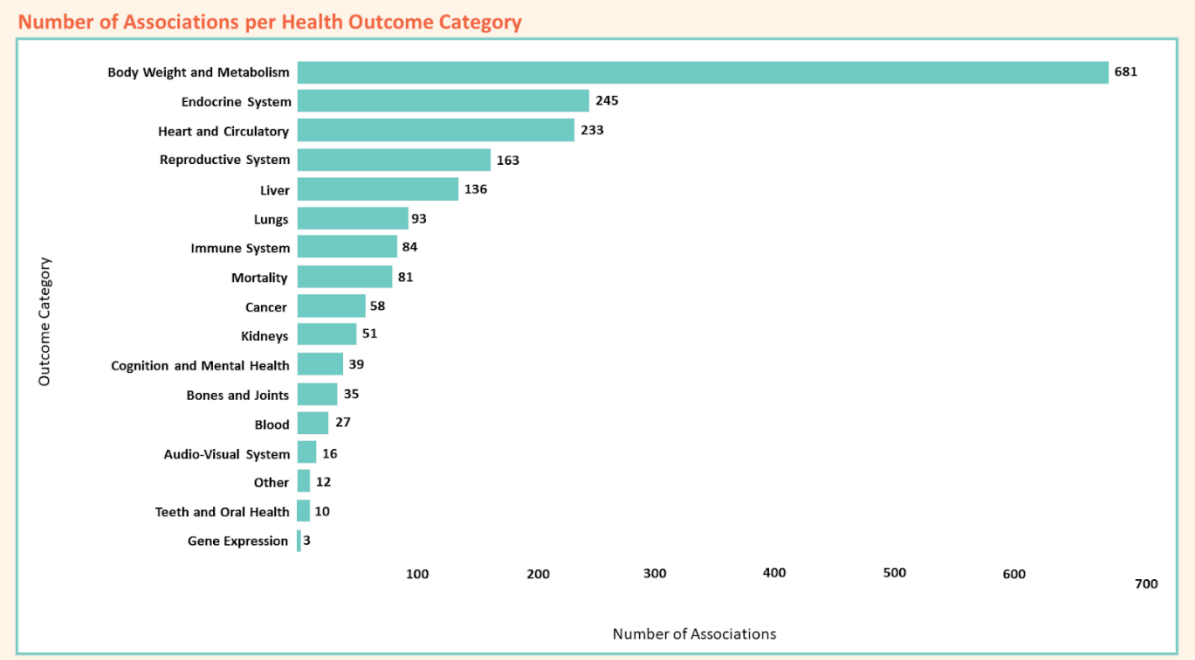

Figure 4: The number of associations (x axis) which investigated a health outcome within each of the outcome categories (y axis), across all included publications.

"Body weight and metabolism" was also the most diverse outcome category, with nearly twice as many specific health outcomes coded with the term as the next most diverse group; "Mortality" (see Supplementary File S3).

4.2.3 Associations

Figure 5A illustrates which exposure and outcome categories were most frequently associated with each other, and which were not studied for association at all. Associations between “metals/metalloids” and “body weight/metabolism” were the most prevalent, followed closely by “dioxins, furans, PCBs” and “body weight/metabolism”. “Metal/metalloids” was the only exposure group to have been associated (at least once) with each of the health outcome categories. Similarly, “body weight/metabolism” and “liver” were the only health categories to have been associated (at least once) with each of the exposure categories. Figure 5B illustrates how these associations are distributed across individual publications - indicating that associations between “metals/metalloids” and “heart and circulatory” outcomes were independently reported in the largest number of publications. The implications for analysing trends at the publication vs. association level are further discussed below.

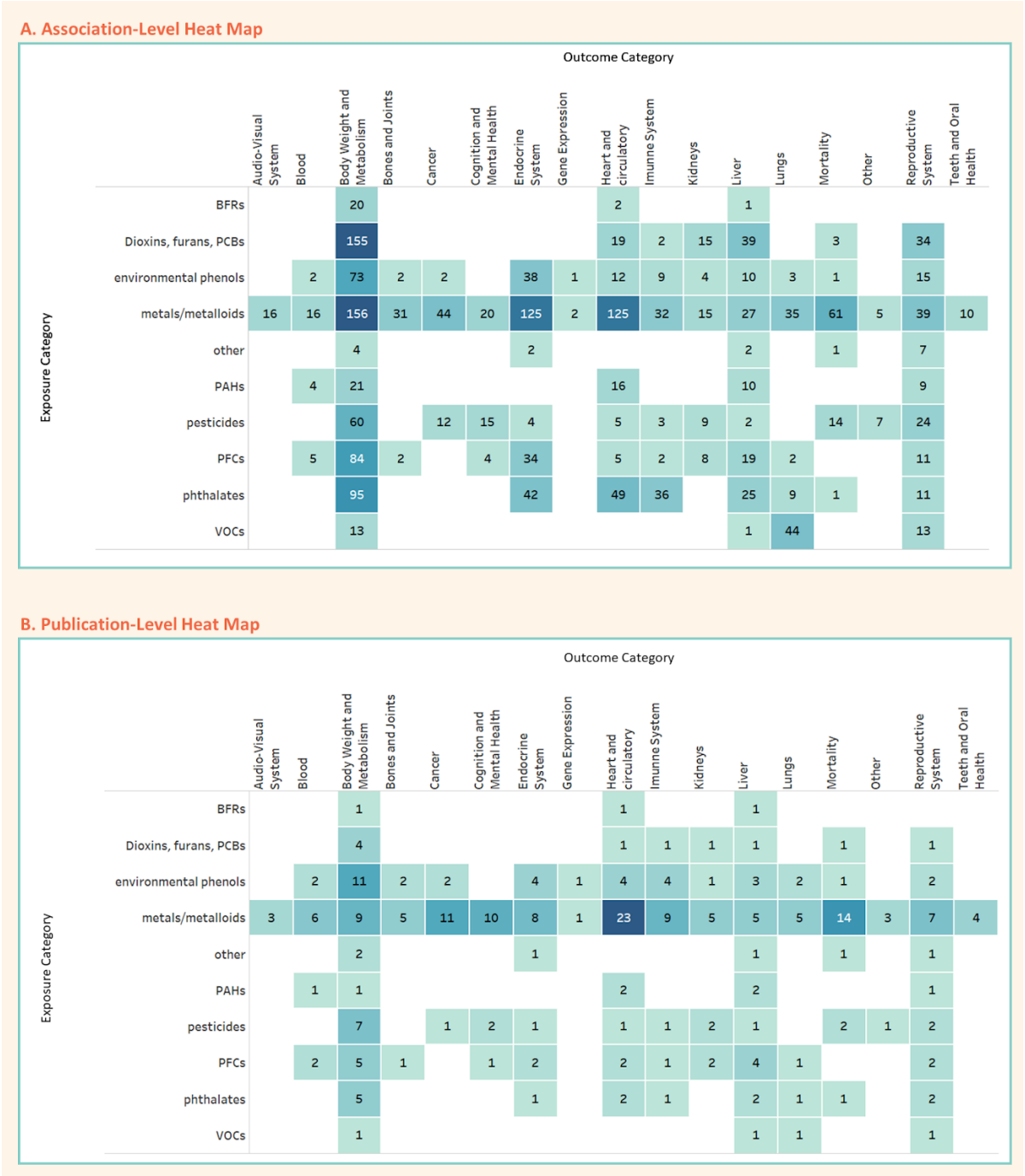

Figure 5: (A) Association-level heat map illustrating the frequency with which exposure and outcome categories have been associated, across all included publications i.e. a single publication reporting associations between “cadmium, blood” and “lipid levels” as well as “lead, blood” and “lipid levels” would count as two associations between “metals/metalloids” and “body weight/metabolism”. (B) Publication-level heat map illustrating the frequency with which independent publications report associations between

exposure and outcome categories i.e. a single publication reporting associations between "cadmium, blood" and "lipid levels" as well as "lead, blood" and "lipid levels" would count only once as reporting associations between "metals/metalloids" and "body weight/metabolism".

5. Discussion

This case-study mapping exercise expanded on the scoping review of Sobus et al (2015) by increasing resolution through extracting additional detail regarding exposure-outcome pairs for each included publication reporting a health outcome. In this exercise we did not present the direction, significance or statistical methods used to study associations - nor did we appraise the methodological integrity of the included studies. Thus, the presence of an association in this map is not indicative of a positive association or causative relationship between exposure and outcome - but simply of the fact that a study has investigated a relationship between those two variables.

Inclusion of specific detail regarding results of included studies within systematic evidence maps is an issue for debate in the field - and has led to the practice of censoring evidence maps (e.g. (Gumbo et al., 2018). This censorship is borne out of a responsibility to ensure that data collated within an evidence map are not misinterpreted or misused, as validation and appraisal of included evidence is often beyond the scope of the evidence surveillance function served by SEMs. Such issues raise questions over the validity of exposure-outcome associations for NHANES datasets given limitations in the survey design and/or analytical methods employed in assessing associations (e.g. Christensen et al., 2014; Stone & Reynolds, 2003). However, the goal of evidence mapping is to make best use of all available data by improving computational access to an evidence landscape such that data *can* be critically analysed. It presents a neutral,

queryable account of what has been done or investigated within a field, regardless of whether and how that field "should" have conducted its investigations. Thus, although it may be appropriate in some cases (e.g. the production of user-interfaces for non-specialist audiences) censoring should not necessarily present a barrier to the resolution of evidence-maps.

Disregarding censoring, limitations in the resolution of evidence maps are imposed by several other factors e.g. the prevalent use of rigid data structures which struggle to uphold referential integrity as the complexity of data increases with resolution (Wolffe et al., 2020). We discuss our simple and accessible exploration of the graph data model for handling this increase in complexity - focusing on the application and potential of graph-based methods for facilitating the production of highly resolved evidence maps. We compare our findings to scoping methods where appropriate, highlighting the remaining challenges which threaten resolution of evidence maps and the future work required to address these challenges through the lens of further expanding this NHANES mapping exercise.

5.1 Mapping vs scoping exposure-outcome associations for the NHANES datasets

In finding that "metals/metalloids" was the most prevalent exposure group, our results echo that of Sobus et al (2015), even without taking into consideration the publications which did not study a health outcome. This raises the question of fitness for purpose of evidence surveillance exercises, as the scoping methods employed by Sobus et al. are sufficient to broadly determine the most dominant features of the evidence landscape and are suited to characterising evidence at the *publication* level.

However, this publication level assessment of the evidence base did illicit some results which were ambiguous, e.g. publications which fall into the "multi-group" category. Additionally, our results show the potential for considerable differences in trends evaluated at the publication versus association level (Figure 5). This is because a single publication can be broadly categorised as being about one thing (e.g. metals/metalloids), but actually present multiple results (e.g. for five different metals) which are uniquely relevant to the context of chemical risk assessment, and the manner in which substances are currently regulated on a single chemical-by-chemical basis. Thus, even minimally increasing resolution, as in our evidence map, is likely to increase the value of the evidence surveillance exercise for regulatory applications (e.g. Wolffe et al, 2019).

Presenting mapped associations as a computationally accessible and queryable output (rather than as a static data table within a pdf) also has advantages over traditional scoping methods of evidence surveillance. Static data tables or visualizations may be valuable and fit for the purpose of identifying trends and specific evidence gaps, but they limit the range of questions which can be asked of the collated data - requiring interested users to re-conduct data extraction efforts should they wish to explore the presented trends in further detail, or from alternate angles. Thus, ensuring that the underlying data is computationally accessible expands the utility of evidence surveillance exercises. This, in combination with increased resolution meant that we were able to identify evidence gaps in an equivalent manner to Sobus et al.'s scoping review (e.g. finding a lack of investigated associations between "PAHs" and "cancer"), but were also able to query the data to learn/infer more about the evidence landscape. For example, even if completely ignorant to the chemistry of included exposures - uncovering the fact that the "Dioxins, Furans, PCBs" group was most frequently studied as a mixture of chemicals - and having access to the constituents of those mixtures, allows users

to learn something about the potential similarities and/or detection of these chemicals - which may be of significance for their regulation.

5.2 Exploring application of a graph-based approach to systematic evidence mapping

Although our mapping exercise only minimally increased the resolution of extracted data, the subsequent increase in complexity and connectedness of the underlying data model was significant and began to present issues for representation in flat data structures. This can be illustrated by comparing the structure of the flat data table which housed the pre-processed raw data (Table S2) to the graph data model which stored this data. The flat data table contained a combination of expanding rows and columns, where only the authors (as producers of the map), are cognizant of the relationships between the attributes and entities housed in various rows and columns. Contrastingly, the graph data model (Fig. 1) makes these relationships explicit to end-users and maintains referential integrity.

We opted to use Neo4j's (community edition) graph database implementation for this initial exploration of a graph-based approach to evidence-mapping due to its accessibility and the availability of resources designed for non-technical audiences to familiarise themselves with the graph database (Robinson et al., 2015; Sasaki et al., 2018). Neo4j implements a labelled-property graph model, whereby nodes can be assigned labels and properties e.g. labels of "HealthOutcome", "SingleChemicalExposure" etc. were applied to nodes in our evidence map, and properties of "Title", "Authors", "Year" etc. assigned specifically to the publication nodes in our evidence map. We found this graph data model amenable to manual evidence mapping efforts, utilising labels to facilitate categorising and querying the evidence base. Populating and querying the graph was intuitive, but did require some technical knowledge in the form of the cypher querying language. This technical knowledge indicates a key barrier to the wider

uptake of graph-based approaches for evidence mapping in environmental health. However, a growing volume of resources (Robinson et al., 2015; Sasaki et al., 2018), tools designed for non-specialists (e.g. Neo4j, 2020) and increasing computational literacy of evidence mappers will aide overcoming this barrier.

We found a more severe barrier to the graph-based approach in attempting to balance the potential for higher resolution with the limitations set by a manual data extraction process. In the initial planning phase of this exercise, we hoped to expand our definition of an "association" as "occurring between any two entities for which the results of a statistical measure/model of associated were reported" - including assessments between e.g. stratified and non-stratified populations, adjusted and non-adjusted models etc. as independent associations and extracting detail on the statistical approach, significance and direction of the associations. However, this dramatically increased the manual data extraction burden of the exercise e.g. Dye et al. (2002) which reports 4 associations in our lower resolution map would report 59 associations according to this expanded definition. A lack of sufficient time and resources meant that we were unable to pursue such high-resolution mapping manually. This challenge is likely reflected by the limited resource availability of chemicals policy workflows. This highlights the need for more automated approaches to data screening and especially extraction if the full potential of high-resolution graph-based evidence maps are to be realised.

Preserving resolution and facilitating automation appears well aligned with the storage of data as semantic triples within a knowledge graph, discussed in our previous work (see Wolffe et al., 2020). This is because all information expressed with language within a publication can be captured as a series of subject-predicate-object triples, where subjects and objects occupy nodes of a graph, and predicates form the relationships which connect subjects and objects.

The designation of subjects, predicates and objects is dependent on rules and conventions inherent to the structure of language - thus mappers need not manually "choose" information from a publication for extraction. Instead, all plain text within a publication can be parsed into sets of semantic triples using natural language processing (Rusu et al., 2014). Although such parsers are available in other fields (e.g. Gangemi et al., 2009) we are unaware of their successful application in evidence mapping or environmental health contexts. It can also be argued that distilling unstructured text into a series of semantic triples is not a data extraction task, but is the data standard to which publications should adhere in the first instance.

However, with higher resolution and automated extraction workflows comes increased complexity for graph outputs. Much of this complexity may be noise - i.e. information which is irrelevant to the interests of users wishing to query the evidence landscape for a particular application. This makes accessing trends within the graph and/or discovering underlying data models more challenging than in our lower resolution, manually produced exercise. This is where binding data to ontologies is vital for distilling data relevant to a particular domain, and for characterising the evidence space around a domain. However, this is still an area of active research - and another aspect of the graph-based approach limited by a greater demand for technical knowledge.

5.3 Expanding this evidence mapping exercise

Pending further advances and applications of the automated approaches required to facilitate high resolution knowledge graphs, it remains important to continue exploration of graph-based approaches to evidence mapping as a means of upholding data integrity. To this end, there is significant scope to expand and improve the NHANES mapping exercise presented in this manuscript.

While data extraction workflows are still largely manual, ensuring the process is conducted in duplicate by two independent mappers will help to protect against human error and improve the accuracy and consistency of extracted information. Additionally, expanding the search strategy used to find relevant publications e.g. by removing restrictive terms such as those relating to specific biomarkers ("urine", "blood" etc.), and updating the date-range of the search to encompass current literature, will ensure a more comprehensive coverage of the evidence-landscape within the map. Extracting even minimal additional detail from each of the included studies would also serve to increase the resolution and utility of the map. Similarly, extracting exposure-exposure associations from the exposure studies (excluded for the purposes of this mapping exercise) will begin to facilitate more complex graph queries, e.g. where a path from an exposure source such as personal care products can be traversed to a health outcome through related biomarkers - even if the exposure source and the health outcome were not reported within the same study. However, screening and extracting a larger, updated dataset in duplicate will incur the same challenges regarding manual workflows and resource availability as discussed above.

Even without further screening or data extraction, there is scope for expanding the utility of the evidence map. Incorporating a relevant ontology will help to further categorise and organise data, such that the map can be queried against the topics of interest to a particular domain (e.g. cancer biology). This will also ensure that the map is interoperable and will facilitate the incorporation of data from sources beyond the publications collated in the mapping exercise, adding greater contextual value to the interpretation of trends. For example, incorporating evidence from environmental monitoring studies into the map of NHANES associations may begin to elucidate potential exposure pathways.

There is also scope to better characterise the use (or gaps in the use) of NHANES data by developing an ontology derived from the NHANES datasets themselves. A potential workflow for the development of such an ontology is briefly outlined in Fig. 6. Incorporating all data variables available in the NHANES datasets into an ontology will facilitate the identification of specific variables which might be available for analysis but have been under-utilised in studies of association. Maintaining relationships between specific variables and the NHANES survey cycle in which those variables were studied will facilitate detailed exploration and inferencing of how trends have changed over time, e.g. allowing fast identification of whether a sudden increase in associations studied for a particular chemical is due to a corresponding sudden availability of data within a new NHANES survey cycle, or whether such trends can be attributed to other factors. Ensuring that NHANES variables are related through a hierarchy of terms will facilitate further, variably resolved querying of included associations e.g. if "Cadmium, blood" and "Cadmium, urine" are categorised as "Cadmium" before "metals/metalloids", trends can be analysed at three levels of resolution. This may be particularly useful for chemical substances which are currently grouped in very broad categories e.g. "Dioxins, furans, PCBs".

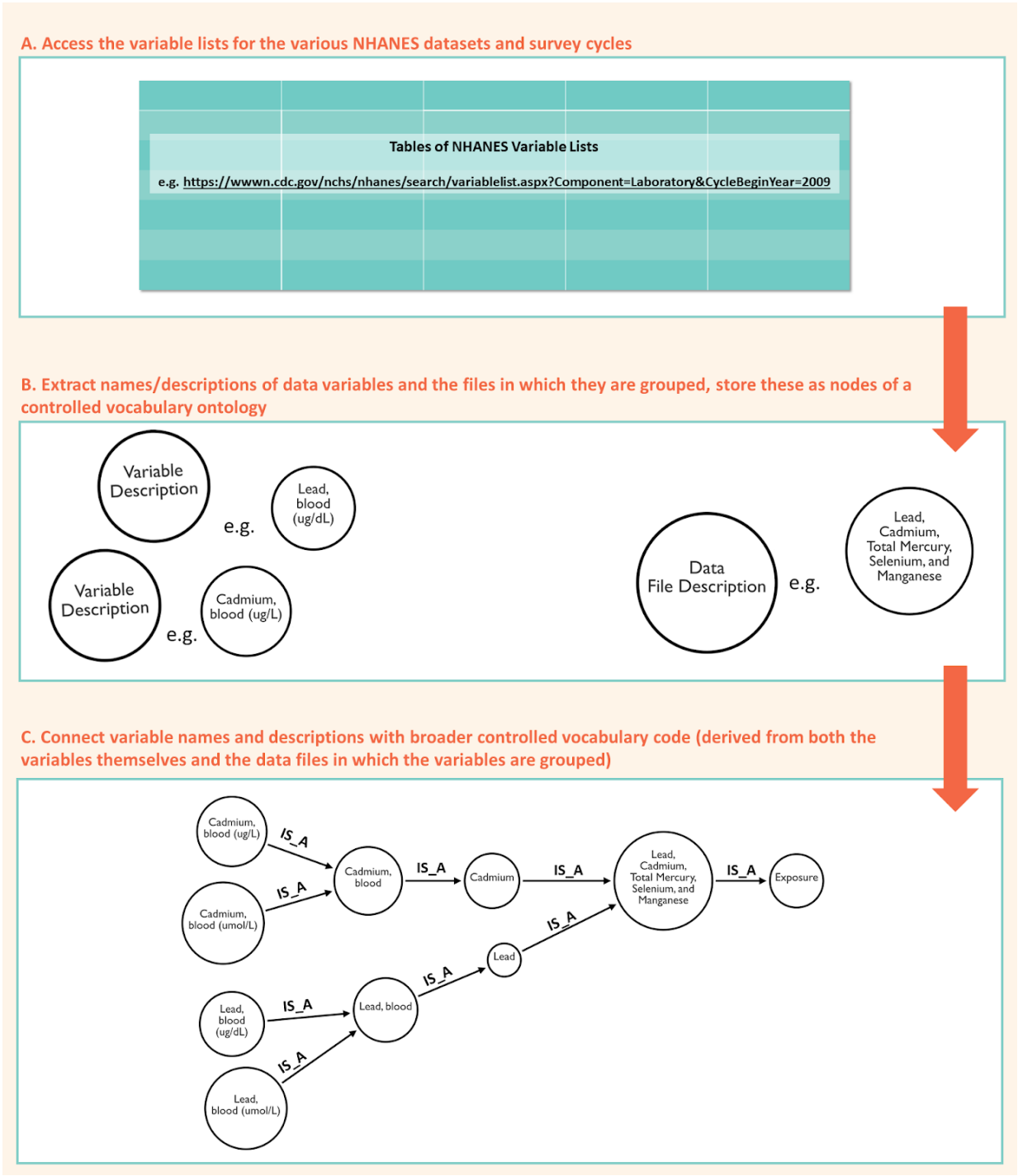

Figure 6: Brief outline of a potential workflow for devising a controlled vocabulary ontology which describes the availability of data within the NHANES datasets. Variables which are more specific to particular survey cycles (e.g. dates) might also be incorporated into the organisational structure of the ontology.

Incorporating health related variables from the NHANES datasets would prove similarly beneficial for the resolution of queries and interpretation of identified trends. We noticed that the health outcomes reported in many included studies were not explicitly reported as NHANES variables, but were defined by authors based on several more specific NHANES variables. For example, Muntner et al. (2005) defined "hypertension" as "...based on the average of all available blood pressure measurements, hypertension was defined as systolic or diastolic blood pressure of at least 140 mm Hg or 90 mm Hg, respectively, and/or self-reported current use of blood pressure– lowering medication." Thus, maintaining a link between the individual variables within the NHANES datasets which constitute a defined health outcome will improve insights into the use of NHANES data, and will allow the appropriateness of these uses to be appraised.

6. Conclusion

In this manuscript, we presented an exploration of the implementation of graph-based approaches to evidence mapping using a context of relevance to decision-making in environmental health, and a dataset accessible to others wishing to learn from, or further expand this work. The graph data model is a flexible and intuitive means of maintaining data integrity when extracting, storing and querying increasingly complex, higher resolution datasets. It has significant potential application for evidence surveillance within regulatory workflows - and when coupled with SEM methodology, offers greater transparency and reusability than current scoping approaches.

However, our exploration of the application of graphs to current evidence mapping workflows identified two key challenges on which to focus future work;

1. Although the graph data model is arguably more intuitive than flat, or relational data models - graph-based approaches demand a greater level of technical expertise to implement. This may present a barrier for evidence mappers who are unfamiliar with the programming and querying languages required for successful implementation.
2. Graphs are capable of upholding referential integrity for complex and highly connected datasets and facilitate highly resolved queries. However, the manual nature of data extraction within SEM methodology limits the resolution and complexity of the datasets to be stored within a graph- preventing graphs from being exploited to their full potential.

Research into the automation of evidence synthesis workflows is ongoing (van Altena et al., 2019; Connor et al., 2019; Marshall & Wallace, 2019) - and will facilitate the production of large and informative graph datasets. In the interim, it is vital to continue increasing familiarity with graph-based approaches and their associated data standards. Continued research into the application of graphs for evidence surveillance will allow independently conducted, manual mapping efforts to be amalgamated. This will facilitate a deeper understanding not only of the toxicological evidence landscape, but of the methods required to implement evidence mapping at scale.

7. Acknowledgements

The authors would like to thank Jon Sobus for helpful correspondence during the formulation of this mapping exercise. We also thank Matthew Barnes, senior library assistant at Lancaster University, for help with retrieval of publications for inclusion in this exercise.

8. Funding Sources

TW's PhD is financially supported by the Centre for Global Eco-Innovation (a programme funded by the European Regional Development Fund) and Yordas Group, a global consultancy in the area of chemical safety, regulations and sustainability. PW's contribution to the manuscript was funded by the Evidence-Based Toxicology Collaboration at Johns Hopkins Bloomberg School of Public Health.

9. Author Contributions

TW established the principle ideas for the mapping exercise. All authors refined the scope of the mapping exercise according to resource availability. TW screened studies at full text, extracted data, populated and queried the graph database, and wrote the first draft of the manuscript. All authors reviewed and edited the manuscript and contributed to its development.

10. References

Ahluwalia, N., Dwyer, J., Terry, A., Moshfegh, A., & Johnson, C. (2016). Update on NHANES Dietary Data : Focus on Collection , Release , Analytical Considerations , and Uses to Inform Public Policy 1 , 2, (9). https://doi.org/10.3945/an.115.009258.status

Bell, S., & Edwards, S. (2015). Identification and Prioritization of Relationships between Environmental Stressors and Adverse Human Health Impacts, *123*(11), 1193–1199.

Beverly, B. (2019). Abstract 3267: Potential Alternatives to Systematic Review: Evidence

Maps and Scoping Reviews. Retrieved from https://www.toxicology.org/events/am/AM2019/program-details.asp

Brown, N., van Rongen, J., van de Velde, J., & Williams, M. (2019). A Reproduction of the Results of Onyike et al. (2003). *PsyArXiv*, 6–8. https://doi.org/https://doi.org/10.31234/osf.io/hbnm9

CDC. (2001). *Blood and Hair Mercury Levels in Young Children and Women of Childbearing Age --- United States, 1999*. Retrieved from https://www.cdc.gov/mmwr/preview/mmwrhtml/mm5008a2.htm

Centers for Disease Control and Prevention. (2019a). *National Health and Nutrition Examination Survey: Overview*. https://doi.org/10.4135/9781483346663.n401

Centers for Disease Control and Prevention. (2019b). National Health and Nutrition Examination Survey 1999 – 2014 Survey Content Brochure 1999 – 2014.

Christensen, K., Sobus, J., Phillips, M., Blessinger, T., Lorber, M., & Tan, Y. (2014). Changes in epidemiologic associations with different exposure metrics : A case study of phthalate exposure associations with body mass index and waist circumference ☆ , ☆☆. *Environment International*, *73*, 66–76. https://doi.org/10.1016/j.envint.2014.07.010

Connor, A. M. O., Tsafnat, G., Gilbert, S. B., Thayer, K. A., Shemilt, I., Thomas, J., … Wolfe, M. S. (2019). Still moving toward automation of the systematic review process : a summary of discussions at the third meeting of the International Collaboration for Automation of Systematic Reviews (ICASR), 1–5.

Dye, B. A., Hirsch, R., & Brody, D. J. (2002). The Relationship between Blood Lead Levels and Periodontal Bone Loss in the United States , 1988 – 1994, *110*(10), 997–1002.

Easthope, T., & Valeriano, L. (2007). Phase Out Persistent, Bioaccumulative or Highly Toxic Chemicals. *New Solutions*, *17*(3), 193–207. https://doi.org/10.2190/NS.17.3.d

EPA. (2003). EPA Handbook for Use of Data from the National Health and Nutrition Examination Surveys (NHANES): A Goldmine of Data for Environmental Health Analysis, 1–180.

Gangemi, A., Presutti, V., Recupero, D. R., Nuzzolese, A. G., & Draicchio, F. (2009). Semantic Web Machine Reading with FRED, *1*, 1–5.

Gumbo, D., Clendenning, J., Martius, C., Moombe, K., Grundy, I., Nasi, R., … Petrokofsky, G. (2018). How have carbon stocks in central and southern Africa ' s miombo woodlands changed over the last 50 years ? A systematic map of the evidence. *Environmental Evidence*, 1–19. https://doi.org/10.1186/s13750-018-0128-0

Gupta, P., Goel, A., Lin, J., Sharma, A., Wang, D., & Zadeh, R. (2013). WTF : The Who to Follow Service at Twitter, 505–514.

Hitzler, P. (2012). Facebook Linked Data via the Graph API, *1*, 1–6.

James, K. L., Randall, N. P., & Haddaway, N. R. (2016). A methodology for systematic mapping in environmental sciences. *Environmental Evidence*, *5*(1), 7. https://doi.org/10.1186/s13750-016-0059-6

Kobrosly, R. W., Parlett, L. E., Stahlhut, R. W., Barrett, E. S., & Swan, S. H. (2012). Socioeconomic factors and phthalate metabolite concentrations among United States women of reproductive age. *Environmental Research*, *115*, 11–17. https://doi.org/10.1016/j.envres.2012.03.008

Markowitz, L. E., Hariri, S., Lin, C., Dunne, E. F., Steinau, M., Mcquillan, G., & Unger, E. R. (2013). Reduction in Human Papillomavirus (HPV) Prevalence Among Young Women Following HPV Vaccine Introduction in the United States , National Health and Nutrition Examination Surveys , 2003 – 2010, *208*, 385–393. https://doi.org/10.1093/infdis/jit192

Marshall, I. J., & Wallace, B. C. (2019). Toward systematic review automation : a practical

guide to using machine learning tools in research synthesis, *9*, 1–10.

Muntner, P., Menke, A., DeSalvo, K. B., Rabito, F. A., & Batuman, V. (2005). Continued Decline in Blood Lead Levels Among Adults in the United States. *Archives of Internal Medicine*, *165*.

Nelson, J. W., Scammell, M. K., Hatch, E. E., & Webster, T. F. (2012). Social disparities in exposures to bisphenol A and polyfluoroalkyl chemicals : a cross-sectional study within NHANES 2003-2006, 1–15.

Neo4j. (2019). Neo4j. Retrieved from https://neo4j.com/

Neo4j. (2020). Neo4j Bloom. Retrieved from http://neo4j.com/bloom/

Pelch, K. E., Reade, A., Wolffe, T. A. M., & Kwiatkowski, C. F. (2019). PFAS health effects database: Protocol for a systematic evidence map. *Environment International*, *130*(May). https://doi.org/10.1016/j.envint.2019.05.045

Project Jupyter. (2019). Jupyter. Retrieved from https://jupyter.org/

Robinson, I., Webber, J., & Eifrem, E. (2015). *Graph Databases*. O'Reilly. https://doi.org/10.1016/B978-0-12-407192-6.00003-0

Rusu, D., Dali, L., Fortuna, B., Grobelnik, M., & Mladeni, D. (2014). Triplet extraction from sentences.

Sasaki, B. M., Chao, J., Howard, R., Sasaki, B. M., Chao, J., & Howard, R. (2018). Graph Databases for Beginners Graph Databases For Beginners.

Singhal, A. (2012). Introducing the Knowledge Graph: things, not strings. Retrieved from https://googleblog.blogspot.com/2012/05/introducing-knowledge-graph-things-not.html

Small, N. (2019). The Py2neo v4 Handbook. Retrieved from http://py2neo.org/v4/

Sobus, J. R., DeWoskin, R. S., Tan, Y. M., Pleil, J. D., Phillips, M. B., George, B. J., … Edwards, S. W. (2015). Uses of NHANES biomarker data for chemical risk assessment: Trends, challenges, and opportunities. *Environmental Health Perspectives*, *123*(10), 919–927. https://doi.org/10.1289/ehp.1409177

Stone, B. M., & Reynolds, C. R. (2003). Can the National Health and Nutrition Examination Survey III (NHANES III) data help resolve the controversy over low blood lead levels and neuropsychological development in children ?, *18*, 219–244. https://doi.org/10.1016/S0887-6177(03)00005-2

Taboureau, O., & Audouze, K. (2017). Human environmental disease network: A computational model to assess toxicology of contaminants. *Altex*, *34*(2), 289–300. https://doi.org/10.14573/altex.1607201

Tyrrell, J., Melzer, D., Henley, W., Galloway, T. S., & Osborne, N. J. (2013). Associations between socioeconomic status and environmental toxicant concentrations in adults in the USA : NHANES 2001 – 2010. *Environment International*, *59*, 328–335. https://doi.org/10.1016/j.envint.2013.06.017

van Altena, A. J., Spijker, R., Olabarriaga, S. D. (2019). Usage of automation tools in systematic reviews. *Research Synthesis Methods*, 10 (1), 72–82. https://doi.org/10.1002/jrsm.1335

Wolffe, T. A. M., Vidler, J., Halsall, C., Hunt, N., & Whaley, P. (2020). A survey of systematic evidence mapping practice and the case for knowledge graphs in environmental health & toxicology. *Manuscript Submitted for Publication.*

Wolffe, T. A. M., Whaley, P., Halsall, C., Rooney, A. A., & Walker, V. R. (2019). Systematic evidence maps as a novel tool to support evidence-based decision-making in chemicals policy and risk management. *Environment International*, *130*(March), 104871.

https://doi.org/10.1016/j.envint.2019.05.065

Conclusion

Evidence-based approaches such as systematic review, which have transformed medical and social sciences, have much to offer environmental health. Interest in their uptake, especially for chemical risk assessment and regulatory purposes, is growing – and is representative of an overall push to reform the resource efficiency, representativeness and rigour in developing chemicals policy. These sentiments are increasingly important as the chemicals industry continues to grow, and as the availability of data relevant to chemical risk assessment exponentially increases.

However, uptake of systematic review for chemical risk assessment is not without its challenges. Some of these challenges are more fundamental to the systematic review process (e.g. the need to appropriately manage subjectivity when assessing risk of bias of included studies, see Chapter One); while others are specific to regulatory decision-making (e.g. limited resource availability and a broader set of information requirements than can be addressed by a single systematic review, see Chapter Two); or to environmental health data itself (e.g. managing the integrity of highly complex and connected data, see Chapter Three). Facilitating uptake of evidence-based approaches to chemical risk assessment requires that these challenges are understood and addressed.

Continuing to develop and communicate best practice for environmental health systematic reviews serves to address the challenges associated with accessing the methodology itself. This is a key focus for several working groups dedicated to the EBT movement (Johns Hopkins Bloomberg School of Public Health, 2019; NTP-OHAT, 2019; UCSF Program on Reproductive Health and the Environment, 2019 etc.) who produce comprehensive systematic review guidance, training, standards and methodological tools tailored for the environmental health

context (e.g. Hoffmann et al., 2017; NTP, 2015; OHAT, 2015; Schaefer & Myers, 2017; Woodruff & Sutton, 2014 etc.). Open and constructive dialogue regarding best practice for EH SRs may help to establish standards and consensus within the field, or at the least – equip stakeholders with the understanding required to critically appraise current evidence synthesis practice. Such ongoing communication is an important aspect of the move toward more evidence-based approaches to chemical risk assessment, especially as ill-defined systematic review practices begin to appear in regulatory frameworks (e.g. Scientific Committee on Consumer Safety SCCS, 2019) or as aspects of the methodology adapted for regulatory frameworks deviate from best-practice (e.g. the numerical scoring system recommended for assessing risk of bias in the EPA's systematic review methodology for TSCA risk evaluations (EPA, 2018)).

Understanding the needs and limitations of regulatory decision-making is key to devising tools and/or workflows which facilitate the uptake of evidence-based approaches. In this thesis, systematic evidence mapping is identified as a methodological solution which addresses these needs and limitations (see Chapter Two). By providing a much broader overview of the evidence-landscape, SEMs facilitate the identification of trends (including issues of emerging regulatory concern), on which to focus resources. The computationally accessible and easily updated format of SEMs as queryable databases renders them multi-purpose and "re-useable", ensuring that any data collated, characterised and stored is available for varied present and/or future uses. This creates larger returns on the resources invested when developing a SEM. The evidence-surveillance function served by SEMs is an integral component of existing regulatory decision-making frameworks – and thus the methodology can be more readily incorporated into current chemical risk assessment workflows. Similarly, SEMs are able to serve the information retrieval steps of systematic review. Thus, not only do SEMs facilitate the formulation of informative (rather than empty) systematic review research

questions through identification of research clusters, but they also potentially reduce the workload associated with conducting systematic reviews.

As interest in, and application of, systematic evidence mapping beings to emerge – it is important to understand what underpins the utility of the methodology for chemical risk assessment and regulatory decision-making. On a fundamental level, systematic evidence mapping transforms unstructured, textual data which is heterogeneous and distributed over disparate sources – into a single, organised and machine-readable resource. It is this accessibility of data which allows trends across vast quantities of evidence to be programmatically explored, quickly and efficiently. Thus, ensuring that the data management practices of systematic evidence maps do not compromise this accessibility is vital for the successful application of the methodology. In this thesis, the familiar, tabular-based data management practices of evidence mapping in other fields were found to be poorly suited to the complex and highly connected nature of toxicology and environmental health data. The rigidity of these data structures was found to compromise data integrity and consequently reduced the utility of evidence maps for varied application. The graph data model was identified as a flexible alternative, capable of directly storing the relationships between highly connected toxicology data (see Chapter Three).

Modelling toxicology data as a graph, and storing relationships as queryable entities, has significant implications for trend-spotting – facilitating complex and highly resolved graph queries which traverse patterns of information. These complex queries have the potential to move systematic evidence mapping beyond the identification of broad trends such as research gaps and research clusters, and toward more highly resolved applications such as the identification of adverse outcome pathways (Villeneuve et al., 2018). This may serve to facilitate a more predictive (rather than simply proactive) approach to chemical risk assessment. However, graph-based technologies are novel and unfamiliar to stakeholders

working within environmental health. There is therefore a need to bridge the gap between those with expertise in the implementation of graph-based data management, and those with expertise in the potential applications of graph-based data management (i.e. regulatory decision-making within toxicology and environmental health). Continued exploration and communication of the potential gains in data integrity, transparency and interoperability offered by the application of graphs within environmental health will serve to increase familiarity within the field and resolve the future research required to implement evidence mapping at scale (see Chapter Four) - expediting the uptake of resource-efficient evidence-based methods within chemicals policy and wider environmental health.

Future Work

Successfully implementing evidence-based approaches to chemical risk assessment requires that the resource burden associated with these approaches is lessened. Whilst systematic evidence mapping offers a more resource-efficient framework for pursuing evidence-based decision-making - its manual workflow continues to present barriers to wider uptake. Thus, a key focus of future work within the field is the development of automated approaches to evidence-mapping.

Automation is a topic of increasing interest for evidence synthesis applications, with several ongoing research efforts (van Altena et al., 2019; Connor et al., 2019; Marshall & Wallace, 2019). These efforts have largely manifested as tools which assist the screening and/or literature tagging aspects of evidence synthesis workflows and are beginning to appear as standard features of review management software (e.g. Evidence Partners, 2019; Sciome, 2018). Screening literasture for inclusion in a systematic review represents a typical case for application of automated approaches. This is because, in addressing specific research questions, the inclusion/exclusion criteria for systematic reviews are more clearly and specifically defined. Thus, machine learning classifiers are able to learn from a manually screened training set where all of the included studies are likely to have very similar features. In the field of medicine, incorporating automation tools into the systematic review workflow has been reported to reduce the time and workload required to complete a systematic review e.g. by 40% for the Rayyan tool (Ouzzani, 2017), among many others.

However, as discussed in this thesis – resolving a specific research question for chemical risk assessment applications is more difficult – and the range of potentially included evidence

considerably more heterogeneous. Similarly, the inclusion criteria for systematic evidence maps are much broader. As systematic evidence mapping is fundamentally concerned with making data accessible for querying, the data extraction phase of the methodology is most important and demanding. Thus, focusing automation efforts on the screening stage of the methodology is insufficient for reducing the manual workload of SEMs. Developing automated approaches to data extraction, which extend beyond the simple identification of key words and toward the consideration of context, is therefore a challenge on which to focus future research.

However, it can be argued that the issue of data accessibility which systematic evidence mapping targets, is not a challenge for machine learning, but is an issue concerning data standards. Ensuring that environmental health evidence is published in a machine-readable format (rather than as unstructured text) in the first instance will alleviate the need to manually process and store data in a machine accessible format. Similar issues can be found motivating the Semantic Web movement (Berners-Lee et al., 2001), which fundamentally strives to make unstructured data published on the web machine-accessible – allowing for greater connectivity and interoperability. A variety of tools have emerged from this movement, including: the resource description framework (RDF) data standard (Manola & Miller, 2004), in which data are stored as a graph of semantic triples (see Chapter Three); the SPARQL Protocol and RDF Query Language (SPARQL) for querying data stored in RDF format (The W3C SPARQL Working Group, 2013); and linked open data libraries (e.g. DBpedia, 2020), which allow RDF datasets to be linked through semantic triples to other datasets for greater contextual value. Exploring and/or exploiting these tools for evidence mapping applications and understanding the overlap/applicability of the Semantic Web movement to the EBT movement, represents a key area for future work.

Once data are machine-readable, evidence mapping approaches can move away from data extraction and focus on deriving value from accessible data. Modelling, characterising and querying data will become the key focus for automation efforts, resolving the "signal" for a particular research application from what will become a considerably "noisy" computationally accessible and interconnected evidence landscape. Facilitating automated approaches to querying or deriving value from accessible data requires implementing graph-based controlled vocabulary ontologies which organise data for applications within a particular domain (see Chapter Three). Such ontologies also form a key aspect of the Semantic Web toolkit. More established applications of ontologies for querying and inferencing over graphs of data can be found in the pharmaceutical industry (Samwald et al., 2011; Wild et al., 2012; Yankulov, 2019), and domain-specific ontologies available for biological fields e.g. (Ashburner et al., 2000; National Center for Biomedical Ontology, 2019). These use cases represent valuable learning opportunities for evidence-based approaches to chemical risk assessment and risk management decision-making. Future work within this area will require the development of toxicological ontologies which are relevant to regulatory workflows (Hardy et al., 2012). This represents a largely manual, consensus-building exercise and reiterates the importance of continued communication of evidence-based methods within chemicals policy.

In the interim, the true value of pursuing these avenues of future research can be explored and refined by conducting slightly more narrowly focused evidence mapping exercises within environmental health and toxicology, in which the utility of the methodology can still be demonstrated and developed without exceeding resource availability. Producing evidence maps which can be incorporated directly into chemical risk assessment workflows will promote interest and uptake of the methodology, as well as resolve further issues in need of future research e.g. the development of user interfaces for evidence maps which are both fit-for-purpose and accessible to chemical risk assessors, decision-makers and/or the public.

Given that current regulatory workflows operate on a chemical-by-chemical basis, producing evidence maps which collate all available and relevant information on a particular chemical (or class of chemicals) offers a more immediately achievable means of exploring and adapting the methodology – providing much-needed "case studies". These exemplar evidence-maps will provide the foundations from which future research efforts can be successfully developed.

References

Abelkop, A. D. K., & Graham, J. (2014). Regulation of Chemical Risks: Lessons for TSCA Reform from Canada and the European Union. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.2499309

Aggarwal, C. C., & Wang, H. (2010). Managing and Mining Graph Data (Vol. 40). Springer. https://doi.org/10.1007/978-1-4419-6045-0

Ahluwalia, N., Dwyer, J., Terry, A., Moshfegh, A., & Johnson, C. (2016). Update on NHANES Dietary Data : Focus on Collection , Release , Analytical Considerations , and Uses to Inform Public Policy 1 , 2, (9). https://doi.org/10.3945/an.115.009258.status

Aiassa, E., Higgins, J. P. T., Frampton, G. K., Greiner, M., Afonso, A., Amzal, B., … Verloo, D. (2015). Applicability and Feasibility of Systematic Review for Performing Evidence-Based Risk Assessment in Food and Feed Safety. Critical Reviews in Food Science and Nutrition, 55(7), 1026–1034. https://doi.org/10.1080/10408398.2013.769933

Amec Foster Wheeler Environment & Infrastructure UK Limited. (2017). A study to gather insights on the drivers, barriers, costs and benefits for updating REACH registration and CLP notification dossiers.

Applegate, J. S. (2008). Synthesizing TSCA and REACH: Practical principles for chemical regulation reform. Ecology Law Quarterly, 35(4), 721–769.

ArangoDB. (2019). Graphs and ArangoDB. Retrieved from https://www.arangodb.com/arangodb-training-center/graphs/

Armijo-Olivo, S., Stiles, C. R., Hagen, N. A., Biondo, P. D., & Cummings, G. G. (2012). Assessment of study quality for systematic reviews: A comparison of the Cochrane

Collaboration Risk of Bias Tool and the Effective Public Health Practice Project Quality Assessment Tool: Methodological research. Journal of Evaluation in Clinical Practice, 18(1), 12–18. https://doi.org/10.1111/j.1365-2753.2010.01516.x

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., … Rubin, Gerald M. Sherlock, G. (2000). Gene Ontology: tool for the unification of biology. Nature Genetics, 25(1), 25–29. https://doi.org/10.1038/75556.Gene

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., … Rubin, Gerald M. Sherlock, G. (2000). The Gene Ontology Consortium, Michael Ashburner1, Catherine A. Ball3, Judith A. Blake4, David Botstein3, Heather Butler1, J. Michael Cherry3, Allan P. Davis4, Kara Dolinski3, Selina S. Dwight3, Janan T. Eppig4, Midori A. Harris3, David P. Hill4, Laurie Is. Nature Genetics, 25(1), 25–29. https://doi.org/10.1038/75556.Gene

Baker, N., Boobis, A., Burgoon, L., Carney, E., Currie, R., Fritsche, E., … Daston, G. (2018). Building a developmental toxicity ontology. Birth Defects Research, 110(6), 502–518. https://doi.org/10.1002/bdr2.1189

Balk, E., Bonis, P., Moskowitz, H., Schmid, C. H., Ioannidis, J. P. A., Wang, C., & Lau, J. (2002). Correlation of QualityMeasures With Estimates of Treatment Effect inMeta-analyses of Randomized Controlled Trials. JAMA, 287(22), 2973–2982.

Barra Caracciolo, A., de Donato, G., Finizio, A., Grenni, P., Santoro, S., & Petrangeli, A. B. (2013). A New Online Database on Chemicals in Accordance with REACH Regulation. Human and Ecological Risk Assessment, 19(6), 1682–1699. https://doi.org/10.1080/10807039.2012.708271

Bell, S., & Edwards, S. (2015). Identification and Prioritization of Relationships between Environmental Stressors and Adverse Human Health Impacts, 123(11), 1193–1199.

Berkman, N. D., Lina Santaguida, M. P., Viswanathan, M., & Morton, S. C. (2014). The Empirical Evidence of Bias in Trials Measuring Treatment Differences. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK253181/pdf/Bookshelf_NBK253181.pdf

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web. Scientific American, 284(5), 34–43.

Bernes, C., Bullock, J. M., Jakobsson, S., Rundlöf, M., Verheyen, K., & Lindborg, R. (2017). How are biodiversity and dispersal of species affected by the management of roadsides? A systematic map. Environmental Evidence, 6(1), 1–16. https://doi.org/10.1186/s13750-017-0103-1

Bernes, C., Jonsson, B. U., Junninen, K., Lõhmus, A., Macdonald, E., Müller, J., & Sandström, J. (2015). What is the impact of active management on biodiversity in boreal and temperate forests set aside for conservation or restoration? A systematic map. Environmental Evidence, 4(1), 1–22. https://doi.org/10.1186/s13750-015-0050-7

Beronius, A., & Hanberg, A. (2017). SciRAP workshop report Bridging the gap between academic research and Workshop report : " Bridging the gap between academic research and. Retrieved from http://www.diva-portal.org/smash/get/diva2:1074866/FULLTEXT02.pdf

Beronius, A., & Vandenberg, L. N. (2015). Using systematic reviews for hazard and risk assessment of endocrine disrupting chemicals. Reviews in Endocrine and Metabolic Disorders, 16(4), 273–287. https://doi.org/10.1007/s11154-016-9334-7

Beverly, B. (2019). Abstract 3267: Potential Alternatives to Systematic Review: Evidence Maps and Scoping Reviews. Retrieved from https://www.toxicology.org/events/am/AM2019/program-details.asp

Blaha, M. R., Premerlani, W. J., & Rumbaligh, J. E. (1988). Relational database design using an object-oriented methodology. Computing Practices, 31(4).

Bolden, A. L., Rochester, J. R., Schultz, K., & Kwiatkowski, C. F. (2017). Polycyclic aromatic hydrocarbons and female reproductive health: A scoping review. Reproductive Toxicology, 73, 61–74. https://doi.org/10.1016/j.reprotox.2017.07.012

Bolvig, J., Juhl, C. B., Boutron, I., Tugwell, P., Ghogomu, E. A. T., Pardo, J. P., … Christensen, R. (2018). Some Cochrane risk-of-bias items are not important in osteoarthritis trials: a meta-epidemiological study based on Cochrane reviews. Journal of Clinical Epidemiology, 95, 128–136. https://doi.org/10.1016/j.jclinepi.2017.11.026

Borah, R., Brown, A. W., Capers, P. L., & Kaiser, K. A. (2017). Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry. BMJ Open, 7(2), 1–7. https://doi.org/10.1136/bmjopen-2016-012545

Brown, N., van Rongen, J., van de Velde, J., & Williams, M. (2019). A Reproduction of the Results of Onyike et al. (2003). PsyArXiv, 6–8. https://doi.org/https://doi.org/10.31234/osf.io/hbnm9

Brown, V. J. (2003). REACHing for Chemical Safety. Environmental Health Perspectives, 111(14).

Buxton, L. (2017). Echa clarifies Brexit impacts on UK, EU regulatory obligations. Retrieved October 25, 2017, from https://chemicalwatch.com/59534/echa-clarifies-brexit-impacts-on-uk-eu-regulatory-obligations

CAMARADES-NC3Rs. (2019). Preclinical Systematic Review & Meta-analysis Facility (SyRF). Retrieved February 6, 2019, from http://syrf.org.uk/

Castelltort, A., & Laurent, A. (2016). Rogue behavior detection in NoSQL graph databases. Journal of Innovation in Digital Ecosystems, 3(2), 70–82. https://doi.org/10.1016/j.jides.2016.10.004

CDC. (2001). Blood and Hair Mercury Levels in Young Children and Women of Childbearing Age --- United States, 1999. Retrieved from https://www.cdc.gov/mmwr/preview/mmwrhtml/mm5008a2.htm

Centers for Disease Control and Prevention. (2019a). National Health and Nutrition Examination Survey: Overview. https://doi.org/10.4135/9781483346663.n401

Centers for Disease Control and Prevention. (2019b). National Health and Nutrition Examination Survey 1999 – 2014 Survey Content Brochure 1999 – 2014.

Chalmers, I., Hedges, L., & Cooper, H. (2002). A brief history of research synthesis. Evaluation & the Health …, 25(1), 12–37. https://doi.org/10.1177/0163278702025001003

Cheng, S. H., Macleod, K., Ahlroth, S., Onder, S., Perge, E., Shyamsundar, P., … Miller, D. C. (2019). A systematic map of evidence on the contribution of forests to poverty alleviation. Environmental Evidence, 1–22. https://doi.org/10.1186/s13750-019-0148-4

Christensen, K., Sobus, J., Phillips, M., Blessinger, T., Lorber, M., & Tan, Y. (2014). Changes in epidemiologic associations with different exposure metrics : A case study of phthalate exposure associations with body mass index and waist circumference ☆ , ☆☆. Environment International, 73, 66–76. https://doi.org/10.1016/j.envint.2014.07.010

Clapton, J., Rutter, D., & Sharif, N. (2009). SCIE Systematic mapping guidance April 2009. Retrieved from https://www.scie.org.uk/publications/researchresources/rr03.pdf

Cochrane Community. (2019). MECIR Manual: Assessing risk of bias in included studies (C52-C60). Retrieved September 20, 2001, from https://community.cochrane.org/mecir-manual/standards-conduct-new-cochrane-intervention-reviews-c1-c75/performing-review-c24-c75/assessing-risk-bias-included-studies-c52-c60

Cochrane Methods Group. (2017). Assessing Risk of Bias in Included Studies. Retrieved November 3, 2017, from http://methods.cochrane.org/bias/assessing-risk-bias-included-studies

Collaboration for Environmental Evidence. (2019a). CEE Meetings. Retrieved September 20, 2007, from https://www.environmentalevidence.org/meetings

Collaboration for Environmental Evidence. (2019b). Completed Reviews. Retrieved from http://www.environmentalevidence.org/completed-reviews

Collaboration for Environmental Evidence. (2019c). Environmental Evidence: Systematic Map Submission Guidelines. Retrieved from https://environmentalevidencejournal.biomedcentral.com/submission-guidelines/preparing-your-manuscript/systematic-map

Commission of the European Communities. (1998). Report on the operation of DIRECTIVE 67/548/EEC, DIRECTIVE 88/379/EEC, REGULATION (EEC) 793/93 AND DIRECTIVE 761769/EEC.

Commission of the European Communities. (2001). White Paper: Strategy for a future chemicals policy (Vol. 13). https://doi.org/10.1007/BF03038641

Connor, A. M. O., Tsafnat, G., Gilbert, S. B., Thayer, K. A., Shemilt, I., Thomas, J., … Wolfe, M. S. (2019). Still moving toward automation of the systematic review process : a summary of discussions at the third meeting of the International Collaboration for Automation of Systematic Reviews (ICASR), 1–5.

Cooper, G. S., Lunn, R. M., Ågerstrand, M., Glenn, B. S., Kraft, A. D., Luke, A. M., & Ratcliffe, J. M. (2016). Study sensitivity: Evaluating the ability to detect effects in systematic reviews of chemical exposures. Environment International, 92–93, 605–610. https://doi.org/10.1016/j.envint.2016.03.017

Covidence. (2019). Covidence. Retrieved January 14, 2019, from https://www.covidence.org/home

Cresswell, C. J., Wilcox, A., Randall, N. P., & Cunningham, H. M. (2018). What specific plant traits support ecosystem services such as pollination, bio-control and water quality protection in temperate climates? A systematic map. Environmental Evidence, 7(1), 1–13. https://doi.org/10.1186/s13750-018-0120-8

Crossley, N. A., Sena, E., Goehler, J., Horn, J., Van Der Worp, B., Bath, P. M. W., … Dirnagl, U. (2008). Empirical evidence of bias in the design of experimental stroke studies: A metaepidemiologic approach. Stroke, 39(3), 929–934. https://doi.org/10.1161/STROKEAHA.107.498725

Datawrapper GmbH. (2019). Datawrapper. Retrieved from https://www.datawrapper.de/

DBpedia. (2020). DBpedia. Retrieved from http://wiki.dbpedia.org/

De Vet, H. C. W., De Bie, R. A., Van Der Heijden, G. J. M. G., Verhagen, A. P., Sijpkes, P., & Knipschild, P. G. (1997). Systematic reviews on the basis of methodological criteria. Physiotherapy, 83(6), 284–289. https://doi.org/10.1016/S0031-9406(05)66175-5

Dechartres, A., Trinquart, L., Faber, T., & Ravaud, P. (2016). Empirical evaluation of which trial characteristics are associated with treatment effect estimates. Journal of Clinical Epidemiology, 77, 24–37. https://doi.org/10.1016/j.jclinepi.2016.04.005

Deeks, J. J., Dinnes, J., D'Amico, R., Sowden, A. J., Sakarovitch, C., Song, F., … European Carotid Surgery Trial Collaborative Group. (2003). Evaluating non-randomised intervention studies. Health Technology Assessment (Winchester, England), 7(27). https://doi.org/96-26-99 [pii]

Dillon, T., Chang, E., Hadzic, M., & Wongthongtham, P. (2008). Differentiating conceptual modelling from data modelling, knowledge modelling and ontology modelling and a notation

for ontology modelling. Conferences in Research and Practice in Information Technology Series, 79.

Ding, W., Liang, P., Tang, A., & Van Vliet, H. (2014). Knowledge-based approaches in software documentation: A systematic literature review. Information and Software Technology, 56(6), 545–567. https://doi.org/10.1016/j.infsof.2014.01.008

Downs, S. H., & Black, N. (1998). The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. Journal of Epidemiology and Community Health, 52(6), 377–384. https://doi.org/10.1136/jech.52.6.377

Dye, B. A., Hirsch, R., & Brody, D. J. (2002). The Relationship between Blood Lead Levels and Periodontal Bone Loss in the United States , 1988 – 1994, 110(10), 997–1002.

Easthope, T., & Valeriano, L. (2007). Phase Out Persistent, Bioaccumulative or Highly Toxic Chemicals. New Solutions, 17(3), 193–207. https://doi.org/10.2190/NS.17.3.d

ECHA. (2016). Practical guide How to use alternatives to animal testing to fulfil your information requirements for REACH registration.

ECHA. (2019a). REACH registration results. Retrieved from https://echa.europa.eu/reach-registrations-since-2008

ECHA. (2019b). Weight of evidence. Retrieved from https://echa.europa.eu/support/registration/how-to-avoid-unnecessary-testing-on-animals/weight-of-evidence

Eckley, N., & Selin, H. (2004). All talk, little action: Precaution and European chemicals regulation. Journal of European Public Policy, 11(1), 78–105. https://doi.org/10.1080/1350176042000164316

Edwards, S. W., Tan, Y.-M., Villeneuve, D. L., Meek, M. E., & McQueen, C. A. (2015). Adverse Outcome Pathways--Organizing Toxicological Information to Improve Decision Making. Journal of Pharmacology and Experimental Therapeutics, 356(1), 170–181. https://doi.org/10.1124/jpet.115.228239

Egeghy, P. P., Judson, R., Gangwal, S., Mosher, S., Smith, D., Vail, J., & Cohen Hubal, E. A. (2012). The exposure data landscape for manufactured chemicals. Science of the Total Environment, 414, 159–166. https://doi.org/10.1016/j.scitotenv.2011.10.046

Elmasri, R., & Navathe, S. B. (2016). The Relational Data Model and Relational Database Constraints. Fundamentals of Database Systems, 150–152.

EPA. (2003). EPA Handbook for Use of Data from the National Health and Nutrition Examination Surveys (NHANES): A Goldmine of Data for Environmental Health Analysis, 1–180.

EPA. (2018). Application of Systematic Review in TSCA Risk Evaluations, 1–247. Retrieved from https://www.epa.gov/sites/production/files/2018-06/documents/final_application_of_sr_in_tsca_05-31-18.pdf

EPA's Office of Pollution Prevention and Toxics. (1998). Chemical Hazard Data Availability Study - What Do We Really Know About the Safety of High Production Volume Chemicals? Retrieved from https://noharm-uscanada.org/sites/default/files/documents-files/915/Chemical_Hazard_Data_Availability_Study_1998.pdf

European Chemicals Agency. (2011). Guidance on information requirements and chemical safety assessment Chapter R . 4: Evaluation of available information. Retrieved from https://echa.europa.eu/documents/10162/13643/information_requirements_r4_en.pdf

European Chemicals Agency. (2016). Guidance on registration. https://doi.org/ECHA-12-G-07-EN

European Chemicals Agency. (2017a). Guidance on information requirements and chemical safety assessment: QSARs and grouping of chemicals. Guidance for the Implementation of REACH, 134. https://doi.org/10.2823/43472

European Chemicals Agency. (2017b). Study finds companies lack incentives for updating their REACH registrations. Retrieved November 8, 2018, from https://echa.europa.eu/-/study-finds-companies-lack-incentives-for-updating-their-reach-registrations

European Chemicals Agency. (2018a). Information Requirements. Retrieved November 9, 2018, from https://echa.europa.eu/regulations/reach/registration/information-requirements

European Chemicals Agency. (2018b). Strategy for gathering your data. Retrieved November 9, 2018, from https://echa.europa.eu/support/registration/strategy-for-gathering-your-data

European Comission. (2007). REACH In Brief. https://doi.org/10.1067/j.cpsurg.2007.09.005

European Comission. (2016). REACH. Retrieved October 25, 2017, from http://ec.europa.eu/environment/chemicals/reach/reach_en.htm

European Environment Agency. (2013). Late lessons from early warnings: science, precaution, innovation — European Environment Agency. Retrieved from http://www.eea.europa.eu/publications/late-lessons-2

European Food Safety Authority. (2010). Application of systematic review methodology to food and feed safety assessments to support decision making 1 EFSA Guidance for those carrying out systematic reviews, 8(6). https://doi.org/10.2903/j.efsa.2010.1637.

Evidence Partners. (2019). DistillerSR. Retrieved January 14, 2019, from https://www.evidencepartners.com/products/distillersr-systematic-review-software/

Gangemi, A., Presutti, V., Recupero, D. R., Nuzzolese, A. G., & Draicchio, F. (2009). Semantic Web Machine Reading with FRED, 1, 1–5.

Garg, A. X., Hackam, D., & Tonelli, M. (2008). Systematic review and meta-analysis: When one study is just not enough. Clinical Journal of the American Society of Nephrology, 3(1), 253–260. https://doi.org/10.2215/CJN.01430307

Ghrab, A., Romero, O., Skhiri, S., Vaisman, A., & Zimányi, E. (2016). GRAD: On Graph Database Modeling. Retrieved from http://arxiv.org/abs/1602.00503

Giraudeau, B., Higgins, J. P. T., Tavernier, E., & Trinquart, L. (2016). Sample size calculation for meta-epidemiological studies. Statistics in Medicine, 35(2), 239–250. https://doi.org/10.1002/sim.6627

Glasziou, P., Djulbegovic, B., & Burls, A. (2006). Are systematic reviews more cost-effective than randomised trials? Lancet, 367(9528), 2057–2058. https://doi.org/10.1016/S0140-6736(06)68919-8

Goldacre, B. (2013). Bad Pharma: How Medicine is Broken, and How We Can Fix It. HarperCollins.

Golden, R., Doull, J., Waddell, W., & Mandel, J. (2003). Potential Human Cancer Risks from Exposure to PCBs: A Tale of Two Evaluations. Critical Reviews in Toxicology, 33(5), 543–580. https://doi.org/10.1080/10408440390242414

Grant, M. J., & Booth, A. (2009). A typology of reviews: An analysis of 14 review types and associated methodologies. Health Information and Libraries Journal, 26(2), 91–108. https://doi.org/10.1111/j.1471-1842.2009.00848.x

Greenland, S., & O'Rourke, K. (2001). On the bias produced by quality scores in meta-analysis, and a hierarchical view of proposed solutions. Biostatistics (Oxford, England), 2(4), 463–471. https://doi.org/10.1093/biostatistics/2.4.463

Groenwold, R. H. H., & Rovers, M. M. (2010). The Catch-22 of appraisals on the quality of observational studies. Journal of Clinical Epidemiology, 63(10), 1059–1060. https://doi.org/10.1016/j.jclinepi.2010.04.013

Gumbo, D., Clendenning, J., Martius, C., Moombe, K., Grundy, I., Nasi, R., … Petrokofsky, G. (2018). How have carbon stocks in central and southern Africa ' s miombo woodlands changed over the last 50 years ? A systematic map of the evidence. Environmental Evidence, 1–19. https://doi.org/10.1186/s13750-018-0128-0

Gupta, P., Goel, A., Lin, J., Sharma, A., Wang, D., & Zadeh, R. (2013). WTF : The Who to Follow Service at Twitter, 505–514.

Guyatt, G., Cairns, J., Churchill, D., Cook, D., Haynes, B., Hirsh, J., … Tugwell, P. (1992). Evidence-Based Medicine A New Approach to Teaching the Practice of Medicine. JAMA. https://doi.org/10.1001/jama.1992.03490170092032

Haddaway, N. R., & Westgate, M. J. (2018). Predicting the time needed for environmental systematic reviews and systematic maps. Conservation Biology, 0(0), 1–10. https://doi.org/10.1111/cobi.13231

Haddaway, N. R., Bernes, C., Jonsson, B.-G., & Hedlund, K. (2016). The benefits of systematic mapping to evidence-based environmental management. Ambio, 45(5), 613–620. https://doi.org/10.1007/s13280-016-0773-x

Haddaway, N. R., Brown, C., Eales, J., Eggers, S., Josefsson, J., Kronvang, B., … Uusi-Kämppä, J. (2018). The multifunctional roles of vegetated strips around and within agricultural fields. Environmental Evidence, 7(1), 1–43. https://doi.org/10.1186/s13750-018-0126-2

Haddaway, N. R., Hedlund, K., Jackson, L. E., Kätterer, T., Lugato, E., Thomsen, I. K., … Söderström, B. (2015). What are the effects of agricultural management on soil organic

carbon in boreo-temperate systems? Environmental Evidence, 4(1), 1–29. https://doi.org/10.1186/s13750-015-0049-0

Haddaway, N. R., Macura, B., Whaley, P., & Pullin, A. S. (2018). ROSES RepOrting standards for Systematic Evidence Syntheses : pro forma , flow - diagram and descriptive summary of the plan and conduct of environmental systematic reviews and systematic maps. Environmental Evidence, 4–11. https://doi.org/10.1186/s13750-018-0121-7

Haddaway, N. R., Styles, D., & Pullin, A. S. (2014). Evidence on the environmental impacts of farm land abandonment in high altitude/mountain regions: A systematic map. Environmental Evidence, 3(1), 1–19. https://doi.org/10.1186/2047-2382-3-17

Haddaway, N., & Crowe, S. (editors). (2018). Stakeholder engagement in environmental evidence synthesis. Retrieved from http://www.eviem.se/Documents/projekt/2018/SRbookAll.pdf

Hardy, B., Apic, G., Carthew, P., Clark, D., Cook, D., Dix, I., … Yang, C. (2012). Toxicology ontology perspectives. Altex, 29(2), 139–156.

Hartling, L., Ospina, M., Liang, Y., Dryden, D., Hooton, N., Krebs Seida, J., & Klassen, T. (2009). Risk of bias versus quality assessment of randomised controlled trials: cross sectional study. BMJ (Clinical Research Ed.), 339, 1–6. https://doi.org/10.1136/bmj.b4012

Hartung, T. (2009). Toxicology for the twenty-first century. Nature. 460, 208-212. https://doi.org/10.1038/460208a

Have, C. T., & Jensen, L. J. (2013). Databases and ontologies Are graph databases ready for bioinformatics ?, 29(24), 3107–3108. https://doi.org/10.1093/bioinformatics/btt549

HAWC Project. (2013). HAWC: Health Assessment Workplace Collaborative. Retrieved from https://hawcproject.org/

Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, … JAC, S. (2011). The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. British Medical Journal, 343, 889–893. https://doi.org/doi: 10.1136/bmj.d5928

Higgins J.P.T., Green, S. (editors). (2011). Cochrane Handbook for Systematic Reviews of Interventions (Version 5.1.0).

Higgins, J.P.T., Green, S. (2009). Cochrane Handbook for Systematic Reviews of Interventions (Version 5.0.2).

Hoffmann, S., Hartung, T. (2006). Toward an evidence-based toxicology. Human & Experimental Toxicology, 25 , 497–513.

Hoffmann, S., de Vries, R. B. M., Stephens, M. L., Beck, N. B., Dirven, H. A. A. M., Fowle, J. R., … Tsaioun, K. (2017). A primer on systematic reviews in toxicology. Archives of Toxicology, 91(7), 2551–2575. https://doi.org/10.1007/s00204-017-1980-3

Hooijmans, C. R., Rovers, M., de Vries, R. B., 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. Laboratory Animals, 46(4), 356–357. https://doi.org/10.1258/la.2012.012052

Hooker, R. C. (1997). The rise and rise of evidence-based medicine [20]. Lancet, 349(9061), 1329–1330. https://doi.org/10.1016/S0140-6736(05)62552-4

Hróbjartsson, A., Emanuelsson, F., Thomsen, A. S. S., Hilden, J., & Brorson, S. (2014). Bias due to lack of patient blinding in clinical trials. A systematic review of trials randomizing patients to blind and nonblind sub-studies. International Journal of Epidemiology, 43(4), 1272–1283. https://doi.org/10.1093/ije/dyu115

IBM. (2019). Cognos Analytics. Retrieved from https://www.ibm.com/uk-en/products/cognos-analytics?S_PKG=ov42988&cm_mmc=Search_Google-_-

Hybrid+Cloud_Business+Analytics-_-EP_GB-_-
ibm+cognos_Exact_ov42988&cm_mmca1=000026DO&cm_mmca2=10000844&cm_mmca7=
1006727&cm_mmca8=aud-309367918490:kwd-
296914875002&cm_mmca9=_k_CjwKCAjwwtTmBRBqEiwA-
b6c_8xHeckkUle1yQidObbSFtXDDPvTu2KTsnn5ddPbdT1mA1879S7DgRoC5bEQAvD_BwE_k_
&cm_mmca10=304362079192&cm_mmca11=e&mkwid=_k_CjwKCAjwwtTmBRBqEiwA-
b6c_8xHeckkUle1yQidObbSFtXDDPvTu2KTsnn5ddPbdT1mA1879S7DgRoC5bEQAvD_BwE_k_
%7C542%7C158857&cvosrc=ppc.google.ibm
cognos&cvo_campaign=000026DO&cvo_crid=304362079192&Matchtype=e&gclid=CjwKCAj
wwtTmBRBqEiwA-
b6c_8xHeckkUle1yQidObbSFtXDDPvTu2KTsnn5ddPbdT1mA1879S7DgRoC5bEQAvD_BwE

Ingre-Khans, E., Ågerstrand, M., Beronius, A., & Rudén, C. (2016). Transparency of chemical risk assessment data under REACH. Environmental Science: Processes and Impacts, 18(12), 1508–1518. https://doi.org/10.1039/c6em00389c

Ingre-Khans, E., Ågerstrand, M., Beronius, A., & Rudén, C. (2019). Reliability and relevance evaluations of REACH data. Toxicology Research, 8(1), 46–56. https://doi.org/10.1039/c8tx00216a

Innvaer, S., Vist, G., Trommald, M., & Oxman, A. (2002). Review article Health policy-makers ' perceptions of their use of evidence : a systematic review. Journal of Health Services Research & Policy, 7(4), 239–244. https://doi.org/10.1258/135581902320432778

Jadad, A. R., & Enkin, M. W. (2008). Randomized Controlled Trials: Questions, Answers, and Musings, Second Edition. Retrieved from http://www.blackwellpublishing.com/content/BPL_Images/Content_store/Sample_chapter/9781405132664/9781405132664_4_003.pdf

Jadad, A. R., Moore, R. A., Carroll, D., Jenkinson, C., Reynolds, D. J. M., Gavaghan, D. J., & McQuay, H. J. (1996). Assessing the quality of reports of randomized clinical trials: Is blinding necessary? Controlled Clinical Trials, 17(1), 1–12. https://doi.org/10.1016/0197-2456(95)00134-4

James, K. L., Randall, N. P., & Haddaway, N. R. (2016). A methodology for systematic mapping in environmental sciences. Environmental Evidence, 5(1), 7. https://doi.org/10.1186/s13750-016-0059-6

Janković, S., Mladenović, S., Mladenović, D., Vesković, S., & Glavić, D. (2018). Schema on read modeling approach as a basis of big data analytics integration in EIS. Enterprise Information Systems, 12(8–9), 1180–1201. https://doi.org/10.1080/17517575.2018.1462404

Johns Hopkins Bloomberg School of Public Health. (2019). Evidence-Based Toxicology Collaboration. Retrieved from http://ebtox.org/

Juni, P., Altman, D. G., & Egger, M. (2001). Systematic reviews in health care: Assessing the quality of controlled clinical trials. BMJ, 323(7303), 42–46. https://doi.org/10.1136/bmj.323.7303.42

Jüni, P., Witschi, A., Bloch, R., & Egger, M. (1999). The hazards of scoring the quality of clinical trials for meta-analysis. JAMA : The Journal of the American Medical Association, 282(11), 1054–1060. https://doi.org/10.1001/jama.282.11.1054

Karp, P. D. (1996). Database links are a foundation for interoperability. Trends in Biotechnology, 14(8), 273–279. https://doi.org/10.1016/0167-7799(96)10044-5

Kelly, S. E., Moher, D., & Clifford, T. J. (2016). Quality of conduct and reporting in rapid reviews: An exploration of compliance with PRISMA and AMSTAR guidelines. Systematic Reviews, 5(1), 1–19. https://doi.org/10.1186/s13643-016-0258-9

Kleppmann, M. (2017). Designing Data-Intensive Applications. O'Reilly Media, Inc. Retrieved from https://www.oreilly.com/library/view/designing-data-intensive-applications/9781491903063/

Klimisch, H.-J., Andreae, M., & Tillmann, U. (1997). A Systematic Approach for Evaluating the Quality of Experimental Toxicological and Ecotoxicological Data. Regulatory Toxicology and Pharmacology, 25(1), 1–5. https://doi.org/10.1006/rtph.1996.1076

Kobrosly, R. W., Parlett, L. E., Stahlhut, R. W., Barrett, E. S., & Swan, S. H. (2012). Socioeconomic factors and phthalate metabolite concentrations among United States women of reproductive age. Environmental Research, 115, 11–17. https://doi.org/10.1016/j.envres.2012.03.008

Koch, L., & Ashford, N. A. (2006). Rethinking the role of information in chemicals policy: Implications for TSCA and REACH. Journal of Cleaner Production, 14(1), 31–46. https://doi.org/10.1016/j.jclepro.2005.06.003

Krauth, D., Woodruff, T. J., & Bero, L. (2013). Instruments for assessing risk of bias and other methodological criteria of published animal studies: A systematic review. Environmental Health Perspectives, 121(9), 985–992. https://doi.org/10.1289/ehp.1206389

Lau, J., Rothstein, H., & Stewart, G. (2013). History & progress of meta-analysis (Handbook of Meta-analysis in Ecology and Evolution). Princeton University Press. Retrieved from https://books.google.co.uk/books?id=Zh3GNd9M1oUC&printsec=frontcover&source=gbs_ge_summary_r&cad=0

Lefebvre, C. (1994). The Cochrane Collaboration: the role of the UK Cochrane Centre in identifying the evidence. Health Libraries Review, 11, 235–242.

Leisher, C., Temsah, G., Booker, F., Day, M., Samberg, L., Prosnitz, D., … Wilkie, D. (2016). Does the gender composition of forest and fishery management groups affect resource

governance and conservation outcomes? A systematic map. Environmental Evidence, 5(1), 1–10. https://doi.org/10.1186/s13750-016-0057-8

Lewis, K. A., Tzilivakis, J., Warner, D. J., & Green, A. (2016). An international database for pesticide risk assessments and management. Human and Ecological Risk Assessment, 22(4), 1050–1064. https://doi.org/10.1080/10807039.2015.1133242

Liu, Z. H., & Gawlick, D. (2015). Management of Flexible Schema Data in RDBMSs - Opportunities and Limitations for NoSQL -.

Lundh, A., & Gøtzsche, P. C. (2008). Recommendations by Cochrane Review Groups for assessment of the risk of bias in studies. BMC Medical Research Methodology, 8, 22. https://doi.org/10.1186/1471-2288-8-22

Lyndon, M. (1989). Information Economics and Chemical Toxicity : Designing Laws to Produce and Use Data. Michigan Law Review, 87(7), 1795–1861.

Macleod, M. R., Lawson McLean, A., Kyriakopoulou, A., Serghiou, S., de Wilde, A., Sherratt, N., … Sena, E. S. (2015). Risk of Bias in Reports of In Vivo Research: A Focus for Improvement. PLoS Biology, 13(10), 1–12. https://doi.org/10.1371/journal.pbio.1002273

Macura, B., Secco, L., & Pullin, A. S. (2015). What evidence exists on the impact of governance type on the conservation effectiveness of forest protected areas? Knowledge base and evidence gaps. Environmental Evidence, 4(1). https://doi.org/10.1186/s13750-015-0051-6

Mandrioli, D., Schlünssen, V., Ádám, B., Cohen, R. A., Colosio, C., Chen, W., … Scheepers, P. T. J. (2018). WHO/ILO work-related burden of disease and injury: Protocol for systematic reviews of occupational exposure to dusts and/or fibres and of the effect of occupational exposure to dusts and/or fibres on pneumoconiosis. Environment International, 119, 174–185. https://doi.org/10.1016/j.envint.2018.06.005

Manola, F., & Miller, E. (2004). RDF Primer, 1–107.

Mara-eves, A. O., Thomas, J., Mcnaught, J., Miwa, M., & Ananiadou, S. (2015). Using text mining for study identification in systematic reviews: a systematic review of current approaches. Systematic Reviews, 1–22. https://doi.org/10.1186/2046-4053-4-5

Markowitz, L. E., Hariri, S., Lin, C., Dunne, E. F., Steinau, M., Mcquillan, G., & Unger, E. R. (2013). Reduction in Human Papillomavirus (HPV) Prevalence Among Young Women Following HPV Vaccine Introduction in the United States , National Health and Nutrition Examination Surveys , 2003 – 2010, 208, 385–393. https://doi.org/10.1093/infdis/jit192

Marshall, I. J., & Wallace, B. C. (2019). Toward systematic review automation : a practical guide to using machine learning tools in research synthesis, 9, 1–10.

Marshall, I., Marshall, R., Wallace, B., Brassey, J., & Thomas, J. (2018). Rapid reviews may produce different results to systematic reviews: a meta-epidemiological study. Journal of Clinical Epidemiology. https://doi.org/10.1016/j.jclinepi.2018.12.015

Martin, O. V., Geueke, B., Groh, K. J., Chevrier, J., Fini, J.-B., Houlihan, J., … Muncke, J. (2018). Protocol for a systematic map of the evidence of migrating and extractable chemicals from food contact articles. https://doi.org/10.5281/zenodo.2525277

Martin, P., Bladier, C., Meek, B., Bruyere, O., Feinblatt, E., Touvier, M., … Makowski, D. (2018). Weight of evidence for hazard identification: A critical review of the literature. Environmental Health Perspectives, 126(7), 1–15. https://doi.org/10.1289/EHP3067

McKinnon, M. C., Cheng, S. H., Dupre, S., Edmond, J., Garside, R., Glew, L., … Woodhouse, E. (2016). What are the effects of nature conservation on human well-being? A systematic map of empirical evidence from developing countries. Environmental Evidence, 5(1), 1–25. https://doi.org/10.1186/s13750-016-0058-7

McMichael, A. J. (1976). Standardized Mortality Ratios and the “Healthy Worker Effect”: Scratching Beneath the Surface. Journal of Occupational and Environmental Medicine, 18(3). Retrieved from https://journals.lww.com/joem/Fulltext/1976/03000/Standardized_Mortality_Ratios_and_the__Healthy.9.aspx

Moja, L. P., Telaro, E., D’Amico, R., Moschetti, I., Coe, L., & Liberati, A. (2005). Assessment of methodological quality of primary studies by systematic reviews: results of the metaquality cross sectional study. Bmj, 330(7499), 1053. https://doi.org/10.1136/bmj.38414.515938.8F

Morgan, M. G. (2014). Use (and abuse) of expert elicitation in support of decision making for public policy. Proceedings of the National Academy of Sciences of the United States of America, 111(20), 7176–7184. https://doi.org/10.1073/pnas.1319946111

Morgan, R. L., Thayer, K. A., Bero, L., Bruce, N., Falck-Ytter, Y., Ghersi, D., … Schünemann, H. J. (2016). GRADE: Assessing the quality of evidence in environmental and occupational health. Environment International, 92–93, 611–616. https://doi.org/10.1016/j.envint.2016.01.004

Muntner, P., Menke, A., DeSalvo, K. B., Rabito, F. A., & Batuman, V. (2005). Continued Decline in Blood Lead Levels Among Adults in the United States. Archives of Internal Medicine, 165.

National Academy of Sciences. (1983). Risk Assessment in the Federal Government: Managing the Process. Washington, DC: National Academy Press. https://doi.org/10.17226/366

National Center for Biomedical Ontology. (2019). BioPortal. Retrieved from https://bioportal.bioontology.org/

National Centre for the Replacement Refinement & Reduction of Animals in Research. (2018). The 3Rs. Retrieved November 8, 2018, from https://www.nc3rs.org.uk/the-3rs

National Institute for Health Research. (2018). PROSPERO - International prospective register of systematic reviews. Retrieved December 31, 2018, from https://www.crd.york.ac.uk/prospero/

Neaves, L. E., Eales, J., Whitlock, R., Hollingsworth, P. M., Burke, T., & Pullin, A. S. (2015). The fitness consequences of inbreeding in natural populations and their implications for species conservation - A systematic map. Environmental Evidence, 4(1), 1–17. https://doi.org/10.1186/s13750-015-0031-x

Nelson, J. W., Scammell, M. K., Hatch, E. E., & Webster, T. F. (2012). Social disparities in exposures to bisphenol A and polyfluoroalkyl chemicals : a cross-sectional study within NHANES 2003-2006, 1–15.

Neo4j. (2019). Neo4j. Retrieved from https://neo4j.com/

Neo4j. (2020). Neo4j Bloom. Retrieved from http://neo4j.com/bloom/

Noy, N. F., & Klein, M. (2004). Ontology Evolution: Not the Same as Schema Evolution. Knowledge and Information Systems, 6(4), 428–440. https://doi.org/10.1007/s10115-003-0137-2

NTP. (2015). Handbook for conducting a literature-based health assessment using OHAT approach for systemic review and evidence integration, 1–98. Retrieved from https://ntp.niehs.nih.gov/ntp/ohat/pubs/handbookjan2015_508.pdf

NTP-OHAT. (2019). About the Office of Health Assessment and Translation. Retrieved from http://ntp.niehs.nih.gov/

O'Connor, S. R., Tully, M. A., Ryan, B., Bradley, J. M., Baxter, G. D., & McDonough, S. M. (2015). Failure of a numerical quality assessment scale to identify potential risk of bias in a systematic review: a comparison study. BMC Research Notes, 8, 224. https://doi.org/10.1186/s13104-015-1181-1

OHAT. (2015). OHAT Risk of Bias Rating Tool for Human and Animal Studies, 1–37. Retrieved from https://ntp.niehs.nih.gov/ntp/ohat/pubs/riskofbiastool_508.pdf

Oliver, S., Dickson, K. (2016). Policy-relevant systematic reviews to strengthen health systems: Models and mechanisms to support their production. Evidence and Policy, 12(2), 235–259. https://doi.org/10.1332/174426415X14399963605641

Ontotext. (2019a). Ontotext GraphDB. Retrieved from https://www.ontotext.com/products/graphdb/

Ontotext. (2019b). What is a Knowledge Graph? Retrieved from https://www.ontotext.com/knowledgehub/fundamentals/what-is-a-knowledge-graph/

Ontotext. (2019c). What is Inference? Retrieved from https://www.ontotext.com/knowledgehub/fundamentals/what-is-inference/

Ouzzani, M. (2017). Rayyan — a web and mobile app for systematic reviews. Systematic Reviews, (2016), 1–10. https://doi.org/10.1186/s13643-016-0384-4

Papathanasopoulou, E., Queirós, A. M., Beaumont, N., Hooper, T., & Nunes, J. (2016). What evidence exists on the local impacts of energy systems on marine ecosystem services: a systematic map. Environmental Evidence, 5(1), 1–12. https://doi.org/10.1186/s13750-016-0075-6

Pavlopoulos, G. A., Secrier, M., Moschopoulos, C. N., Soldatos, T. G., Kossida, S., Aerts, J., … Bagos, P. G. (2011). Using graph theory to analyze biological networks, 1–27.

Pega, F., Norris, S. L., Backes, C., Bero, L. A., Descatha, A., Gagliardi, D., Godderis, L., Loney, T., Modenese, A., Morgan, R. L., Pachito, D., Paulo, M. B. S., Scheepers, P. T. J., Schlünssen, V., Sgargi, D., Silbergeld, E. K., Sørensen, K., Sutton, P., Tenkate, T., … Mandrioli, D. (2020). RoB-SPEO : A tool for assessing risk of bias in studies estimating the prevalence of exposure to occupational risk factors from the WHO / ILO Joint Estimates of the Work-related Burden of Disease and Injury. Environment International, 135, https://doi.org/10.1016/j.envint.2019.105039

Pelch, K. E., Bolden, A. L., & Kwiatkowski, C. F. (2019). Environmental Chemicals and Autism: A Scoping Review of the Human and Animal Research. Environmental Health Perspectives, 127(4), 46001. https://doi.org/10.1289/EHP4386

Pelch, K. E., Reade, A., Wolffe, T. A. M., & Kwiatkowski, C. F. (2019). PFAS health effects database: Protocol for a systematic evidence map. Environment International, 130. https://doi.org/10.1016/j.envint.2019.05.045

Pool, Robert, Rusch, E. (2014). Identifying and Reducing Environmental Health Risks of Chemicals in Our Society: Workshop Summary. Washington, DC: The National Academies Press.

Project Jupyter. (2019). Jupyter. Retrieved from https://jupyter.org/

QlikTech International AB. (2019). Qlik Sense. Retrieved from https://www.qlik.com/us/products/qlik-sense

Randall, N. P., & James, K. L. (2012). The effectiveness of integrated farm management, organic farming and agri-environment schemes for conserving biodiversity in temperate Europe - A systematic map. Environmental Evidence, 1(4). https://doi.org/10.1046/j.1365-2664.2002.00695.x

Randall, N. P., Donnison, L. M., Lewis, P. J., & James, K. L. (2015). How effective are on-farm mitigation measures for delivering an improved water environment? A systematic map. Environmental Evidence, 4(1), 18. https://doi.org/10.1186/s13750-015-0044-5

Reynen, E., Robson, R., Ivory, J., Hwee, J., Straus, S. E., Pham, B., & Tricco, A. C. (2018). A retrospective comparison of systematic reviews with same-topic rapid reviews. Journal of Clinical Epidemiology, 96, 23–34. https://doi.org/10.1016/j.jclinepi.2017.12.001

Rhomberg, L. R., Goodman, J. E., Bailey, L. A., Prueitt, R. L., Beck, N. B., Bevan, C., … Becker, R. A. (2013). A survey of frameworks for best practices in weight-of-evidence analyses. Critical Reviews in Toxicology, 43(9), 753–784. https://doi.org/10.3109/10408444.2013.832727

Robinson, I., Webber, J., & Eifrem, E. (2015). Graph Databases. O'Reilly. https://doi.org/10.1016/B978-0-12-407192-6.00003-0

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), 711–718. https://doi.org/10.1289/ehp.1307972

Rooney, A. A., Cooper, G. S., Jahnke, G. D., Lam, J., Morgan, R. L., Boyles, A. L., Ratcliffe, J. M., Kraft, A. D., Schünemann, H. J., Schwingl, P., Walker, T. D., Thayer, K. A., & Lunn, R. M. (2016). How credible are the study results? Evaluating and applying internal validity tools to literature-based assessments of environmental health hazards. Environment International, 92–93, 617–629. https://doi.org/10.1016/j.envint.2016.01.005

Rothman, K. J., & Greenland, S. (1998). Modern Epidemiology (2nd Edition). Lippincott-Raven.

Rudén, C. (2001a). Interpretations of primary carcinogenicity data in 29 trichloroethylene risk assessments. Toxicology, 169(3), 209–225. https://doi.org/10.1016/S0300-483X(01)00525-X

Rudén, C. (2001b). The use and evaluation of primary data in 29 trichloroethylene carcinogen risk assessments. Regulatory Toxicology and Pharmacology, 34(1), 3–16. https://doi.org/10.1006/rtph.2001.1482

Rus, I., & Lindvall, M. (2002). Knowledge management in software engineering. IEEE Software, 19(3), 26–38. https://doi.org/10.1109/MS.2002.1003450

Rusu, D., Dali, L., Fortuna, B., Grobelnik, M., & Mladeni, D. (2014). Triplet extraction from sentences.

Sadowski, G., & Rathle, P. (2014). Fraud Detection: Discovering Connections with Graph Databases. Neo Technology, 17.

Samuel, G. O., Hoffmann, S., Wright, R. A., Lalu, M. M., Patlewicz, G., Becker, R. A., … Stephens, M. L. (2016). Guidance on assessing the methodological and reporting quality of toxicologically relevant studies: A scoping review. Environment International, 92–93, 630–646. https://doi.org/10.1016/j.envint.2016.03.010

Samwald, M., Jentzsch, A., Bouton, C., Kallesøe, C. S., Willighagen, E., Hajagos, J., … Stephens, S. (2011). Linked Open drug data for pharmaceutical research and development. Journal of Cheminformatics, 3(5), 1–6. https://doi.org/10.1186/1758-2946-3-19

Sanderson, S., Tatt, I. D., & Higgins, J. P. T. (2007). Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: A systematic review and annotated bibliography. International Journal of Epidemiology, 36(3), 666–676. https://doi.org/10.1093/ije/dym018

Sasaki, B. M., Chao, J., Howard, R., Sasaki, B. M., Chao, J., & Howard, R. (2018). Graph Databases for Beginners Graph Databases For Beginners.

Schaafsma, G., Kroese, E. D., Tielemans, E. L. J. P., Van de Sandt, J. J. M., & Van Leeuwen, C. J. (2009). REACH, non-testing approaches and the urgent need for a change in mind set. Regulatory Toxicology and Pharmacology, 53(1), 70–80. https://doi.org/10.1016/j.yrtph.2008.11.003

Schaefer, H. R., & Myers, J. L. (2017). Guidelines for performing systematic reviews in the development of toxicity factors. Regulatory Toxicology and Pharmacology, 91, 124–141. https://doi.org/10.1016/j.yrtph.2017.10.008

Schmidt, C. W. (2016). TSCA 2.0 - A New Era in Chemical Risk Management. Environmental Health Perspectives, 124(10), 182–186.

Schneider, K., Schwarz, M., Burkholder, I., Kopp-Schneider, A., Edler, L., Kinsner-Ovaskainen, A., … Hoffmann, S. (2009). "ToxRTool", a new tool to assess the reliability of toxicological data. Toxicology Letters, 189(2), 138–144. https://doi.org/10.1016/j.toxlet.2009.05.013

Schoeters, G. (2010). The reach perspective: Toward a new concept of toxicity testing. Journal of Toxicology and Environmental Health - Part B: Critical Reviews, 13(2–4), 232–241. https://doi.org/10.1080/10937404.2010.483938

Schultz, T. W., Cronin, M. T. D., Walker, J. D., & Aptula, A. O. (2003). Quantitative structure-activity relationships (QSARS) in toxicology: A historical perspective. Journal of Molecular Structure: THEOCHEM, 622(1–2), 1–22. https://doi.org/10.1016/S0166-1280(02)00614-0

Schwarzman, M., & Wilson, M. (2009). New science for chemicals policy. Science, 326(5956), 1065–1066. https://doi.org/10.1126/science.1177537

Science for Nature and People Partnership Evidence-Based Conservation working group, Conservation International, Datakind. (2018). Colandr. Retrieved January 14, 2019, from https://www.colandrapp.com/signin

Scientific Committee on Consumer Safety SCCS. (2019). Guidance on the Safety Assessment of nanomaterials in cosmetics. https://doi.org/10.2773/ISBN

Sciome. (2018). SWIFT-Review. Retrieved January 14, 2019, from https://www.sciome.com/swift-review/

Segaran, T., Evans, C., & Taylor, J. (2009). Programming the Semantic Web: Traditional Data-Modeling Methods. O'Reilly.

Sexton, K., & Hattis, D. (2007). Assessing cumulative health risks from exposure to environmental mixtures - Three fundamental questions. Environmental Health Perspectives, 115(5), 825–832. https://doi.org/10.1289/ehp.9333

Shamliyan, T., Kane, R. L., & Dickinson, S. (2010). A systematic review of tools used to assess the quality of observational studies that examine incidence or prevalence and risk factors for diseases. Journal of Clinical Epidemiology, 63(10), 1061–1070. https://doi.org/10.1016/j.jclinepi.2010.04.014

Silbergeld, E. K., Mandrioli, D., & Cranor, C. F. (2015). Regulating Chemicals: Law, Science, and the Unbearable Burdens of Regulation. Annual Review of Public Health, 36(1), 175–191. https://doi.org/10.1146/annurev-publhealth-031914-122654

Singhal, A. (2012). Introducing the Knowledge Graph: things, not strings. Retrieved from https://googleblog.blogspot.com/2012/05/introducing-knowledge-graph-things-not.html

Small, N. (2019). The Py2neo v4 Handbook. Retrieved from http://py2neo.org/v4/

Sobus, J. R., DeWoskin, R. S., Tan, Y. M., Pleil, J. D., Phillips, M. B., George, B. J., … Edwards, S. W. (2015). Uses of NHANES biomarker data for chemical risk assessment: Trends, challenges, and opportunities. Environmental Health Perspectives, 123(10), 919–927. https://doi.org/10.1289/ehp.1409177

Sola, P., Cerutti, P. O., Zhou, W., Gautier, D., Iiyama, M., Shure, J., … Shepherd, G. (2017). The environmental, socioeconomic, and health impacts of woodfuel value chains in Sub-Saharan Africa: A systematic map. Environmental Evidence, 6(1), 1–16. https://doi.org/10.1186/s13750-017-0082-2

Solomon, K. R., Wilks, M. F., Bachman, A., Boobis, A., Moretto, A., Pastoor, T. P., … Embry, M. R. (2016). Problem formulation for risk assessment of combined exposures to chemicals and other stressors in humans. Critical Reviews in Toxicology, 46(10), 835–844. https://doi.org/10.1080/10408444.2016.1211617

Stardog. (2019). Stardog. Retrieved from https://www.stardog.com/

Stone, B. M., & Reynolds, C. R. (2003). Can the National Health and Nutrition Examination Survey III (NHANES III) data help resolve the controversy over low blood lead levels and neuropsychological development in children ?, 18, 219–244. https://doi.org/10.1016/S0887-6177(03)00005-2

Tableau Software. (2019). Tableau. Retrieved from https://www.tableau.com/en-gb/trial/tableau-software?utm_campaign_id=2017049&utm_campaign=Prospecting-CORE-ALL-ALL-ALL-ALL&utm_medium=Paid+Search&utm_source=Google+Search&utm_language=EN&utm_country=UKI&kw=tableau&adgroup=CTX-Brand-Core-E-control&adused=3

Taboureau, O., & Audouze, K. (2017). Human environmental disease network: A computational model to assess toxicology of contaminants. Altex, 34(2), 289–300. https://doi.org/10.14573/altex.1607201

The Cochrane Collaboration. (2019). Cochrane UK. Retrieved from http://uk.cochrane.org/

The Endocrine Disruption Exchange. (2019). TEDX Publications. Retrieved from https://endocrinedisruption.org/interactive-tools/publications/

The Endocrine Disruption Exchange. (2019b). The Endocrine Disruption Exchange. Retrieved from http://endocrinedisruption.org/

The National Academies of Sciences. (2017). Application of Systematic Review Methods in an Overall Strategy for Evaluating Low-Dose Toxicity from Endocrine Active Chemicals. Washington, DC: The National Academies Press. https://doi.org/10.17226/24758

The W3C SPARQL Working Group. (2013). SPARQL 1.1 Overview. Retrieved from https://www.w3.org/TR/sparql11-overview/

Thomas J, Brunton J, G. S. (2010). EPPI-Reviewer 4.0: software for research synthesis. London: Social Science Research Unit, Institute of Education, University of London. Retrieved from http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=3299#Research impact

Thorn, J. P. R., Friedman, R., Benz, D., Willis, K. J., & Petrokofsky, G. (2016). What evidence exists for the effectiveness of on-farm conservation land management strategies for preserving ecosystem services in developing countries? A systematic map. Environmental Evidence, 5(1), 1–29. https://doi.org/10.1186/s13750-016-0064-9

Tyrrell, J., Melzer, D., Henley, W., Galloway, T. S., & Osborne, N. J. (2013). Associations between socioeconomic status and environmental toxicant concentrations in adults in the USA : NHANES 2001 – 2010 ☆. Environment International, 59, 328–335. https://doi.org/10.1016/j.envint.2013.06.017

U.S. National Library of Medicine. (2016). Unified Medical Language System (UMLS). Retrieved December 31, 2018, from https://www.nlm.nih.gov/research/umls/

UCSF Program on Reproductive Health and the Environment. (2019). Navigation Guide. Retrieved from https://prhe.ucsf.edu/navigation-guide

United States Environmental Protection Agency. (2016). The Frank R. Lautenberg chemical safety for the 21st century act. Retrieved from https://www.epa.gov/assessing-and-managing-chemicals-under-tsca/frank-r-lautenberg-chemical-safety-21st-century-act

van Altena, A. J., Spijker, R., Olabarriaga, S. D. (2019). Usage of automation tools in systematic reviews. Research Synthesis Methods, 10 (1), 72–82. https://doi.org/10.1002/jrsm.1335

van Leeuwen, K., Schultz, T. W., Henry, T., Diderich, B., & Veith, G. D. (2009). Using chemical categories to fill data gaps in hazard assessment. SAR and QSAR in Environmental Research, 20(3–4), 207–220. https://doi.org/10.1080/10629360902949179

Vandenberg, L. N., Ågerstrand, M., Beronius, A., Beausoleil, C., Bergman, Å., Bero, L. A., … Rudén, C. (2016). A proposed framework for the systematic review and integrated assessment (SYRINA) of endocrine disrupting chemicals. Environmental Health: A Global Access Science Source, 15(1), 1–19. https://doi.org/10.1186/s12940-016-0156-6

Villeneuve, D. L., Angrish, M. M., Fortin, M. C., Katsiadaki, I., Leonard, M., Margiotta-Casaluci, L., … Knapen, D. (2018). Adverse outcome pathway networks II: Network analytics. Environmental Toxicology and Chemistry, 37(6), 1734–1748. https://doi.org/10.1002/etc.4124

Vink, S. R., Mikkers, J., Bouwman, T., Marquart, H., & Kroese, E. D. (2010). Use of read-across and tiered exposure assessment in risk assessment under REACH – A case study on a phase-in substance. Regulatory Toxicology and Pharmacology, 58, 64–71. https://doi.org/10.1016/j.yrtph.2010.04.004

Vinken, M., Whelan, M., & Rogiers, V. (2014). Adverse outcome pathways: Hype or hope? Archives of Toxicology, 88(1), 1–2. https://doi.org/10.1007/s00204-013-1093-6

Walker, V. R., Boyles, A. L., Pelch, K. E., Holmgren, S. D., Shapiro, A. J., Blystone, C. R., … Rooney, A. A. (2018). Human and animal evidence of potential transgenerational inheritance of health effects: An evidence map and state-of-the-science evaluation. Environment International, 115, 48–69. https://doi.org/10.1016/j.envint.2017.12.032

Weaver, J., Tarjan, P. (2012). Facebook Linked Data via the Graph API, IOS Press, 1, 1–6.

Webber, J. (2018). Powering Real-Time Recommendations with Graph Database Technology.

Wells, G.A., Shea, B., O'Connell, D., Peterson, J., Welch, V., Losos, M., Tugwell, P. (2014). The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. Retrieved from http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp

Whaley, P., Halsall, C. (2016). Assuring high-quality evidence reviews for chemical risk assessment: Five lessons from guest editing the first environmental health journal special issue dedicated to systematic review. Environment International, 92–93, 553–555. https://doi.org/10.1016/j.envint.2016.04.016

Whaley, P., Halsall, C., Ågerstrand, M., Aiassa, E., Benford, D., Bilotta, G., … Taylor, D. (2016). Implementing systematic review techniques in chemical risk assessment: Challenges, opportunities and recommendations. Environment International, 92–93, 556–564. https://doi.org/10.1016/j.envint.2015.11.002

Wild, D. J., Ding, Y., Sheth, A. P., Harland, L., Gifford, E. M., & Lajiness, M. S. (2012). Systems chemical biology and the Semantic Web: What they mean for the future of drug discovery research. Drug Discovery Today, 17(9–10), 469–474. https://doi.org/10.1016/j.drudis.2011.12.019

Wolffe, T. A. M., Vidler, J., Halsall, C., Hunt, N., & Whaley, P. (2020). A survey of systematic evidence mapping practice and the case for knowledge graphs in environmental health & toxicology. Toxicological Sciences, 175 (1), 35-49. https://doi.org/10.1093/toxsci/kfaa025

Wolffe, T. A. M., Whaley, P., Halsall, C., Rooney, A. A., & Walker, V. R. (2019). Systematic evidence maps as a novel tool to support evidence-based decision-making in chemicals policy and risk management. Environment International, 130, https://doi.org/10.1016/j.envint.2019.05.065

Wood, L., Egger, M., Gluud, L. L., Schulz, K. F., Juni, P., Altman, D. G., … Sterne, J. A. C. (2008). Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. BMJ, 336(7644), 601–605. https://doi.org/10.1136/bmj.39465.451748.AD

Woodruff, T. J., & Sutton, P. (2011). An evidence-based medicine methodology to bridge the gap between clinical and environmental health sciences. Health Affairs, 30(5), 931–937. https://doi.org/10.1377/hlthaff.2010.1219

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), 1007–1014. https://doi.org/10.1289/ehp.1307175

World Health Organization. (2019). Framework for use of Systematic Review methods in chemical risk assessment – authors meeting. Retrieved from http://who.int/ipcs/events/SRmeeting_US/en/

Yankulov, M. (2019). Semantic Search for Smart Data Discovery in the Pharma Industry. Retrieved from http://ontotext.com/blog/semantic-search-for-smart-data-discovery-in-the-pharma-industry/

Zynda, M. R. (2013). The First Killer App: A History of Spreadsheets. Interactions, 68–72. https://doi.org/10.1145/2509224

Appendix: PFAS health effects database: Protocol for a systematic evidence map

The research presented in this Appendix was published in the journal *Environment International* in 2019, volume 130.

Environment International 130 (2019) 104851

Contents lists available at ScienceDirect

Environment International

journal homepage: www.elsevier.com/locate/envint

Review article

PFAS health effects database: Protocol for a systematic evidence map

Katherine E. Pelch[a,*], Anna Reade[b], Taylor A.M. Wolffe[c], Carol F. Kwiatkowski[a,d]

[a] The Endocrine Disruption Exchange, PO Box 54, Eckert, CO, USA
[b] Healthy People & Thriving Communities Program, Natural Resources Defense Council, 111 Sutter Street, Floor 21, San Francisco, CA, USA
[c] Lancaster Environment Centre, Lancaster University, Lancaster LA1 4YQ, UK
[d] Department of Biological Sciences, North Carolina State University, Raleigh, NC, USA

ARTICLE INFO

Edited by Hanna Boogaard

Keywords:
PFAS
Per- and polyfluoroalkyl substances
Systematic evidence map
Toxicology

ABSTRACT

Background: Per- and polyfluoroalkyl substances (PFAS) confer waterproof, greaseproof, and non-stick properties when added to consumer products. They are also used for industrial purposes including in aqueous film forming foams for firefighting. PFAS are ubiquitous in the environment, are widely detected in human biomonitoring studies, and are of growing regulatory concern across federal, state, and local governments. Regulators, scientists, and citizens need to stay informed on the growing health and toxicology literature related to PFAS.
Objectives: The goal of this systematic evidence map is to identify and organize the available health and toxicology related literature on a set of 29 PFAS of emerging and growing concern.
Search and study eligibility: We will search the electronic database PubMed for health or toxicological studies on 29 PFAS of emerging concern. Eligible studies must contain primary research investigating the link between one or more of the PFAS of interest and a health effect, toxicological, or biological mechanistic endpoint.
Study appraisal and synthesis methods: Title and abstract screening and full text review will require a single reviewer for inclusion to the next level and two independent reviewers for exclusion. Study quality will not be conducted for this evidence mapping. Study characteristics will be extracted and coded from the included studies and checked for accuracy by a second reviewer. The extracted and coded information will be visualized in a publicly available, interactive database hosted on Tableau Public. Results of the evidence mapping will be published in a narrative summary.

1. Introduction

1.1. Rationale

Over the past few decades per- and polyfluoroalkyl substances (PFAS) contamination has grown into a serious global health threat. PFAS are a large class of synthetic chemicals that contain an alkyl chain with at least one fully fluorinated carbon atom. Although the class is broad, they are related in their extreme persistence in our environment and are often referred to as “forever chemicals”. PFAS are also highly mobile in the environment and some have been found to bioaccumulate, or build up, in humans and animals.

Best known for their original use in producing the fluoropolymer Teflon and the stain-resistant coating Scotchgard, these chemicals are now used in a wide range of consumer and industrial products where grease or water proofing is desired, or surfactant action is a benefit. These products include food packaging and non-stick cookware, cosmetics, waterproof and stain-proof textiles and carpet, aqueous film forming foam (AFFF) to fight Class B fires, and as part of metal plating processes.

Widespread use of PFAS has resulted in the ubiquitous presence of these chemicals in the environment including in rivers, soil, air, house dust, food and drinking water from surface water and groundwater sources. Virtually all Americans have multiple PFAS at detectable levels in the blood serum (CDC, 2018). Unfortunately, PFAS have been linked to many harmful health effects, including cancer, immune system

Abbreviations: ADME/PK/TK, absorption, distribution, metabolism, excretion, pharmacokinetic or toxicokinetic properties; AFFF, aqueous film forming foam; AI, artificial intelligence; ATSDR, Agency for Toxic Substances and Disease Registry; COI, conflict of interest; EPA, US Environmental Protection Agency; hpf, hours post-fertilization; MCL, maximum contaminant level; NJDWQI, New Jersey Drinking Water Quality Institute; NTP, National Toxicology Program; PECO, populations, exposures, comparators, and outcomes; PFAS, per- and polyfluoroalkyl substances; PFBS, perfluorobutane sulfonic acid; PFOA, perfluorooctanoic acid; PFOS, perfluorooctanesulfonic acid; PND, postnatal day; PPAR, peroxisome proliferator activated receptor; ppt, part per trillion; QC, quality control

* Corresponding author.
E-mail addresses: katiepelch@tedx.org (K.E. Pelch), areade@nrdc.org (A. Reade), t.wolffe@lancaster.ac.uk (T.A.M. Wolffe), carolkw@tedx.org (C.F. Kwiatkowski).

https://doi.org/10.1016/j.envint.2019.05.045
Received 19 February 2019; Received in revised form 15 May 2019; Accepted 16 May 2019
Available online 05 July 2019

dysfunction, liver damage, developmental and reproductive harm, and hormone disruption (ATSDR, 2018).

The most well-known and well-studied PFAS are perfluorooctanoic acid (PFOA) and perfluorooctane sulfonic acid (PFOS) (ATSDR, 2018). Due to increasing concern over the harm these chemicals cause to human health, wildlife, and the environment, the U.S. Environmental Protection Agency initiated the PFOA Stewardship Program in 2006 (US EPA, 2006). Through the program, the major PFAS manufacturing companies committed to phasing out PFOA, its precursor chemicals and related higher homologue chemicals from production in the US by 2015; however, PFOA and PFOS are still produced internationally and consumer products containing PFOA and related-PFAS may still be imported into the US (US EPA, n.d.) This, in combination with their extreme persistence in the environment, ensures that their legacy remains.

The scientific literature on PFAS has increased exponentially in the last decade, which has resulted in a greater understanding of the potential adverse health effects associated with PFOA and PFOS exposure (Grandjean, 2018). For PFOA and PFOS this has resulted in increasingly stricter health thresholds proposed by various agencies (Cordner et al., 2019). In 2016 the EPA issued lifetime drinking water health advisories of 70 ppt for PFOA and PFOS, individually or combined (US EPA, 2016b, 2016c). Recently several states (*e.g.* MN, NH, NJ, VT, MI) have proposed drinking water regulatory or guideline levels below 70 ppt (Cordner et al., 2019; MDHHS, 2019; NHDES, 2019).

For various reasons, including uncertainties in data and biological significance, the EPA did not select the most sensitive health effects currently associated with PFOA and PFOS when generating their 2016 health advisories. There is evidence that both altered mammary gland development for PFOA (Macon et al., 2011; Tucker et al., 2015; White et al., 2011) and immunotoxicity for PFOS (Dong et al., 2009; Grandjean and Budtz-Jorgensen, 2013; Guruge et al., 2006; Peden-Adams et al., 2008) can occur at levels an order of magnitude or lower than the health effects selected by the EPA. Since the EPA issued its 2016 advisories, the National Toxicology Program (NTP) released a report concluding that both PFOA and PFOS are presumed to constitute immune hazards to humans (NTP, 2016). And most recently, the New Jersey Drinking Water Quality Institute (NJDWQI) and the Agency for Toxic Substances and Disease Registry (ATSDR) have either acknowledged or attempted to account for these more sensitive health effects in generating their proposed health standards (ATSDR, 2018; NJDWQI, 2017; NJDWQI, 2018). As a result, both NJDWQI and ATSDR have proposed significantly more protective (5–10 times lower) health thresholds for PFOA and PFOS than the EPA health advisories (ATSDR, 2018; NJDWQI, 2017; NJDWQI, 2018).

The expansion of research on PFAS has also resulted in increasing concern over the rising use of and exposure to replacements for legacy PFAS. Most legacy PFAS, including PFOA and PFOS, are "long-chain" chemicals, meaning their molecular structure contains a chain of six (for perfluoroalkyl sulfonic acids) or seven (for perfluoroalkyl carboxylic acids) or more carbon atoms. While there is less toxicity data on shorter-chain and other alternative PFAS replacing long-chain PFAS, evidence is growing quickly that indicates they collectively pose similar threats to human health and the environment; which, combined with similar concerns over the environmental fate and persistence, have led independent scientists and other professionals from around the globe to express concern about the continued and increasing production and release of PFAS (Blum et al., 2015; Scheringer et al., 2014).

Due to the health concerns related to PFAS exposure and concerns over their environmental fate and persistence, there have been various efforts at the local, state and federal level to regulate PFAS. For example, severe contamination of drinking water with both legacy and alternative PFAS in communities across the nation, has led to considerable efforts at the state-level to set enforceable drinking water maximum contaminant levels (MCLs). It is expected that efforts to regulate PFAS in drinking water (as well as in ground and surface waters, air, consumer products, *etc.*) will continue over the coming years. Staying abreast of the current PFAS health effects literature is a major barrier for setting effective regulations to protect human and environmental health. Further, as additional communities learn of their own PFAS contamination, there is a desire from citizens and citizen-led groups to know more about these chemicals and how they may impact the health of their communities.

The ATSDR Draft Toxicological Profile for Perfluoroalkyls provides estimates concerning the volume of available human and experimental animal studies through May 2016 for PFOA (Fig. 2-1; $n = 271$), PFOS (Fig. 2-2; $n = 218$) and 12 additional PFAS (Fig. 2-3; $n = 127$) (ATSDR, 2018). Though helpful, the figures provided by ATSDR do not allow the end-user much flexibility in sorting, filtering, or deeply exploring the available evidence. Additionally, ATSDR Fig. 2-3 presents the evidence for 12 PFAS of emerging interest, but it is not possible to determine how the identified studies are distributed among the chemicals, which limits its utility to state agencies proposing regulatory values for individual PFAS beyond PFOA and PFOS.

To this end, we will use systematic evidence mapping methodology to improve citizen, scientific and regulatory access to current evidence regarding the health effects associated with exposure to PFAS. Systematic evidence maps collate and characterise evidence available on a broad research topic. They distill a potentially vast, heterogeneous evidence base into a (computationally) accessible, comparable and easily updated format using transparent and reproducible methodology. Systematic evidence maps take the form of searchable databases of references and meta-data, including data extracted and coded from each individual included study. This format removes the barriers associated with manually assessing large volumes of data by affording end users a broad overview of the evidence base, allowing fast identification of emerging trends, including the presence of evidence gaps and evidence clusters (James et al., 2016). As such, systematic evidence maps do not attempt to synthesise or integrate evidence in answer to any one specific research question, but rather provide users with the means of exploring the evidence according to their own varied research interests - identifying trends which might form the basis of future syntheses or further research.

Here, we propose to create a systematic evidence map that transparently and systematically surveys the available health and toxicological evidence associated with PFAS exposure. The result will be an online, interactive, interrogable, and user-friendly database (Miake-Lye et al., 2016). Given the pace at which the evidence base appears to be growing, it would seem that now is a good time to establish a systematically and transparently created interactive database, such as the one proposed in this protocol. A database concerning the health effects of "short-chain PFAS" has been previously suggested, but to our knowledge has not yet been produced (Bowman, 2015).

1.2. Objectives

The objectives of this systematic evidence map are to:

1. Identify and organize the available scientific research on the physiological health effects of a set of 29 PFAS (see Table 1), individually or combined, as measured in human, animal, or *ex vivo/in vitro* models.
2. Present the literature in a user-friendly, online, interactive database that will connect end-users directly to referenced primary studies.
3. Identify data gaps and research needs, and publish a narrative summary of the systematic map.

The protocol described here, serves to document decisions made *a priori* regarding the conduct of the systematic evidence mapping.

Table 1
List of PFAS included in systematic evidence map

Abbreviation	Chemical name	CASRN
PFHxA	Perfluorohexanoic acid	307-24-4
PFHpA	Perfluoroheptanoic acid	375-85-9
PFNA	Perfluorononanoic acid	375-95-1
PFDA	Perfluorodecanoic acid	335-76-2
PFBS	Perfluorobutanesulfonic acid	375-73-5
PFHxS	Perfluorohexanesulfonic acid	355-46-4
PFUnA	Perfluoroundecanoic acid	2058-94-8
PFDoA	Perfluorododecanoic acid	307-55-1
NEtFOSAA	2-(N-ethyl-perfluorooctane sulfanamido) acetic acid	2991-50-6
NMeFOSAA	2-(N-Methyl-perfluorooctane sulfanamido) acetic acid	2355-31-9
GenX	Hexafluoropropylene Oxide (HFPO) Dimer Acid	13252-13-6
PFTA	Perfluorotetradecanoic acid	376-06-7
PFTrDA	Perfluorotridecanoic acid	72629-94-8
ADONA	4,8-dioxa-3H-perfluorononanoic	919005-14-4
6:2 Cl-PFESA	6:2 chlorinated polyfluorinated ether sulfonic acid	73606-19-6
8:2 Cl-PFESA	8:2 chlorinated polyfluorinated ether sulfonic acid	83329-89-9
PFBA	Perfluorobutanoic acid	375-22-4
PFPeA	Perfluoro-n-pentanoic acid	2706-90-3
Nafion BP2	Nafion Byproduct 2	749836-20-2
PFO4DA	Perfluoro-3,5,7,9-tetraoxadecanoic acid	39492-90-5
PFO5DoDA	Perfluoro-3,5,7,9,11-pentaoxadodecanoic acid	39492-91-6
Hydro-Eve	2,2,3,3-Tetrafluoro-3-((1,1,1,2,3,3-hexafluoro-3-(1,2,2,2-tetrafluoroethoxy)propan-2-yl)oxy) propanoic acid	773804-62-9
6:2 FTSA	h,1h,2h,2h-Perfluorooctanesulfonic acid	27619-97-2
8:2 FTSA	2-(Perfluorooctyl)ethane-1-sulfonic acid	39108-34-4
PFPeS	Perfluoropentanesulfonic acid	2706-91-4
PFHpS	Perfluoroheptanesulfonic acid	375-92-8
PFNS	Perfluorononanesulfonic acid	68259-12-1
PFDS	Perfluorodecanesulfonic acid	335-77-3
HFPO-TA	Hexafluoropropylene Oxide (HFPO) Trimer Acid	13252-14-7

2. Methods

This protocol has been prepared in accordance with the ENVINT PRISMA-SM-P report (available at (Elsevier, 2017)) and based on guidance from the Collaboration for Environmental Evidence (Collaboration for Environmental Evidence, 2018). The protocol has been registered at Zenodo (Pelch et al., 2019).

2.1. Information sources

PFAS (Table 1) were prioritized for inclusion in this systematic evidence map due to their inclusion in the ATSDR Draft Toxicological Profile for Perfluoroalkyls (ATSDR, 2018), their presence in US EPA Method 537.1 (Shoemaker and Tettenhorst, 2018), because they were reported to be detected in blood in the GenX exposure study (NC State Center for Human Health and the Environment, 2018a; NC State Center for Human Health and the Environment, 2018b), or because they were suggested to be of interest by members of the NGO community (personal communication). Because PFOA and PFOS have been recently reviewed by US EPA (US EPA, 2016a, 2016c), ATSDR (ATSDR, 2018), and NTP (NTP, 2016), they were not prioritized for incorporation in this systematic evidence map.

The peer-reviewed published literature will be identified by searching PubMed electronic database with no date or language restrictions. If a search update is needed, the PubMed search will be repeated but limited to studies published since the date of the last search using the "date-create" field in the PubMed Advanced Search Builder. The number of studies retrieved from searching will be tracked in a study flow diagram (*e.g.* Fig. 1), which will also track how the studies progress through the review. Any studies identified from sources other than PubMed (*e.g.* identified by hand searching included studies or relevant reviews) will be marked as "Identified from other sources" on the study flow diagram.

2.2. Search strategy

The Pubmed search will include names and synonyms for 29 PFAS of emerging interest. Specific search terms can be found in Appendix 1. There will be no search limitations based on health outcome or other aspects of study design or conduct. Furthermore, the search will be conducted without limit on publication year or language.

Search terms were identified for the PFAS of interest by searching the CASRN for each chemical, the common abbreviation, and full chemical names, which have been identified as synonyms for the chemical in PubChem. The search logic for GenX and PFBS are adapted from the recent EPA GenX and PFBS Draft Toxicity Assessments (US EPA, 2018a, 2018b). The search logic for PFAS in general has been adapted from the search logic used in the NTP monograph (NTP, 2016). When possible, the search will also include CASRN and relevant search terms for associated salts (see Table 2).

2.3. Eligibility criteria

Study eligibility is based on the PECO statement provided in Table 2.

To be included in this systematic evidence map, studies must contain primary research investigating the link between one or more of the PFAS of interest and a health effect, toxicological, or biological mechanistic endpoint. Epidemiological, animal, and *in vitro* and mechanistic evidence will be included. Studies that do not contain health, toxicological, or mechanistic information on the PFAS of interest will be excluded at the title and abstract level and will not be further data extracted.

Studies that investigate aspects of PFAS other than health outcomes will be tagged and categorized as to the nature of the evidence and may be made available upon request or as a downloadable list on the TEDX website (www.tedx.org). This includes studies on environmental detection, environmental fate and transport, biomonitoring, detection in wildlife, reports on the absorption, distribution, metabolism, excretion, pharmacokinetic or toxicokinetic properties (ADME/PK/TK), *in silico* and read across analyses, reviews, and systematic reviews of the PFAS of interest. Though they will be tagged and collated, studies that lack health outcome endpoints will not proceed past title and abstract screening.

Given that this is a systematic evidence map rather than a systematic review, efforts will be made to include non-English language studies if essential information (*i.e.* chemicals tested and health outcomes assessed) can be obtained from the title and abstract. Non-English studies will be denoted with square brackets on the title. Conference abstracts, presentations, posters, and theses/dissertations will not be included in this systematic evidence map.

2.4. Data management

2.4.1. Management of literature updates and study flow diagram

A study flow diagram will be maintained that describes the number of studies evaluated in each step of the review (Fig. 1). Any search updates or modifications to the protocol will also be noted as amendments to the registered protocol.

Literature search results will be imported to EndNote X6. Duplicate records will be identified using EndNote's "Find Duplicates" feature based on title and author fields. All records will receive a unique identification number upon import to EndNote X6 that will be maintained throughout the review. Records will then be exported and uploaded to DistillerSR (Evidence Partners; Ottawa, Ontario, Canada). Customized forms in DistillerSR will be used to manually screen studies at the title and abstract level and to extract study details from full-text documents. Extracted information will be exported from DistillerSR to one of three .csv files that can be directly uploaded to Tableau Desktop Professional Edition *vs* 2018.3 (Tableau; Seattle, WA) for visualization.

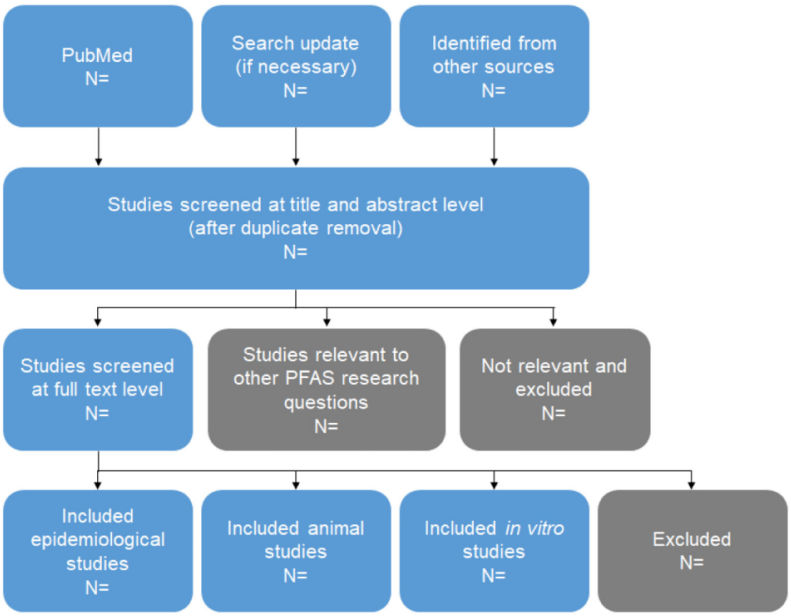

Fig. 1. Example study flow diagram
The example study flow diagram shows how studies will proceed through the review.

Table 2
Populations, **E**xposures, **C**omparators, and **O**utcomes (PECO) Statement.

PECO element	Evidence
Populations	Any human, animal (whole organism including experimental and observational studies), or *ex vivo/in vitro* models utilizing organs, tissues, cell lines, or cellular components (*e.g.* cell-free receptor binding assays).
Exposures	Exposure to at least one of the PFAS or the associated salts listed in Table 1 (*e.g.* perfluorobutane sulfonic acid (PFBS; CASRN 375-73-5) and potassium perfluorobutane sulfonate (K + PFBS; CASRN 29420-49-3)). Exposures may include, for example: biomarkers of exposure, modeling of potential exposures, and/or administered exposures. Mixtures of PFAS will also be included and listed as $PFAS_{mix}$. There are no limitations on the timing, route, level, or determination of estimated exposure.
Comparators	Humans, animals, organs, tissues, cell lines, or cellular components exposed to a lower level of a PFAS than the more highly exposed subjects or treatment groups, or vehicle-only treatment.
Outcomes	Any health outcome or type of biological response measured in the exposed population.

The three .csv files will represent the three evidence streams: human, animal, and *in vitro*. The .csv files will also be submitted as supplemental files to the journal with the final report.

The systematic evidence map will be hosted on TEDX's public profile on Tableau Public, which is available at https://public.tableau.com/profile/the.endocrine.disruption.exchange#!/. A link to the visualization will also be found on the TEDX website along with additional systematic evidence map details including links to the published and registered protocols.

2.5. Selection and data collection processes

Title and abstract screening will be performed in DistillerSR by senior researchers (KEP, AR, TW), none of which have authored peer reviewed articles that would be relevant for inclusion in this systematic evidence map. DistillerSR's artificial intelligence (AI) text mining functionality may be utilized to prioritize studies for title and abstract screening. Title and abstract screening and full text review will require a single reviewer for inclusion to the next level and two independent reviewers for exclusion. Discrepant screening results will be resolved by discussion. Likewise, full text review, data extraction, and coding will be conducted by a single reviewer with a secondary reviewer confirming the accuracy and completeness of extracted and coded data using DistillerSR's quality control (QC) feature. We will attempt to contact study authors *via* email if it is unclear which PFAS was investigated (*e.g.* missing CASRN or structure, or ambiguous chemical name). Other missing information will be flagged as missing, but study authors will not be contacted. Prior to commencing the search, DistillerSR forms will be piloted by KEP, AR, and TW on a small set of studies to ensure ease and accuracy of data extraction and export for visualization in Tableau.

2.6. Data coding strategy

Data extraction will be conducted on full-text studies using structured forms in DistillerSR. The following information will be collected from all included studies: authors, journal, reference information, year of publication, which evidence streams were investigated (human, animal, or *in vitro)*, conflict of interest statement (COI), funding statement, acknowledgements statement, chemicals evaluated, and the health outcome category (see Table 3). Data specific to each evidence stream will also be collected as outlined in Table 3. All data will be captured at the study level rather than at the level of each individual endpoint. In other words, for each study, data extractors will be instructed to select all responses that apply to each question.

Table 3
Data coding and recording.

Data category	Data captured
Bibliographic information	● authors ● year of publication ● journal ● title ● reference information ● study URL ● COI statement ● authors' acknowledgments statement ● funding source
Evidence stream	*Evidence stream is defined by the type of subject or population being exposed to the chemical.* ● Human epidemiological studies ● Animal (including experimental and observational whole animal studies) ● *In vitro* (includes mechanistic studies in humans and other species, *ex vivo*, and cell free)
Health effects studied	*Health outcomes will be tagged as follows (these headings were derived from the MedLinePlus ontology, which is available with definitions from the Unified Medical Language Systems Database (US NLM (United States National Library of Medicine), 2016):* ● Blood, heart, and circulation ● Bones, joints, and muscles ● Brain and nerves ● Cancers ● Digestive system ● Ear, nose, and throat ● Endocrine system ● Eyes and vision ● Female reproductive system ● Genetics/birth defects ● Immune system ● Injuries and wounds ● Kidneys and urinary system ● Lungs & Breathing ● Male reproductive system ● Mental health and behavior ● Metabolic problems ● Mouth and teeth ● Mortality ● Pregnancy and reproduction ● Sexual health issues ● Skin, hair, and nails
Chemicals studied	Data will be collected on the 29 PFAS listed in Table 1. If PFAS other than those listed in Table 1 are studied in included studies, they will be permanently added to the list of options so that they might be tracked for any future updates or expansions to this systematic evidence map. Mixtures of PFAS or Σ_{PFAS} presented in a study will be categorized as $PFAS_{mix}$ in addition to the component PFAS.
Human study elements	Study type: ● Case control ● Cohort ● Cross-sectional ● Ecological/community Study location: ● US (list US state abbreviation) ● Non-US ● *The city, state, and/or country of study location will be captured as free text* Exposure type: ● General population ● Known or suspected point source pollution ● Occupational Study population sex: ● Male ● Female ● Both Study N: ● *The study N will be collected as free text for the total number of study participants (e.g. all cases and controls)* Timing of exposure assessment: ● *The timing of exposure according to study authors will be captured as free text and will also be further categorized as:* ● Preconception ● Pregnancy ● birth-1 years of age ● > 1–3 years of age ● > 3–12 years of age ● > 12–20 years of age ● > 20 years of age Exposure assessment: ● *The exposure assessment method as described by the study authors will be captured as free text and will also be further categorized as follows, with controlled additions allowed as needed:* ● Adipose tissue ● Amniotic fluid ● Breast milk ● Cord blood ● Distance to source ● Drinking water concentration ● Hair ● Nails ● Serum ● Urine ● Whole blood Exposure level: ● Minimum reported exposure ● Maximum reported exposure ● Reported units of measured exposures Timing of outcome assessment: ● *The timing of outcome assessment according to study authors will be captured as free text and will also be further categorized as:* ● Pregnancy ● Birth −1 years of age ● > 1–3 years of age ● > 3–12 years of age ● > 12–20 years of age ● > 20 years of age
Animal study elements	Animal subjects: ● Species - species will be categorized as follows, with controlled additions allowed as needed: ○ Daphnia ○ Monkey ○ Mouse ○ Rat ○ Frog ○ Fish ● Strain - will be captured as free text Study population sex: ● Male ● Female ● Both Study N: ● *The study N will be collected as free text for the range of N from different experimental groups assessed throughout the study* Timing of exposure: ● *The timing of exposure according to study authors will be captured as free text and will also be further categorized as:* ● For rodents:

(*continued on next page*)

Table 3 (*continued*)

Data category	Data captured
	○ Gestational ○ Postnatal (for rodents postnatal day (PND)0-PND14) ○ Developmental (gestational + postnatal) ○ Juvenile (for rodents PND15–40) ○ Adult (for rodents PND41 +) ● For zebrafish: ○ Embryonic (hpf 0–72) ○ Larval (hpf 72–30 days) ○ Adult (> 30 days) ● For other model systems: ○ Will develop as needed with expert consultation Route of exposure: ● *The exposure assessment method as described by the study authors will be categorized as follows, with controlled additions allowed as needed:* ● Inhalation ● Intraperitoneal injection ● Embryonic injection (*e.g.* zebrafish, xenopus) ● Subcutaneous: injection ● Subcutaneous: mini osmotic pump ● Subcutaneous: silastic capsule ● Oral: drinking water ● Oral: gavage ● Oral: feed/diet/treat ● In treatment water (*e.g.* zebrafish, xenopus) ● Dermal ● Ocular Exposure assessment: ● *When relevant (i.e. observational animal studies), the exposure assessment method as described by the study authors will be categorized as follows, with controlled additions allowed as needed:* ● Adipose tissue ● Amniotic fluid ● Breast milk ● Cord blood ● Feces ● Hair ● Nails ● Serum ● Urine ● Whole blood ● Whole organism Exposure/dose range: ● Minimum reported exposure/dose ● Maximum reported exposure/dose ● Reported units of measured exposures/dose Timing of assessment: ● *The timing of outcome assessment according to study authors will be captured as free text and will also be further categorized as:* ● For rodents: ○ Gestational ○ Postnatal (for rodents PND0-PND14) ○ Juvenile (for rodents PND15–40) ○ Adult (for rodents PND41 +) ● For zebrafish: ○ Embryonic (hpf 0–72) ○ Larval (hpf > 72–30 days) ○ Adult (> 30 days +) ● For other model systems: ○ Will develop as needed with expert consultation
In vitro study elements	Cell species: ● *Cell species will be categorized as follows, with controlled additions allowed as needed:* ● Chicken ● *E. coli* ● Frog ● Guinea pig ● Hampster ● Human ● Mouse ● Rabbit ● Rat ● Yeast ● Zebrafish Cell line name: ● *Example cell line names are provided below. Controlled additions to this list will be allowed as needed:* ● 3T3L-1 ● BG-1 ● CHO ● COS-7 ● DT40 ● GH3 ● H295R ● HeLa ● HepaRG ● HepG2 ● Ishikawa ● MCF-7 ● MDA-kb2 ● NIH3T3 ● PC3 ● PZFH ● U2OS ● ZLF Cell type: ● *Example cell types are provided below. Controlled additions to this list will be allowed as needed:* ● Leukocytes ● Oocytes ● Neuronal ● Kidney ● Breast cancer ● Normal breast Exposure timing: ● *The range of exposure lengths used for the various experiments in a study will be recorded as free-text.* Endpoint description: ● *In vitro endpoints will be broadly categorized. Examples of broad categories are provided below. Controlled additions to this list will be allowed as needed:* ● Estrogen related ● Androgen related ● Thyroid related ● Glucocorticoid related ● Peroxisome proliferator activated receptor (PPAR) related ● Cell and molecular dysfunction (*e.g.* oxidative stress) Dose range: ● Minimum reported dose ● Maximum reported dose ● Reported units of measured exposures

2.7. Data mapping method

Studies will be collated by evidence stream, PFAS studied, and health outcome. The systematic evidence map will be hosted on TEDX's public profile on Tableau Public, which is available at https://public.tableau.com/profile/the.endocrine.disruption.exchange#!/. An example of how the data will be presented in shown in Fig. 2.

The display in Tableau Public will be an interactive evidence map that contains an evidence map as shown in Fig. 2, a list of all included studies, and a filter to limit the display based on evidence stream. In the freely available, online interactive display, it will be possible to filter the data to only see the studies for selected evidence streams, health

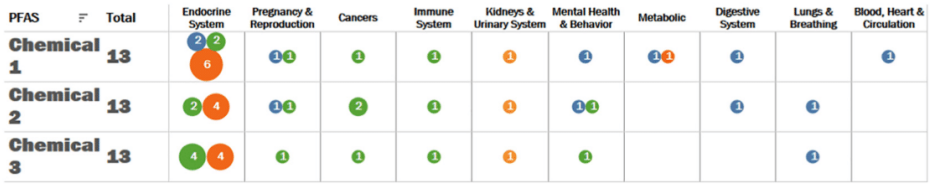

Fig. 2. Example evidence mapping
The example evidence mapping shows one aspect of how the data is expected to be presented in Tableau Public. In this example the different colored circles represent the three different evidence streams (human, animal, *in vitro*). The size of and number in each circle represents the number of studies for that specific chemical and health outcome category in that evidence stream. The rows are each different PFAS chemicals and the columns are different health outcome categories. A list of included studies is presented in another panel of the interactive figure not shown here.

outcome categories or chemicals. Users will be able to easily identify papers of interest by clicking on one of the colored circles to see a list of only those papers evaluating that specific PFAS and health outcome category. Users will be able to find additional study details (*e.g.* timing of exposure and outcome assessment, conflict of interest statement, *etc.*) and read the abstract by hovering over the name of the study in the study list. Further, clicking on a study of interest will take the user directly to the PubMed entry (or the entry on the publisher's page if the paper is not in PubMed).

2.8. Study quality assessment

Study quality will not be assessed in this systematic evidence map.

2.9. Synthesis of results

Results of this systematic evidence map will be summarized narratively and prepared as a manuscript for peer review. We anticipate discussing the overall results of the literature search (to be described in the study flow diagram, Fig. 1) and providing an analysis of the trends in PFAS publications by year. A list/lists of studies that investigate aspects of PFAS other than health outcomes (*i.e.* environmental detection, environmental fate and transport, biomonitoring, detection in wildlife, reports on the ADME/PK/TK, *in silico* and read across analyses, reviews, and systematic reviews) for the 29 PFAS of interest may be made available upon request or as a downloadable list on the TEDX website (www.tedx.org). The human evidence will be discussed in terms of chemicals evaluated to-date, the frequency of use of different study types and locations of the studies, the frequency of use and timing of various exposure assessments, the ranges of reported exposures and the different health outcomes evaluated to-date. The animal evidence will be discussed similarly but separately for observational studies and experimental studies, and will include a discussion on the chemicals studied to-date, the frequency of study of different species, and different experimental aspects including the timing, route, and level of exposure and health outcomes evaluated. The *in vitro* evidence will be discussed in terms of the chemicals and exposure levels studied to-date, the cell or model systems used, and different types of questions addressed by the *in vitro* studies.

Financial support

Financial support for this work was provided by the Arkansas Community Foundation, Winslow Foundation, Marisla Foundation and Tides Foundation. TAMW is financially supported by the Centre for Global Eco-Innovation (a programme funded by the European Regional Development Agency) and Yordas Group, a consultancy based at Lancaster University. AR is employed by the Natural Resources Defense Council, a nonprofit environmental advocacy group.

The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Author contributions

KEP and CFK conceived the protocol. KEP, AR, TAMW scoped the project and wrote the first draft. KEP, AR, TAMW, CFK reviewed and revised the protocol for submission and in response to reviewer requests.

Acknowledgements

Thank you to Paul Whaley for helpful conversations about systematic evidence maps, to Michele Dedeo for helpful conversations regarding search terminology, to Michelle Angrish for helpful conversations about health outcome ontologies, and Keshia Rose from Tableau Service Corps for guidance on creating the interactive web browser.

Declaration of Competing Financial Interests

The authors declare they have no actual or potential competing financial interests.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envint.2019.05.045.

References

ATSDR (Agency for Toxic Substances and Disease Registry), 2018. Toxicological Profile for Perfluoroalkyls (Draft for Public Comment). Department of Health and Human Services, Public Health Service. https://www.atsdr.cdc.gov/toxprofiles/tp.asp?id=1117&tid=237, Accessed date: 2 November 2018.

Blum, A., Balan, S.A., Scheringer, M., Trier, X., Goldenman, G., Cousins, I.T., Diamond, M., Fletcher, T., Higgins, C., Lindeman, A.E., Peaslee, G., de Voogt, P., Wang, Z., Weber, R., 2015. The Madrid statement on poly- and perfluoroalkyl substances (PFASs). Environ. Health Perspect. 123, A107–A111.

Bowman, J.S., 2015. Fluorotechnology is critical to modern life: the FluoroCouncil counterpoint to the Madrid statement. Environ. Health Perspect. 123, A112–A113.

CDC (Centers for Disease Control and Prevention), 2018. Fourth National Report on Human Exposure to Environmental Chemicals Updated Tables, March 2018. https://www.cdc.gov/exposurereport/pdf/fourthreport_updatedtables_volume1_mar2018.pdf, Accessed date: 20 January 2019.

Collaboration for Environmental Evidence, 2018. Guidelines and Standards for Evidence Synthesis in Environmental Management. www.environmentalevidence.org/information-for-authors, Accessed date: 9 January 2019.

Cordner, A., De La Rosa, V.Y., Schaider, L.A., Rudel, R.A., Richter, L., Brown, P., 2019. Guideline levels for PFOA and PFOS in drinking water: the role of scientific uncertainty, risk assessment decisions, and social factors. J. Expo. Sci. Environ. Epidemiol. 29, 157–171.

Dong, G.H., Zhang, Y.H., Zheng, L., Liu, W., Jin, Y.H., He, Q.C., 2009. Chronic effects of perfluorooctanesulfonate exposure on immunotoxicity in adult male C57BL/6 mice. Arch. Toxicol. 83, 805–815.

Elsevier, 2017. Guidance notes for authors of systematic reviews, systematic maps and other related manuscripts. Elsevierhttps://www.elsevier.com/journals/environment-international/0160-4120/guidance-notes, Accessed date: 9 January 2019.

Grandjean, P., 2018. Delayed discovery, dissemination, and decisions on intervention in

environmental health: a case study on immunotoxicity of perfluorinated alkylate substances. Environ. Health 17, 62.
Grandjean, P., Budtz-Jorgensen, E., 2013. Immunotoxicity of perfluorinated alkylates: calculation of benchmark doses based on serum concentrations in children. Environ. Health 12, 35.
Guruge, K.S., Yeung, L.W., Yamanaka, N., Miyazaki, S., Lam, P.K., Giesy, J.P., Jones, P.D., Yamashita, N., 2006. Gene expression profiles in rat liver treated with perfluorooctanoic acid (PFOA). Toxicol. Sci. 89, 93–107.
James, K.L., Randall, N.P., Haddaway, N.R., 2016. A methodology for systematic mapping in environmental sciences. Environ. Evid. 5.
Macon, M.B., Villanueva, L.R., Tatum-Gibbs, K., Zehr, R.D., Strynar, M.J., Stanko, J.P., White, S.S., Helfant, L., Fenton, S.E., 2011. Prenatal perfluorooctanoic acid exposure in CD-1 mice: low-dose developmental effects and internal dosimetry. Toxicol. Sci. 122, 134–145.
MDHHS (Michigan Department of Health and Human Services). Public Health Drinking Water Screening Levels for PFAS https://www.michigan.gov/documents/pfasresponse/MDHHS_Public_Health_Drinking_Water_Screening_Levels_for_PFAS_651683_7.pdf (accessed 10 April 2019). Division of Environmental Health 2019.
Miake-Lye, I.M., Hempel, S., Shanman, R., Shekelle, P.G., 2016. What is an evidence map? A systematic review of published evidence maps and their definitions, methods, and products. Syst. Rev. 5, 28.
NC State Center for Human Health and the Environment. The GenX Exposure Study. https://chhe.research.ncsu.edu/the-genx-exposure-study/ (accessed 22 January 2019). North Carolina State University; 2018a.
NC State Center for Human Health and the Environment, 2018b. GenX Exposure Study: PFAS Blood Sample Results. https://chhe.research.ncsu.edu/wordpress/wp-content/uploads/2018/11/Community-event-BLOOD-slides.pdf, Accessed date: 22 January 2019.
NHDES (New Hampshire Department of Enviornmental Services), 2019. Summary report on the New Hampshire Department of Environmental Services development of maximum contaminant levels and ambient groundwater quality standards for perfluorooctanesulfonic acid (PFOS), perfluorooctanoic acid (PFOA), perfluorononanoic acid (PFNA), and perfluorohexanesulfonic acid (PFHxS). https://www.des.nh.gov/organization/commissioner/pip/publications/documents/r-wd-19-01.pdf, Accessed date: 4 April 2019.
NJDWQI (New Jersey Drinking Water Quality Institute), 2018. Maximum Contaminant Level Recommendation for Perfluorooctane Sulfonate in Drinking Water. https://www.state.nj.us/dep/watersupply/pdf/pfos-recommendation-summary.pdf, Accessed date: 26 October 2018.
NJDWQI, 2017. Maximum Contaminant Level Recommendation for Perfluorooctanoic Acid in Drinking Water. https://www.nj.gov/dep/watersupply/pdf/pfoa-recommend.pdf, Accessed date: 26 October 2018.
NTP (National Toxicology Program). Monograph on Immunotoxicity Associated with Exposure to Perfluorooctanoic Acid (PFOA) and Perfluorooctane Sulfonate (PFOS). Office of Health Assessment and Translation. Research Triangle Park, NC: National Toxicology Program; https://ntp.niehs.nih.gov/ntp/ohat/pfoa_pfos/pfoa_pfosmonograph_508.pdf (accessed 5 November 2018). 2016.
Peden-Adams, M.M., Keller, J.M., Eudaly, J.G., Berger, J., Gilkeson, G.S., Keil, D.E., 2008. Suppression of humoral immunity in mice following exposure to perfluorooctane sulfonate. Toxicol. Sci. 104, 144–154.
Pelch, K.E.; Reade, A.; Wolffe, T.; Kwiatkowski, C.F. PFAS Health Database: A Protocol for a Systematic Evidence Map. https://zenodo.org/record/2832470#.XNt2Lo5Ki70 (accessed 14 May 2019). Zenodo; 2019.
Scheringer, M., Trier, X., Cousins, I.T., de Voogt, P., Fletcher, T., Wang, Z., Webster, T.F., 2014. Helsingor statement on poly- and perfluorinated alkyl substances (PFASs). Chemosphere 114, 337–339.
Shoemaker, J.A., Tettenhorst, D.R., 2018. In: Development O.o.R.a (Ed.), Method 537.1: Determination of Selected per- and Polyflrorinated Alkyl Substances in Drinking Water by Solid Phase Extraction and Liquid Chromatography/Tandem Mass Spectrometry (LC/MS/MS), Cincinnati, OH.
Tucker, D.K., Macon, M.B., Strynar, M.J., Dagnino, S., Andersen, E., Fenton, S.E., 2015. The mammary gland is a sensitive pubertal target in CD-1 and C57Bl/6 mice following perinatal perfluorooctanoic acid (PFOA) exposure. Reprod. Toxicol. 54, 26–36.
US EPA, 2006. PFOA Stewardship Program. Risk Management for Per- and Polyfluoroalkyl Substances (PFASs) under TSCA. United States Environmental Protection Agency. https://www.epa.gov/assessing-and-managing-chemicals-under-tsca/risk-management-and-polyfluoroalkyl-substances-pfass#tab-3, Accessed date: 27 April 2019.
US EPA (United States Environmental Protection Agency), 2016a. Drinking Water Health Advisory for Perfluorooctane Sulfonate (PFOS). Office of Water, Washington, D.C.. https://www.epa.gov/sites/production/files/2016-05/documents/pfos_health_advisory_final_508.pdf, Accessed date: 20 January 2019.
US EPA, 2016b. Drinking Water Health Advisory for Perfluorooctanesulfonate (PFOS). Office of Water, Washington DC. https://www.epa.gov/sites/production/files/2016-05/documents/pfos_health_advisory_final-plain.pdf, Accessed date: 8 January 2019.
US EPA, 2016c. Drinking Water Health Advisory for Perfluorooctanoic Acid (PFOA). Office of Water, Washington DC. https://www.epa.gov/sites/production/files/2016-05/documents/pfoa_health_advisory_final-plain.pdf, Accessed date: 8 January 2019.
US EPA, 2018a. Toxicity Assessment: Human Health Toxicity Values for Hexafluoropropylene Oxide (HFPO) Dimer Acid and its Ammonium Salt (CASRN 13252-13-6 and CASRN 62037-80-3). Office of Water, Washington DC. https://www.epa.gov/sites/production/files/2018-11/documents/genx_public_comment_draft_toxicity_assessment_nov2018-508.pdf, Accessed date: 10 January 2019.
US EPA, 2018b. Toxicity Assessment: Human Health Toxicity Values for Perfluorobutane Sulfonic Acid (CASRN 375-73-5) and Related Compound Potassium Perfluorobutane Sulfonate (CASRN 29420-49-3). Office of Water, Washington DC. https://www.epa.gov/sites/production/files/2018-11/documents/pfbs_public_comment_draft_toxicity_assessment_nov2018-508.pdf, Accessed date: 10 January 2019.
US EPA. Basic Information on PFAS. PFOA, PFOS and Other PFASs: United States Environmental Protection Agency https://www.epa.gov/pfas/basic-information-pfas (accessed 27 April 2019).
US NLM (United States National Library of Medicine). Unified Medical Language System (UMLS). Bethesda, MD: United States National Library of Medicine; https://www.nlm.nih.gov/research/umls/ (accessed 10 January 2019). 2016.
White, S.S., Stanko, J.P., Kato, K., Calafat, A.M., Hines, E.P., Fenton, S.E., 2011. Gestational and chronic low-dose PFOA exposures and mammary gland growth and differentiation in three generations of CD-1 mice. Environ. Health Perspect. 119, 1070–1076.

Supplemental Materials

for

PFAS Health Effects Database: A Protocol for a Systematic Evidence Map

Authors: Katherine Pelch, Anna Reade, Taylor Wolffe, Carol Kwiatkowski

1.0. Appendix 1. Search strings for PubMed

Search #	Search String	# records retrieved 5/13/19
#1	375-22-4[rn] OR PFBA[tw] OR "Perfluorobutyric acid"[nm] OR Perfluorobutanoic[tw] OR Heptafluorobutanoic[tw] OR Heptafluorobutyric[tw] OR heptaflurorbutyric[tw] OR Perfluorobutyric[tw] OR "Heptafluoro-1-butanoic"[tw] OR Perfluoropropanecarboxylic[tw] OR "heptafluoro-butanoic"[tw] OR "Heptafluoro-n-butyric"[tw] OR Heptafluorobutyricacid[tw] OR (c4 [tw] AND perfluorinated [tw])	841
#2	2706-90-3[rn] OR PFPeA[tw] OR "Perfluoropentanoic acid"[nm] OR Perfluoropentanoic[tw] OR Perfluorovaleric[tw] OR Nonafluoropentanoic[tw] OR Nonafluorovaleric[tw] OR "n-Perfluoropentanoic"[tw] OR "Perfluoro-n-pentanoic"[tw] OR "Nonafluoro-valeric"[tw] OR (c5 [tw] AND perfluorinated [tw])	160
#3	307-24-4[rn] OR PFHxA[tw] OR "Perfluorohexanoic acid"[nm] OR Perfluorohexanoic[tw] OR "Perfluoro-hexanoic"[tw] OR "Perfluoro hexanoate"[tw] OR "IPC-PFFA-6"[tw] OR "undecafluoro-hexanoic"[tw] OR "Undecafluoro-1-hexanoic"[tw] OR Undecafluorohexanoic[tw] OR (c6 [tw] AND perfluorinated [tw])	313
#4	375-85-9[rn] OR 20109-59-5[rn] OR 6130-43-4[rn] OR PFHpA[tw] OR "Perfluoroheptanoic acid"[nm] OR Perfluoroheptanoic[tw] OR Perfluoroheptanoicacid[tw] OR Perfluoroheptanoate[tw] OR Tridecafluoroheptanoic[tw] OR "Perfluoro-n-heptanoic"[tw] OR Perfluoroenanthic[tw] OR "tridecafluoro-heptanoic"[tw] OR "Tridecafluoro-1-heptanoic"[tw] OR "n-perfluoroheptanoic"[tw] OR Tridecafluoroenanthic[tw] OR (c7 [tw] AND perfluorinated [tw])	245
#5	375-95-1[rn] OR 4149-60-4[rn] OR PFNA[tw] OR "perfluorononanoic acid"[nm] OR perfluorononanoic[tw] OR "Perfluoro-n-nonanoic"[tw] OR Perfluornonansaeure[tw] OR "Perfluorononan-1-oic"[tw] OR Perfluoropelargonic[tw] OR Heptadecafluorononanoic[tw] OR Heptadecafluornonansaeure[tw] OR "heptadecafluoro-nonanoic"[tw] OR Heptadecafluoropelargonic[tw] OR "n-Heptadecafluorononanoic"[tw] OR "heptadecafluoro-n-nonanoic"[tw] OR (c9 [tw] AND perfluorinated [tw])	888
#6	335-76-2[rn] OR PFDA[tw] OR "Perfluorodecanoic Acid"[nm] OR Perfluorodecanoic[tw] OR Nonadecafluorodecanoic[tw] OR Ndfda[tw] OR "Perfluoro-N-decanoic"[tw] OR perfluorocaprylic[tw] OR "Nonadecafluoro-n-decanoic"[tw] OR Perfluorocapric[tw] OR "n-perfluorodecanoic"[tw] OR (c10 [tw] AND perfluorinated [tw])	543
#7	2058-94-8[rn] OR PFUnA[tw] OR "Perfluoroundecanoic Acid"[nm] OR Perfluoroundecanoic[tw] OR heneicosafluoroundecanoic[tw] OR "Perfluoro-n-undecanoic"[tw] OR "heneicosafluoro-undecanoic"[tw] OR "C11-PFA"[tw] OR (c11 [tw] AND perfluorinated [tw])	222
#8	307-55-1[rn] OR PFDoA[tw] OR "Perfluorododecanoic Acid"[nm] OR Perfluorododecanoic[tw] OR Perfluorolauric[tw] OR Tricosafluorododecanoic[tw] OR Tricosafluorolauric[tw] OR "tricosafluoro-	170

	Dodecanoic"[tw] OR "n-perfluorododecanoic"[tw] OR (c12 [tw] AND perfluorinated [tw])	
#9	72629-94-8[rn] OR PFTrDA[tw] OR "perfluorotridecanoic acid"[nm] OR perfluorotridecanoic[tw] OR Pentacosafluorotridecanoic[tw] OR "Pentacosafluoro-tridecanoic"[tw] OR (c13[tw] AND perfluorinated [tw])	65
#10	376-06-7[rn] OR PFTeA[tw] OR PFTA[tw] OR "perfluorotetradecanoic acid"[nm] OR "perfluoromyristic acid"[nm] OR perfluorotetradecanoic[tw] OR Heptacosafluorotetradecanoic[tw] OR perfluoromyristic[tw] OR "heptacosafluoro-tetradecanoic acid"[tw] OR (c14 [tw] AND perfluorinated [tw])	92
#11	375-73-5[rn] OR 59933-66-3[rn] OR 29420-49-3[rn] OR 68259-10-9[rn] OR 45187-15-3[rn] OR PFBS[tw] OR PFBuS[tw] OR "Eftop FBSA"[tw] OR "nonafluorobutane-1-sulfonic acid"[nm] OR "Perfluorobutanesulfonic acid"[nm] OR "1-Butanesulfonic acid, 1,1,2,2,3,3,4,4,4-nonafluoro-"[tw] OR "1-Butanesulfonic acid, nonafluoro-"[tw] OR "1-Perfluorobutanesulfonic acid"[tw] OR "1-Perfluorobutanesulfonic"[tw] OR "1,1,2,2,3,3,4,4,4-Nonafluoro-1-butanesulfonic acid"[tw] OR "1,1,2,2,3,3,4,4,4-Nonafluorobutane-1-sulphonic acid"[tw] OR "Nonafluoro-1-butanesulfonic acid"[tw] OR "nonafluoro-1-butanesulfonic"[tw] OR "nonafluoro-butanesulfonic acid"[tw] OR "nonafluorobutane sulfonic"[tw] OR "nonafluorobutane-1-sulfonic"[tw] OR "Nonafluorobutanesulfonic acid"[tw] OR "nonafluorobutane-1-sulphonic"[tw] OR "pentyl perfluorobutanoate"[tw] OR "Perfluoro-1-butanesulfonate"[tw] OR "perfluoro-1-butanesulfonic"[tw] OR "perfluorobutane sulfonate"[tw] OR "Perfluorobutane sulfonic acid"[tw] OR "perfluorobutane sulfonic"[tw] OR "Perfluorobutanesulfonate"[tw] OR "Perfluorobutanesulfonic acid"[tw] OR "perfluorobutyl sulfonic"[tw] OR "Perfluorobutylsulfonate"[tw] OR "perfluorobutane-1-sulfonic"[tw] OR FC-98[tw] OR Nonaflate[tw] OR nonafluorobutanesulfonic[tw] OR nonafluorobutanesulphonic[tw] OR perfluorobutanesulphonic[tw]	328
#12	2706-91-4[rn] OR PFPeS[tw] OR "perfluoropentanesulfonic acid"[nm] OR perfluoropentanesulfonic[tw] OR "perfluoropentane-1-sulphonic"[tw] OR "Perfluoropentane-1-sulfonic"[tw] OR "1-Pentanesulfonic"[tw] OR "perfluoropentane sulfonic"[tw] OR "Undecafluoro-1-pentanesulfonic"[tw] OR "undecafluoropentane-1-sulfonic"[tw]	54
#13	355-46-4[rn] OR 3871-99-6[rn] OR 68259-08-5[rn] OR pfhxs[tw] OR "Perfluorohexanesulfonic Acid"[nm] OR "Perfluorohexane sulfonic"[tw] OR "Perfluorohexane-1-sulphonic"[tw] OR "tridecafluoro-1-Hexanesulfonic"[tw] OR "Tridecafluorohexane-1-sulfonic"[tw] OR Perfluorohexanesulfonic[tw]	520
#14	375-92-8[rn] OR 60270-55-5[rn] OR PFHpS[tw] OR "Perfluoroheptanesulfonic acid"[nm] OR Perfluoroheptanesulfonic[tw] OR "Perfluoroheptane sulfonic"[tw] OR "Pentadecafluoro-1-heptanesulfonic"[tw] OR "pentadecafluoroheptane-1-sulfonic"[tw]	26
#15	68259-12-1[rn] OR 17202-41-4[rn] OR PFNS[tw] OR "Perfluorononanesulfonic acid"[nm] OR Perfluorononanesulfonic[tw] OR "Nonadecafluoro-1-nonanesulfonic"[tw] OR "nonadecafluoro-1-Nonanesulfonic"[tw]	38
#16	335-77-3[rn] OR 67906-42-7[rn] OR PFDS[tw] OR "Perfluorodecane sulfonic acid"[nm] OR "Perfluorodecane sulfonic"[tw] OR "Perfluorodecane sulphonic"[tw] OR henicosafluorodecanesulphonicacid[tw] OR perfluordecansulfonsaure[tw] OR "Perfluorodecane sulfonate"[tw] OR "henicosafluorodecane-1-sulfonic"[tw] OR "heneicosafluoro-1-decanesulfonic"[tw] OR henicosafluorodecanesulphonic[tw] OR "henicosafluorodecane sulphonic"[tw] OR "henicosafluorodecane sulfonic"[tw]	230
#17	919005-14-4[rn] OR 958445-44-8[rn] OR ADONA[tw] OR "3H-perfluoro-3-[(3-methoxy-propoxy)propanoic]"[tw] OR "ammonium 4,8-dioxa-3H-perfluorononanoate"[tw] OR "4,8-dioxa-3H-perfluorononanoic"[tw] OR "2,2,3-trifluoro-3-[1,1,2,2,3,3-hexafluoro-3-	18

	(trifluoromethoxy)propoxy]"[tw] OR "3H-perfluoro-3-[(3-methoxy-propoxy)propanoate]"[tw]	
#18	13252-13-6[rn] OR 62037-80-3[rn] OR 236-236-8[rn] OR 26099-32-1[rn] OR GenX[tw] OR "2-(Heptafluoropropoxy)-2,3,3,3-tetrafluoro-propionic"[tw] OR "2-(Heptafluoropropoxy)-2,3,3,3-tetrafluoropropanoic"[tw] OR "2-(Heptafluoropropoxy)-2,3,3,3-tetrafluoropropionic"[tw] OR "2-(Heptafluoropropoxy)tetrafluoropropionic acid"[nm] OR "2-(Heptafluoropropoxy)tetrafluoropropionic"[tw] OR "2-(Heptafluoropropoxy)tetrafluoropropionicacid"[tw] OR "2,3,3,3-tetrafluoro-2-(heptafluoropropoxy)propanoic acid"[nm] OR "2,3,3,3- tetrafluoro-2-(heptafluoropropoxy)propanoic acid"[tw] OR "2,3,3,3-tetrafluoro-2- (1,1,2,2,3,3,3-heptafluoropropoxy)-Propanoic acid"[tw] OR "2,3,3,3-tetrafluoro-2- (1,1,2,2,3,3,3-heptafluoropropoxy)propanoic acid"[tw] OR "2,3,3,3-Tetrafluoro-2- (heptafluoropropoxy)propionic acid"[tw] OR "2,3,3,3-tetrafluoro-2- (perfluoro propoxy) propanoic"[tw] OR "2,3,3,3-tetrafluoro-2-(perfluoro propoxy)propanoic"[tw] OR "2,3,3,3-Tetrafluoro-2-(1,1,2,2,3,3,3-heptafluoropropoxy)propanoic acid, ammonium salt"[tw] OR "2,3,3,3-tetrafluoro-2-(1,1,2,2,3,3,3-heptafluoropropoxy) propanoic"[tw] OR "2,3,3,3-tetrafluoro-2-(1,1,2,2,3,3,3-heptafluoropropoxy)propanoic"[tw] OR "2,3,3,3-tetrafluoro-2-(1,1,2,2,3,3,3-heptafluoropropoxy)propionic"[tw] OR "2,3,3,3-tetrafluoro-2-(heptafluoropropoxy) propanoic"[tw] OR "2,3,3,3-tetrafluoro-2-(heptafluoropropoxy)propanoate"[tw] OR "2,3,3,3-tetrafluoro-2-(heptafluoropropoxy)propanoic"[tw] OR "2,3,3,3-Tetrafluoro-2-(heptafluoropropoxy)propionic"[tw] OR "2,3,3,3-tetrafluoro-2-(perfluoropropoxy) propanoic"[tw] OR "2,3,3,3-tetrafluoro-2-(perfluoropropoxy)propanoic"[tw] OR "Ammonium 2-(perfluoropropoxy)perfluoropropionate"[tw] OR "Ammonium 2,3,3,3- tetrafluoro-2-(heptafluoropropoxy)propanoate"[tw] OR "ammonium perfluoro(2-methyl-3-oxahexanoate)"[tw] OR "Ammonium perfluoro(2-methyl-3-oxahexanoic) acid"[tw] OR "FRD902"[tw] OR "GenX-H3N"[tw] OR "HFPO-DA"[tw] OR "hexafluoropropylene oxide dimer"[tw] OR "Perfluorinated aliphatic carboxylic acid, ammonium salt"[tw] OR "Perfluorinated aliphatic carboxylic acid"[tw] OR "perfluoro-2-methyl-3- oxahexanoic acid"[tw] OR "perfluoro-2-propoxypropanoic acid"[tw] OR "perfluoro-2-propoxypropionic acid"[tw] OR "perfluoro-2-propoxypropionic"[tw] OR "perfluoro-αpropoxypropionic acid"[tw] OR "Perfluoro(2- methyl-3-oxahexanoic) acid"[tw] OR "Perfluoro(2-methyl-3-oxahexanoate) "[tw] OR "Perfluoro(2-methyl-3-oxahexanoic)"[tw] OR "perfluoro2-(propyloxy)propionic acid"[tw] OR "propanoic acid, 2,3,3,3-tetrafluoro-2- (heptafluoropropoxy)-"[tw] OR "Propanoic acid, 2,3,3,3-tetrafluoro-2-(1,1,2,2,3,3,3-heptafluoropropoxy)- "[tw] OR "Propanoic acid, 2,3,3,3-tetrafluoro-2-(1,1,2,2,3,3,3-heptafluoropropoxy)-, ammonium salt"[tw] OR "Propanoic acid, 2,3,3,3-tetrafluoro-2-(heptafluoropropoxy)-, ammonium salt"[tw] OR "propionic acid, 2,3,3,3-tetrafluoro-2-(heptafluoropropoxy)-"[tw] OR "tetrafluoro-2-(1,1,2,2,3,3,3-heptafluoropropoxy)propanoic"[tw] OR "tetrafluoro-2-(heptafluoropropoxy)propanoate"[tw] OR "tetrafluoro-2-(heptafluoropropoxy)propanoic"[tw] OR "Undecafluoro-2-methyl-3-oxahexanoic acid"[tw] OR (("2,3,3,3-Tetrafluoro-2- (heptafluoropropoxy)propionic"[tw] OR "2,3,3,3-tetrafluoro-2-(1,1,2,2,3,3,3-heptafluoropropoxy)-Propanoic"[tw] OR "Perfluorinated aliphatic carboxylic"[tw] OR "Perfluoro(2-methyl-3- oxahexanoic)"[tw] OR "2,3,3,3-tetrafluoro-2-(1,1,2,2,3,3,3-heptafluoropropoxy)propanoic"[tw] OR "2,3,3,3-tetrafluoro-2- (heptafluoropropoxy)propanoic"[tw] OR "perfluoro-2- (propyloxy)propionic"[tw] OR "perfluoro-2-methyl-3- oxahexanoic"[tw] OR "perfluoro-2-propoxypropanoic"[tw] OR "perfluoro-2-propoxypropionic"[tw] OR "perfluoro-αpropoxypropionic"[tw]) AND (acid[tw] OR acids[tw])) OR (("Undecafluoro-2- methyl-3-oxahexanoic"[tw] OR "Ammonium perfluoro(2-methyl-3- oxahexanoic)"[tw] OR "2,3,3,3-Tetrafluoro-2-(1,1,2,2,3,3,3- heptafluoropropoxy)"[tw] OR "Perfluorinated aliphatic carboxylic"[tw]) AND (salt[tw] OR salts[tw] OR acid[tw] OR acids[tw]))))) OR (((Undecafluoro AND oxahexanoic) OR (Ammonium AND perfluoro AND oxahexanoic) OR (Tetrafluoro AND heptafluoropropoxy) OR "Perfluorinated aliphatic carboxylic"[tw] OR "Perfluorinated aliphatic carboxylic"[tw]) AND (salt[tw] OR salts[tw] OR acid[tw] OR acids[tw])) OR (GenX AND (fluorocarbon*[tw] OR fluorotelomer*[tw] OR polyfluoro*[tw] OR perfluoro-*[tw] OR perfluoroa*[tw] OR perfluorob*[tw] OR perfluoroc*[tw] OR perfluorod*[tw] OR perfluoroe*[tw] OR perfluoroh*[tw] OR perfluoron*[tw] OR perfluoroo*[tw] OR perfluorop*[tw] OR perfluoros*[tw] OR perfluorou*[tw] OR perfluorinated[tw] OR	42

	fluorinated[tw]))	
#19	13252-14-7[rn] OR "HFPO-TA"[tw] OR HFPO[tw] OR (Hexafluoropropylene[tw] AND ("oxides"[MeSH Terms] OR oxide*[tw])) OR ("hexafluoropropene"[tw] AND ("oxides"[MeSH Terms] OR oxide*[tw])) OR "hexafluoropropylene oxide"[tw] OR "HFPO trimer"[tw]	62
#20	73606-19-6[rn] OR "6:2 ClPFESA"[tw] OR "6:2 Cl PFESA"[tw] OR "6:2 chlorinated polyfluorinated ether sulfonic acid"[nm] OR "chlorinated polyfluorinated ether sulfonic"[tw] OR "2-[(6-Chloro-1,1,2,2,3,3,4,4,5,5,6,6-dodecafluorohexyl)oxy]-1,1,2,2-tetrafluoroethanesulfonic"[tw] OR "2-(6-Chlorododecafluorohexyloxy)-1,1,2,2-tetrafluoroethanesulfonic"[tw] OR "6:2 Cl-PFESA" OR "Cl-PFESA" OR (Cl[tw] AND (PFESA[tw] OR PFESAs[tw])) OR (((chlorinated[tw] AND polyfluorinated[tw] AND ("sulfonic acids"[MeSH Terms] OR ("sulfonic"[tw] AND "acids"[tw]) OR "sulfonic acids"[tw] OR ("sulfonic"[tw] AND "acid"[tw]) OR "sulfonic acid"[tw])))) OR "2-(6-chloro-1,1,2,2,3,3,4,4,5,5,6,6-dodecafluorohexoxy)-1,1,2,2-tetrafluoroethanesulfonic"[tw]	33
#21	83329-89-9[rn] OR "8:2 Cl:PFESA"[tw] OR "8:2 Cl PFESA"[tw] OR "8:2 Cl-PFESA"[tw] OR "Cl:PFESA" OR "Cl PFESA" OR "8:2 chlorinated polyfluorinated ether sulfonic acid"[nm] OR "8:2 chlorinated polyfluorinated ether sulfonic acid"[tw] OR "2-[oxyl]-1,1,2,2-tetrafluoro-ethanesulfonicacid"[tw] OR "2-[oxyl]-1,1,2,2-tetrafluoro-ethanesulfonic"[tw] OR (chlorinated[tw] AND polyfluorinated[tw] AND ether[tw] AND (sulphonic acid*[tw] OR sulfonic acid*[tw])) OR "2-(8-chloro-1,1,2,2,3,3,4,4,5,5,6,6,7,7,8,8-hexadecafluorooctoxy)-1,1,2,2-tetrafluoroethanesulfonate"[tw] OR "2-(8-chloro-1,1,2,2,3,3,4,4,5,5,6,6,7,7,8,8-hexadecafluorooctoxy)-1,1,2,2-tetrafluoroethanesulfonic"[tw]	28
#22	27619-97-2[rn] OR 59587-39-2[rn] OR "6:2 FTSA"[tw] OR "6:2 FTSA" OR "6:2 FtS"[tw] OR ("6:2"[tw] AND FTSA[tw]) OR ("6:2"[tw] AND FtS[tw]) OR "6:2 fluorotelomer sulfonic"[tw] OR "6:2 fluorotelomer sulphonic"[tw] OR "fluorotelomer sulfonic"[tw] OR "fluorotelomer sulphonic"[tw] OR "3,3,4,4,5,5,6,6,7,7,8,8,8-tridecafluorooctane-1-sulfonic"[tw]	48
#23	39108-34-4[rn] OR 254-295-8[rn] OR "8:2 FTSA" OR "8:2 fluorotelomer sulfonic"[tw] OR "8:2 fluorotelomer sulphonic"[tw] OR "2-(Perfluorooctyl)ethane-1-sulphonic"[tw] OR "2-(Perfluorooctyl)ethane-1-sulfonic"[tw] OR ("8:2"[tw] AND fluorotelomer[tw]) OR ("8:2"[tw] AND FTSA[tw]) OR ("8:2"[tw] AND FtS[tw]) OR "Heptadecafluorodecanesulphonic"[tw] OR "heptadecafluorodecane-1-sulfonic"[tw] OR "Perfluorodecanesulfonic"[tw] OR "Heptadecafluorodecane-1-sulphonic"[tw] OR "heptadecafluoro-1-Decanesulfonic"[tw] OR "3,3,4,4,5,5,6,6,7,7,8,8,9,9,10,10,10-Heptadecafluorodecanesulphonic"[tw] OR "3,3,4,4,5,5,6,6,7,7,8,8,9,9,10,10,10-heptadecafluorodecane-1-sulfonic"[tw] OR "1H,1H,2H,2H-Perfluorodecanesulphonic acid"[tw] OR "1-Decanesulfonic acid,3,3,4,4,5,5,6,6,7,7,8,8,9,9,10,10,10-heptadecafluoro-"[tw]	1917
#24	2991-50-6[rn] OR NEtFOSAA[tw] OR "2-(N-ethyl-perfluorooctane sulfanamido) acetic acid"[tw] OR "2-(N-ethyl-perfluorooctane sulfonamido) acetic acid"[tw] OR "2-(N-ethyl-perfluorooctanesulfonamido) acetic acid"[tw] OR "2-(N-Ethylperfluorooctanesulfonamido)acetic acid"[tw] OR "2-[ethyl(1,1,2,2,3,3,4,4,5,5,6,6,7,7,8,8,8-heptadecafluorooctylsulfonyl)amino]acetic acid"[tw] OR "N-(ethyl)-N-(perfluorooctylsulfonyl)-aminoacetic acid"[tw] OR "n-(ethyl)n-(perfluorooctylsulfonyl)-aminoacetic acid"[tw] OR "N-(Heptadecafluorooctylsulfonyl)-N-ethylglycine"[tw] OR "N-ethyl perfluorooctanesulfonamidoacetic acid"[tw] OR "N-ethyl-N-((1,1,2,2,3,3,4,4,5,5,6,6,7,7,8,8,8-heptadecafluorooctyl)sulfonyl)-"[tw] OR "N-ethyl-N-((heptadecafluorooctyl)sulfonyl)-"[tw] OR "N-Ethyl-N-((heptadecafluorooctyl)sulphonyl)glycine"[tw] OR "N-ethyl-N-[(heptadecafluorooctyl)sulfonyl]-"[tw] OR "N-ethylperfluorooctane sulfonamidoacetic acid"[tw] OR "N-ethylperfluorooctane sulfonamidoacetic"[tw]	2
#25	2355-31-9[rn] OR NMeFOSAA[tw] OR "2-(N-Methyl-perfluorooctane sulfanamido) acetic acid"[tw] OR "2-(N-methyl-perfluorooctane sulfonamido) acetic acid"[tw] OR "2-(N-Methylperfluorooctanesulfonamido)acetic acid"[tw] OR "N-(Heptadecafluorooctylsulfonyl)-N-	4

	methylglycine"[tw] OR "N-[(heptadecafluorooctyl)sulfonyl]-N-methyl-"[tw] OR "N-[(heptadecafluorooctyl)sulfonyl]-N-methylglycine"[tw] OR "N-methyl perfluorooctanesulfonamidoacetic acid"[tw] OR "N-methylperfluorooctane sulfonamidoacetic acid"[tw]	
#26	749836-20-2[rn] OR Nafion[tw] OR "1,1,2,2-Tetrafluoro-2-{[1,1,1,2,3,3-hexafluoro-3-(1,2,2,2-tetrafluoroethoxy)propan-2-yl]oxy}ethane-1-sulfonic"[tw] OR "Perfluoro-2-{[perfluoro-3-(perfluoroethoxy)-2-propanyl]oxy}ethanesulfonic"[tw]	2,180
#27	39492-90-5[rn] OR PFO4DA[tw] OR "1,1,1,3,3,5,5,7,7,9,9-Undecafluoro-2,4,6,8-tetraoxadecan-10-oic acid"[tw] OR "3,5,7,9-Tetraoxadecanoicacid, 2,2,4,4,6,6,8,8,10,10,10-undecafluoro-"[tw] OR "Perfluoro-3,5,7,9-butaoxadecanoic"[tw] OR "Perfluoro-3,5,7,9-tetraoxadecanoic"[tw]	1
#28	39492-91-6[rn] OR PFO5DoDA[tw] OR "1,1,1,3,3,5,5,7,7,9,9,11,11-Tridecafluoro-2,4,6,8,10-pentaoxadodecan-12-oic"[tw] OR "3,5,7,9,11-Pentaoxadodecanoicacid, 2,2,4,4,6,6,8,8,10,10,12,12,12-tridecafluoro-"[tw] OR "Perfluoro-3,5,7,9,11-pentaoxadodecanoic"[tw]	0
#29	773804-62-9[rn] OR "Hydro-Eve"[tw] OR "Hydro Eve"[tw] OR "HydroEve"[tw] OR "2,2,3,3-tetrafluoro-3-((1,1,1,2,3,3-hexafluoro-3-(1,2,2,2-tetrafluoroethoxy)propan-2-yl)oxy)propanoic"[tw]	0
#30	PFAS*[tiab] OR PFCs[tiab] OR PFAA*[tiab] OR perfluorochemical*[tiab] OR perfluorinated[tiab] OR ("per-"[tiab] AND polyfluoroalkyl[tiab]) OR "perfluorinated alkyl"[tiab] OR "Perfluorinated carboxylic"[tiab] OR "perfluorinated chemicals"[tiab] OR "perfluoroalkyl acid"[tiab] OR "perfluoroalkyl acids"[tiab] OR ("perfluoroalkyl sulfonic"[tiab] AND (acid*[tiab] OR acid[tiab] OR acids[tiab])) OR ("perfluoroalkyl sulphonic"[tiab] AND (acid*[tiab] OR acid[tiab] OR acids[tiab])) OR "perfluoroalkyl sulphonic"[tiab] OR ("poly-"[tiab] AND perfluoroalkyl[tiab]) OR "polyfluorinated alkyl"[tiab] OR "polyfluorinated chemicals"[tiab] OR ("polyfluorinated"[tiab] AND substance*[tiab]) OR “fluorinated polymer”[tiab] OR “fluorinated polymers”[tiab] OR (fluorinated[tiab] AND (polymer[tiab] OR polymers[tiab])) OR (fluorinated[tiab] AND surfactant*[tiab]) OR (fluorinated[tiab] AND telomer*[tiab]) OR fluoro-telomer*[tiab] OR (fluorocarbon[tiab] AND (polymer[tiab] OR polymers[tiab])) OR Fluoropolymer*[tiab] OR Fluorosurfactant*[tiab] OR fluorotelomer*[tiab] OR (Perfluorinated[tiab] AND carboxylic[tiab]) OR (perfluorinated[tiab] AND substance*[tiab]) OR perfluoroalkyl[tiab] OR (perfluoroalkyl[tiab] AND acid[tiab]) OR (perfluoroalkyl[tiab] AND acids[tiab]) OR (perfluoroalkyl[tiab] AND substance*[tiab]) OR (perfluorocarbon*[tiab] AND (chemical*[tiab] OR compound*[tiab])) OR perfluorocarboxylic[tiab] OR perfluorosulfonic[tiab] OR polyfluorinated[tiab] OR polyfluoroalkyl[tiab] OR (polyfluoroalkyl[tiab] AND acid[tiab]) OR (polyfluoroalkyl[tiab] AND acids[tiab]) OR (polyfluoroalkyl[tiab] AND substance*[tiab]) OR polyflurochemical*[tiab]	8,948
#31	#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12 OR #13 OR #14 OR #15 OR #16 OR #17 OR #18 OR #19 OR #20 OR #21 OR #22 OR #23 OR #24 OR #25 OR #26 OR #27 OR #28 OR #29 OR #30	12,490

PRISMA-P Report (modified) for Systematic Map Protocols Submitted to *Environment International*

Version 1.0 | 15 Mar 2017. This form is to be completed as supplemental information alongside any systematic review protocol submitted to *Environment International*. Authors are asked to provide relevant quotes in addition to page numbers.

Title of submitted paper and corresponding author: PFAS Health Effects Database: Protocol for a Systematic Evidence Map

#	Item	Guidance	On page #	Author Comments
Title				
1	Identification	Identify the report as a systematic map.	1	The title identifies this as a protocol for a systematic evidence map.
2	Update	If the protocol is for an update of a previous systematic map, identify as such.	N/A	
Registration				
3	Registration	If registered, provide the name of the registry and registration number.	6	The protocol, as submitted to the journal, was published to Zenodo.
Authors				
4	Contact	Provide name, institutional affiliation, and e-mail address of all protocol authors; provide physical mailing address of corresponding author.	1	
5	Contributions	Describe contributions of protocol authors and identify the guarantor of the review.	8	Listed under "Author Contributions"
Amendments				
6	Amendments	If the protocol represents an amendment of a previously completed or published protocol, identify as such and list changes; otherwise, state plan for documenting important protocol amendments.	N/A	
Support				
7	Sources	Indicate sources of financial or other support for the map.	7	Listed under "Financial Support"
8	Sponsor	Provide name for the map funder/s and/or sponsor/s.	N/A	
9	Roles	Describe roles of funder(s), sponsor(s), and/or institution(s), if any, in developing the protocol.	8	Listed under "Financial Support"
Introduction				
10	Rationale	Describe the rationale for the map in the context of what is already known.	1-2	Section 1.1
11	Objectives	Define primary and secondary questions for the systematic map.	6	Section 1.2

PRISMA-P Report (modified) for Systematic Map Protocols Submitted to *Environment International*

Version 1.0 | 15 Mar 2017. This form is to be completed as supplemental information alongside any systematic review protocol submitted to *Environment International*. Authors are asked to provide relevant quotes in addition to page numbers.

#	Item	Guidance	On page #	Author Comments
Methods				
12	Eligibility criteria	Specify the study characteristics and report characteristics to be used as criteria for eligibility for the map.	6-7	Section 2.3
13	Information sources	Describe all intended information sources (e.g., electronic databases, contact with study authors, trial registers, or other grey literature sources) with planned dates of coverage.	6	Section 2.1
14	Search strategy	Present draft of search strategy to be used for at least one electronic database, including planned limits, such that it could be repeated.	6	Section 2.2. The specific search strategy is in Appendix A in Supplemental Materials
15	Data management	Describe the mechanism(s) that will be used to manage records and data throughout the review.	7	Section 2.4
16	Selection process	State the process that will be used for selecting studies (e.g., two independent reviewers) through each phase of the review (i.e., screening, eligibility, and inclusion in meta-analysis).	7	Section 2.5
17	Data collection process	Describe planned method of extracting data from reports (e.g., piloting forms, done independently, in duplicate), any processes for obtaining and confirming data from investigators.	7	Section 2.5
18	Data coding strategy	List and define all variables for which data were sought and any pre-planned assumptions and simplifications made.	7, Table 3	Section 2.6
19	Study quality assessment	If it is to be conducted, describe methods for assessing quality of individual studies.	7	Section 2.8

Environment International modified PRISMA-P report adapted from: Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart LA. Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) 2015 statement. Syst Rev. 2015;4(1):1. and *Environmental Evidence* "Preparing your manuscript: Systematic Map" http://environmentalevidencejournal.biomedcentral.com/submission-guidelines/preparing-your-manuscript/systematic-map (retrieved 24 February 2017)