Environment International xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Environment International

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

Weight of evidence evaluation and systematic review in EU chemical risk assessment: Foundation is laid but guidance is needed

Marlene Ågerstrand *, Anna Beronius

Department of Environmental Science and Analytical Chemistry (ACES), Stockholm University, Stockholm, Sweden

ARTICLE INFO

Article history:
Received 23 July 2015
Received in revised form 25 September 2015
Accepted 14 October 2015
Available online xxxx

Keywords:
Weight of evidence
Systematic review
Evidence integration
Chemical risk assessment
Chemical legislation

ABSTRACT

The aim of this review was to investigate if and how the application of weight of evidence (WoE) evaluation or systematic review (SR) in chemical risk assessment is promoted within different regulatory frameworks in the European Union. Legislative and relevant guidance documents within nine regulatory frameworks were scrutinized and compared. WoE evaluation or SR is promoted in seven of the investigated frameworks but sufficient guidance for how to perform these processes is generally lacking. None of the investigated frameworks give enough guidance for generating robust and reproducible WoE evaluations or SRs. In conclusion, the foundation for use of WoE evaluation and SR is laid in the majority of the investigated frameworks, but there is a need to provide more structured and detailed guidance. In order to make the process of developing guidance as efficient as possible, and to ensure smooth transfer of risk assessment's between frameworks if a chemical is risk assessed both as, for example, a biocide and an industrial chemical, it is recommended that guidance is developed jointly by the European regulatory agencies.

1. Introduction

Environmental and health risk assessment is a key step in the regulation of chemicals in the European Union (EU), i.e. for approving or restricting the use of chemicals. Risk assessment is a step-wise procedure that involves evaluation and interpretation of scientific data, as well as policy-influenced practices such as use of default assumptions, for example regarding assessment factors and thresholds for effect, and case by case judgments, for example concerning relevance of data. Within the EU chemicals are risk assessed within different regulatory frameworks depending on their intended use, e.g. as cosmetics, plant protection products or pharmaceuticals. This means that the risk assessment process, including policy-influenced practices, may vary for different compounds even if the nature of their toxicity is similar and similar risks to human health and the environment can be expected.

Risk assessment can be considered to consist of three main parts: hazard assessment (including hazard identification and hazard characterization), exposure assessment and risk characterization. In this review we focus on hazard assessment. Traditionally, hazard assessment entails identifying one or a few key toxicity studies, upon which the identification and characterization of the critical (most relevant and sensitive) adverse effects of the compound will be based. In risk assessment conducted for regulatory purposes the key study is often an in vivo (eco)toxicity study conducted according to standardized and internationally validated test guidelines, such as the OECD test guidelines, and Good Laboratory Practices (GLP) (European Chemicals Agency, 2008). Standardized test guidelines and GLP have been put in place to promote high reliability of (eco)toxicity test results and are therefore often preferred by agencies conducting risk assessment for regulatory decision making. In practice then, the regulation of a chemical will potentially be based on the results and conclusions from a single study.

Different approaches for assessment of *whole* data sets, often referred to as weight of evidence (WoE) evaluation or systematic review (SR), have been promoted (Koustas et al., 2014; Rooney et al., 2014; European Food Safety Authority, 2010; Whaley et al., 2015; IARC, 2006). In general terms, WoE evaluation and SR are processes of summarizing, synthesizing and interpreting a body of evidence to draw conclusions, e.g. regarding the relationship between a chemical exposure and adverse health effect. As such, these processes differ from the traditional method for risk assessment by promoting the use and integration of information from all available evidence instead of focusing on a single key study. WoE evaluation has established use in several different disciplines, such as economics and law (Krimsky, 2005). SR has also been used for over 30 years in the field of medicine, for example in the Cochrane collaboration (Higgins and Green, 2009).

In the environmental health field, as well as in EU chemicals regulation, the terms WoE evaluation and SR are sometimes used interchangeably and sometimes with slightly different meanings. Historically, the WoE-concept has been used in many different ways, often without providing a clear definition (Weed, 2005; Linkov et al., 2009; Krimsky, 2005). WoE evaluation has for example been used to describe the

* Corresponding author.
E-mail address: marlene.agerstrand@aces.su.se (M. Ågerstrand).

http://dx.doi.org/10.1016/j.envint.2015.10.008
0160-4120/

whole assessment process, from assembling available studies to evaluating, interpreting and integrating the whole body of evidence to reach conclusion, while others use WoE evaluation when describing the process that occurs after assembling studies (Rhomberg et al., 2013). Recently, the US National Research Council decided to replace the term WoE with "evidence integration" due to the vague and varied use, and since it is sometimes used in a manner that oversimplifies the actual situation (National Research Council, 2014). In turn, the key characteristics of a SR are according to e.g. the Cochrane collaboration: a clearly stated objective with pre-defined eligibility criteria for studies; an explicit, reproducible methodology; a systematic search that attempts to identify all studies that would meet the eligibility criteria; an assessment of the validity of the findings of the included studies; and a systematic presentation, and synthesis of the characteristics and findings of the included studies (Higgins and Green, 2009). In this review we use the terms WoE or SR as they are used in the respective EU legislations and guidance documents for risk assessments of chemicals included in this investigation. The main point is that both concepts provide an alternative to the traditional praxis of identifying a key study and instead promote the use of entire bodies of evidence to reach conclusions regarding health and environmental hazards and risks.

2. Review of frameworks

The aim of this review was to investigate if the application of either WoE evaluation or SR in chemical risk assessment is specifically promoted within different regulatory frameworks in the EU and, in that case, when and how such a process should be applied according to current policy. To this end, legislative documents regulating the risk assessment of chemicals, as well as current guidance documents relevant for risk assessment, within nine EU regulatory frameworks were scrutinized (Table 1). These nine frameworks were chosen since they represent the most prominent areas within chemicals risk assessment in the EU. The following search terms were used to systematically extract information from each document: "weight of evidence", "weight-of-evidence", "woe", "systematic review", "evidence" and "evidence integration". In addition, the tables of contents for each document were read carefully and sections where relevant information could be found were scrutinized to minimize the risk that the search using specific search terms missed relevant and critical information.

In order for the risk assessment procedure to be consistent across substances and provide sufficient protection for human health and the environment adequate guidance for conducting the different steps of this procedure has to be available to risk assessors. Another goal of this review was therefore to investigate if sufficiently detailed guidance for conducting WoE evaluation and SR is available within the different frameworks. For this, the identified guidance documents were scrutinized and a comparison to the overall steps in SR as described by the Cochrane Collaboration (Higgins and Green, 2009) was performed. This was made in order to investigate differences between the selected regulatory frameworks, but also as a comparison to a different research field (i.e. medicine) where SR has been used for a longer period of time. The guidelines from the Cochrane Collaboration were chosen for these comparisons since they have relatively established use in the clinical field and have provided a basis for recently developed approaches to WoE evaluation and SR for the purpose of chemicals risk assessment (Rooney et al., 2014; Koustas et al., 2014).

3. How is WoE evaluation and SR promoted and defined?

WoE evaluation is mentioned in four of the nine investigated legislations: the REACH regulation, the Biocides directive, the Cosmetics regulation, and the regulation for Classification, Labelling and Packaging (CLP) (Table 2). WoE evaluation or SR is also mentioned in guidance documents for conducting risk assessment following these four legislations,

Table 1
Overview of nine regulatory frameworks included in this review summarizing relevant legislative and guidance documents and the responsible authorities. Guidance documents in italic are documents that are specific for WoE evaluation or SR.

Regulatory framework	Legislation relevant to risk/safety assessment	Guidance document relevant to risk/safety assessment	Responsible EU authority	Body conducting assessment
Industrial chemicals	Regulation (EC) No. 1907/2006 (REACH)	Guidance on information requirements and chemical safety assessment (European Chemicals Agency, 2008; European Chemicals Agency, 2011) *Practical guide 2: How to report weight of evidence* (European Chemicals Agency, 2010)	ECHA	Producing or importing industry
Plant protection products	Regulation (EC) No. 1107/2009	Guidelines on Active Substances and Plant Protection Products (European Food Safety Authority, 2009; EFSA, 2012; European Commission, 2002; EC, 2002, 2004, 2006, 2009) Submission of scientific peer-reviewed open literature for the approval of pesticide active substances under Regulation (EC) No 1107/2009 (European Food Safety Authority, 2011)	EFSA	MS competent authority ("rapporteur")
Biocides	Regulation (EU) No. 528/2012	Guidance on the Biocidal Products Regulation (European Chemicals Agency, 2014; European Chemicals Agency, 2015a; European Chemicals Agency, 2015b)[a]	ECHA	MS competent authority
Cosmetics	Regulation (EC) No. 1223/2009	SCCS/1501/12 The SCCS's notes of guidance for the testing of cosmetic substances and their evaluation, 8th revision (Scientific Committee on Consumer Safety, 2012)	EC	Producing or importing industry
Human pharmaceuticals in the environment	Directive 2001/83/EC	CPMP/SWP/4447/00 Guideline on the environmental risk assessment of medicinal products for human use (EMA, 2006) Questions and answers on the guideline on the environmental risk assessment (EMA, 2015)[b]	EMA	Producing industry
Veterinary pharmaceuticals in the environment	Directive 2001/82/EC	CVMP/VICH/592/1998 Guideline on environmental impact assessment for veterinary medicinal products Phase I (EMA, 2000) CVMP/VICH/790/2003 Guideline on environmental impact assessment for veterinary medicinal products Phase II (EMA, 2005)	EMA	Producing industry
Contaminants in food	Regulation (EC) No. 178/2002	*Application of systematic review methodology to food and feed safety assessments to support decision making* (European Food Safety Authority, 2010)	EFSA	EFSA panel
Water framework directive	Directive 2000/60/EC	Guidance Document No. 27 Technical Guidance For Deriving Environmental Quality Standards (European Commission, 2011)	EC	Member states
Classification, labelling and packaging	Regulation 1272/2008/EC	Guidance on the Application of the CLP Criteria, Version 4.1 (European Chemicals Agency, 2015a; European Chemicals Agency, 2015b)	ECHA	Producing industry

ECHA = European Chemicals Agency; EFSA = European Food Safety Authority; EC = European Commission; EMA = European Medicines Agency.
[a] Some guidance documents are still under development.
[b] The EMA guidance also refers to the guidance for industrial chemicals.

Please cite this article as: Ågerstrand, M., Beronius, A., Weight of evidence evaluation and systematic review in EU chemical risk assessment: Foundation is laid but guidance is needed, Environ Int (2015), http://dx.doi.org/10.1016/j.envint.2015.10.008

Table 2
Use and definition of WoE evaluation and SR in the nine regulatory frameworks reviewed. Footnotes specify in which guidance document/legislation the information can be found.

Regulatory framework	Is WoE/SR mentioned in the legislation?	Is WoE/SR mentioned in guidance documents?	How is WoE/SR defined?	When should WoE/SR be applied?
Industrial chemicals	Yes, the term WoE is used.	Yes, the term WoE is used.[a]	"*The process of considering the strengths and weaknesses of various pieces of information in reaching and supporting a conclusion concerning a property of the substance.*"[a]	When several studies are available with conflicting results; when a key study (e.g. standard study) is missing; when using information that individually may be regarded as insufficient.[a]
Plant protection products	No	Yes, both WoE[b,c,d] and SR[e] are used.	"*This process of combining available lines of evidence to form an integrated conclusion or risk characterization is frequently referred to as weight-of-evidence assessment. This term reflects the principle that the contribution of each line of evidence should be considered in proportion to its weight.*"[d] "*A systematic review is an overview of existing evidence pertinent to a clearly formulated question, which uses pre-specified and standardised methods to identify and critically appraise relevant research, and to extract, report and analyse data from the studies that are included in the review.*"[e]	When there are several sources of qualitative information; when conflicting results are present; when micro−/mesocosm studies, QSAR, read-across, non-standard test data, and toxicodynamic/toxicokinetic are used; when performing higher-tier assessments.[b,c,d,e]
Biocides	Yes, the term WoE is used.	Yes, the term WoE is used.[f]	"*A Weight of Evidence assessment involves the consideration of all data that is available and may be relevant to reproductive toxicity.*"[f]	When combining information of different type; when deciding on testing strategies; for waiving; when using QSAR, in vitro, read-across, human data, toxicokinetic and/or mechanistic data; to maximize use of existing data; when having data of lower quality.[f,g]
Cosmetics	Yes, the term WoE is used.	Yes, the term WoE is used.[h]	Not mentioned	When combining information of different types.[i]
Human pharmaceuticals in the environment	No	No	Not mentioned	Not mentioned
Veterinary pharmaceuticals in the environment	No	No	Not mentioned	Not mentioned
Contaminants in food	No	Yes, the term SR is used.[j]	"*A systematic review is an overview of existing evidence pertinent to a clearly formulated question, which uses pre-specified and standardized methods to identify and critically appraise relevant research, and to collect, report and analyze data from the studies that are included in the review.*"[j]	When there is a large amount of evidence available or when the evidence is scarce; for controversial topics; when evaluating studies that is publicly available; for understanding apparently conflicting results.[j]
Water framework directive	No	Yes, the term WoE is used.[k]	Not mentioned	When using read-across, grouping or QSAR.[k]
Classification, labelling and packaging	Yes, the term WoE is used.	Yes, the term WoE is used.[l]	"*A weight of evidence determination means that all available information bearing on the determination of hazard is considered together.*"[l]	Where no or inadequate test data are available; when assessing mixtures; when using non-standard tests; when both positive and negative data is available; when using in vitro tests, read-across, grouping or QSAR; when using occupational data, data from accident databases, epidemiological, clinical studies and case reports and observations.[l]

[a] European Chemicals Agency (2010).
[b] European Commission (2002).
[c] European Food Safety Authority (2009).
[d] European Food Safety Authority (2013).
[e] European Food Safety Authority (2011).
[f] European Chemicals Agency (2015b).
[g] Regulation (EU) No. 528/2012.
[h] Scientific Committee on Consumer Safety (2012).
[i] Regulation (EC) No 1223/2009.
[j] European Food Safety Authority (2010).
[k] European Commission (2011).
[l] European Chemicals Agency (2015a).

as well as in guidance documents for risk assessment of plant protection products, contaminants in food, and contaminants regulated under the Water Framework Directive (WFD). Notably, most of these guidance documents provide general guidance for the risk assessment process; specific guidance documents for conducting and reporting WoE evaluation and SR are only available for industrial chemicals and contaminants in food, respectively.

The process of WoE evaluation is defined in similar ways in the guidance documents for industrial chemicals, plant protection products, biocides and CLP. The definitions give a general idea of what should be done, i.e. "consider all available data", but give no further guidance on how this process should be carried out. The process of SR is defined in the guidance documents for plant protection products and contaminants in food. These two practically identical definitions for SR are more specific than the definitions for WoE evaluation as they provide statements for how the SR process should be conducted and by specifying critical steps in the process: "*…uses pre-specified and standardized methods to identify and critically appraise relevant research, and to extract, report and analyze data from the studies that are included in the review*". This definition reflects aspects of the Cochrane Collaboration's description of SR used when assessing effects of health care (Higgins and Green, 2009).

Further, WoE evaluation is mentioned in the guidance documents for cosmetics and the WFD but without providing any definition. Omitting to provide a definition of a critical concept is problematic since it opens up for multiple interpretations and misunderstandings, especially if that concept previously has been identified as unclear (Weed, 2005). None of the investigated frameworks discuss any differences

Please cite this article as: Ågerstrand, M., Beronius, A., Weight of evidence evaluation and systematic review in EU chemical risk assessment: Foundation is laid but guidance is needed, Environ Int (2015), http://dx.doi.org/10.1016/j.envint.2015.10.008

between WoE evaluation and SR, and none of them clarify why one concept was chosen instead of the other. Further, in contrast to the US National Research Council report (National Research Council, 2014), there is no discussion in any of the frameworks concerning the problems with vague or missing definitions of WoE evaluation and SR. Altogether, this makes it difficult to know whether the two concepts are considered to be fundamentally different methods and choosing one over the other was a conscious decision, or if it merely reflects a tradition within that particular framework.

4. When should WoE evaluation and SR be applied?

None of the regulatory frameworks reviewed here dictates that WoE evaluation or SR should be performed in *every* hazard/risk assessment. Instead, specific situations where these processes should be applied are exemplified (Table 2). The REACH guidance documents state that WoE evaluation should be used when the standard testing regime does not appear scientifically necessary, i.e. when the assessment can be performed on data already available. Situations where WoE evaluation is recommended under the REACH regulation include when there are several studies with conflicting results, and when a key study cannot be identified. Examples of studies that cannot be considered to be key studies include non-standard studies and studies evaluated to be "not relevant", "not assignable", and in some cases "reliable without restrictions", according to the evaluation method described by Klimisch et al. (1997). The guidance document for CLP states similar application of WoE evaluation.

In the new Cosmetics regulation animal testing is prohibited. However, animal studies can be performed for such substances under the REACH regulation in order to assess risks to the environment, for workers exposed during production, and for non-cosmetic use of the substance (European Chemicals Agency, 2014). Since animal studies are banned for cosmetics focus are put on alternative methods such as in vitro methods, in silico methods and read across (in its most simple form read across means that studies from chemical A are used to predict the same endpoint for chemical B). In the Cosmetics regulation it is stated that an appropriate WoE evaluation should be used in safety assessment when reviewing data from all existing sources. In addition, it is stated that a number of promising alternative test methods are under development and that results from such studies could be included in WoE evaluation in the future.

In the regulation directing risk assessment of contaminants in food under the auspices of the European Food Safety Authority (EFSA) it is stated that "*risk assessment shall be based on the available scientific evidence and undertaken in an independent, objective and transparent manner*". However, WoE evaluation or SR is not specifically mentioned in this regulation. Nonetheless, the EFSA guidance document states that SR should be used when there is an increased complexity in the assessment, such as conflicting results and peer-reviewed studies.

The guidance document for WFD states that WoE evaluation should be used when there are limitations in experimental data and non-testing approaches, such as read-across, grouping and QSAR-methods, are used. The guidance document for biocides also states the already mentioned reasons and in addition stresses that WoE evaluation could be used when deciding on testing strategies and when justifying waiving, i.e. omitting to follow standard test procedures. The guidance documents for plant protection products also suggest that WoE evaluation could be used in higher tiered assessments.

In general it can be said that SR is a newer concept in risk assessment of chemicals in the EU than WoE evaluations. Due to the recent implementations of such approaches as well as the low transparency in many of the investigated frameworks, which is due to low demand on documentation of the risk assessment process and that parts of the assessment and studies provided by the industry are considered confidential information, it is not yet possible to evaluate when, how often, and how WoE evaluation or SR actually are performed in chemicals regulation within EU.

5. How should WoE evaluation and SR be performed?

Guidance for how to perform WoE evaluation and SR is available in the guidance documents for five of the regulatory frameworks under investigation. However, the guidance varies in level of detail for the different frameworks (Table 3). In general, stepwise guidance that guides risk assessors through the WoE evaluation or SR process is rare. Aspects that, according to Cochrane's method for SR, should be considered in a stepwise guidance include: development of protocol, search strategy, criteria for including and excluding studies in the assessment, and evaluation methods for single studies and groups of studies, and synthesis of evidence. Considering these aspects, none of the investigated EU frameworks alone gives enough guidance for generating robust and reproducible WoE evaluations or SRs.

The REACH-guidance on how to report WoE is short-spoken regarding how the assessment should be performed; it only provides general recommendations to gather all relevant information, assess the overall package to conclude on an endpoint, and pool the information. No guidance is given for development of the protocol, search strategies, selection criteria or conducting evidence integration. Additional REACH-guidance documents do not add any substantial guidance regarding WoE evaluation or SR. EFSA's guidance for conducting SR provides relatively detailed guidance for setting up the review question, developing a protocol for the review, dictating the literature search strategy and collecting the relevant literature, in particular. It also provides some guidance, although less detailed, for the consequent steps of the review, i.e. evaluating data quality, synthesizing the data, as well as presenting and interpreting the results.

Several of the guidance documents available for the risk assessment of plant protection products lack specific guidance on WoE evaluation and SR, but three guidance documents contain at least some information on how to handle contradictory results and data of different reliability and relevance. The importance of evaluating and documenting uncertainties in a transparent manner is also emphasised in the guidance, and the risk assessor is reminded that the process involves expert judgement and that care should be taken to avoid subjective assessments and documentation. The guidance document for risk assessment of biocides highlights that WoE evaluation is a qualitative assessment, and that similarity of effects between humans and animals should be given more weight. The guidance document for CLP also gives some guidance regarding weighting of types of studies by saying that experimental studies should be given higher weight than read across.

6. Discussion

The main conclusion from this review is that there is limited guidance for how to perform WoE evaluation and SR in the regulatory frameworks investigated here. If the EU regulatory agencies want risk assessors to use these methods when performing chemical risk assessment, which they in general state, improved guidance is needed. Current legislation and guidance documents do not give sufficient directions for handling the complicated situations that can arise when dealing with the heterogeneous pool of data that can be expected in risk assessments of chemicals. This is especially important considering that future risk assessment likely will depend on non-animal data to a greater extent (e.g. within the cosmetics regulation and REACH legislation).

To safeguard reproducibility, and thereby credibility, of a decision making process, two things are needed. First, clear guidance that instructs and helps risk assessors when preparing and performing assessments in a scientific and unbiased manner. Second, detailed documentation of the process so that external evaluation and use of the assessment is made possible. This is especially important for processes

Please cite this article as: Ågerstrand, M., Beronius, A., Weight of evidence evaluation and systematic review in EU chemical risk assessment: Foundation is laid but guidance is needed, Environ Int (2015), http://dx.doi.org/10.1016/j.envint.2015.10.008

Table 3
Descriptions of WoE evaluation and SR in the different regulatory frameworks reviewed. Footnotes specify in which guidance document the information can be found. The available guidance has been compared to the guidance steps from the Cochrane Collaboration (Higgins and Green, 2009) and an overall comment on the content is provided.

Regulatory framework	How should WoE/SR be conducted?	Overall comment on the content of the WoE/SR guidance
Industrial chemicals	The following guidance is given: 1. Gather all relevant information: published literature, read across from chemical analogues/homologues, (Q)SAR predictions, data from existing studies, in vitro studies, epidemiological data/human experience, etc. 2. Assess the overall package to conclude on an endpoint. Pooling of information.[a]	Almost non-existent guidance. Guidance on development of protocol, search and selection of studies, and evidence integration are lacking.
Plant protection products	Examples of general guidance given: • Consider all relevant lines of evidence. • Evaluate the uncertainties associated with each line of evidence. • Form overall conclusions by using expert judgement to combine all lines of evidence, weighted according to their certainty, and give more weight to the most certain, but also take due account of the less certain. • Be sure to take full account of the uncertainties and to include a fair description of the range of possible outcomes in the final risk characterization. Identify the outcome that is considered most likely, but do not give it more emphasis than is justified by the evidence. If different lines of evidence conflict, this should be considered a form of uncertainty. No line of evidence should be completely discounted unless it is wholly invalid or irrelevant. • If the overall characterization of risk is expressed qualitatively, choose words very carefully to describe the outcome and its uncertainty as clearly as possible. • A weight-of-evidence assessment is inevitably subjective. Different assessors may vary in their weighing of the evidence, especially when uncertainty is high. Therefore, it is essential to document the assessment in detail, including the outcome and uncertainty for each line of evidence considered, and explaining how they were combined to reach conclusions about the overall outcome and its uncertainty. • A systematic tabular approach to documenting the weight-of-evidence assessment is proposed.[b]	Some stepwise instructions are given. Guidance on development of protocol is lacking. Some guidance on evidence integration is given. Importance of proper documentation is stressed.
	Guidance is based on three initial steps of the systematic review process: 1. Clarification of the objective of the review of the scientific literature and setting of the criteria for study relevance to the dossier 2. Searching for scientific literature 3. Selection of relevant scientific literature for inclusion in the dossier.[c]	Stepwise instructions are lacking for the later steps of SR. Guidance on evidence integration are lacking.
Biocides	Examples of general guidance given: • Evaluate (relevance and reliability) all available evidence. • Evaluate the severity, adversity and reversibility of effects. • Studies of high quality are given more weight than those of lower quality. • Studies enabling the identification of a NOAEL, and a robust hazard identification have a greater weight. • Similarity of effects between humans and animals is given more weight. • Well characterized mechanism or mode of action is used in the interpretation of observed effects. • WoE is not to be interpreted as simply tallying the number of positive and negative studies, nor does it imply an averaging of the doses or exposures identified in individual studies that may be suitable as starting points for risk assessment.[d]	Stepwise instructions are lacking. Guidance on development of protocol, search and selection of studies are lacking. Some guidance on evidence integration is given.
Cosmetics	Not mentioned	–
Human pharmaceuticals in the environment	Not mentioned	–
Veterinary pharmaceuticals in the environment	Not mentioned	–
Contaminants in food	Guidance is given for the following steps: 1. Develop review protocol and set review logistics. 2. Search for studies. 3. Select studies for inclusion or exclusion. 4. Collect data from included studies in evidence tables. 5. Assess methodological quality of included studies. 6. Synthesize data from included studies/meta-analysis. 7. Present data and results. 8. Interpret results and draw conclusions.[e]	Stepwise instructions are given. Some guidance on evidence integration is given.
Water framework directive	Not mentioned	–
Classification, labelling and packaging	Examples of general guidance given: • Using all available data. • The quality and consistency of the data shall be given appropriate weight. • Both positive and negative results should be assembled together. • Good quality data on the substance itself have more weight than such data extrapolated from similar substances • WoE assessment can be divided into two stages: assessment of each single test result and comparison of the weighed single test results.[f]	Stepwise instructions are lacking. Guidance on development of protocol is lacking. Some guidance on evidence integration is given.

[a] European Chemicals Agency, 2010.
[b] European Food Safety Authority, 2013.
[c] European Food Safety Authority, 2011.
[d] European Chemicals Agency, 2015a; European Chemicals Agency, 2015b.
[e] European Food Safety Authority, 2010.
[f] European Chemicals Agency, 2015a; European Chemicals Agency, 2015b.

Please cite this article as: Ågerstrand, M., Beronius, A., Weight of evidence evaluation and systematic review in EU chemical risk assessment: Foundation is laid but guidance is needed, Environ Int (2015), http://dx.doi.org/10.1016/j.envint.2015.10.008

such as chemical risk assessment, as well as other decision-making processes, that are influenced by expert judgement. Evaluations of traditional risk assessments, where one key study is used as the main basis for risk assessment, show that experts often disagree concerning the health or environmental risks of a chemical, or the risk management necessary to reduce that risk (Schenk, 2010; Beronius et al., 2010; Rudén, 2002; Ågerstrand and Rudén, 2010; Whaley et al., 2015). Absence of (or unclear) guidance is one possible reason for disagreement since it is then replaced by expert judgement, which may differ depending on e.g. previous knowledge and experience. For the majority of the chemicals on the EU-market today health and environmental risk assessments are performed by the producing or importing company (Table 1). This implies that there is a large number of risk assessors with varied education and experience responsible for these assessments. It is consequently reasonable to assume that clear guidance for conducting and reporting WoE evaluation and SR would improve the robustness, reproducibility, transparency, and thereby applicability of such processes.

There are differences between the nine regulatory frameworks scrutinized in this review concerning how WoE evaluation and SR in chemical risk assessment is promoted, defined and described. The use of WoE evaluation or SR is mentioned in seven of the nine investigated frameworks, and explained in slightly more detail in five. However, despite the differences there seems to be little contradiction between the frameworks, i.e. similar use is recommended and the guidance address similar aspects, but in different level of detail. Many of the chemicals legislations within the EU have been updated and implemented within the last few years, during that time discussions on the use of WoE evaluations and SR within chemicals regulation has evolved.

The authority in the EU that provides the most detailed guidance for SR is EFSA, and their guidance shows the clearest influence from the fundamental principles of SR: methodological rigour, transparency, and reproducibility (Higgins and Green, 2009; Eden et al., 2011). In contrast, ECHA promotes WoE evaluation but provides very little guidance. This implies to stakeholders that WoE evaluation has not been prioritized, and/or that there is a belief that no guidance in needed for such assessment. This can also be seen in two quotes from ECHA's guidance documents: "*from daily life everybody is familiar with the essence of Weight of Evidence reasoning and its basic mechanism may be regarded as a matter of common sense*" (European Chemicals Agency, 2011) and "*there can be no firm rules to the conduct of a Weight of Evidence assessment as this process involves expert judgement and because the mix and reliability of information available for a particular substance will probably be unique*" (European Chemicals Agency, 2015b). The viewpoint that detailed guidance for WoE (or SR) is not needed or appropriate, is not shared by several other institutions providing guidance on WoE evaluation and SR in chemical risk assessment. The European Commission's Scientific Committee on Emerging and Newly Identified Health Risks (SCENIHR) has produced a detailed guidance for WoE evaluation as used by that committee (SCENIHR, 2012). Similarly, the International Agency for Research on Cancer (IARC) has also, after long experience of producing monographs, published guidance for WoE evaluation (IARC, 2006). In US two similar methods for SR have been developed by the National Toxicology Programme's Office of Health Assessment and Translation (OHAT) and an independent research group at University of California (Navigation Guide) (Rooney et al., 2014; Koustas et al., 2014). The experiences and guidance from these four institutions could, for example, be used in a future update of the nine EU regulatory frameworks investigated in this review.

The benefits of using SR for decision making in health care are many (Eden et al., 2011) and can in some sense be transferred to WoE evaluations and SR in chemicals regulations even though data tend to be more heterogeneous in the later. The SR methodology promotes a structural and clearly documented process that carefully assesses and communicates the scientific data to decision makers (Whaley et al., 2015), something that is needed especially if a large amount of studies, with possible conflicting results, are to be included in the assessment. How resource effective and successful in terms of providing decision support these methods are remains to be seen since only a few WoE evaluations or SRs have been performed in chemical risk assessments to date.

Other benefits with moving away from the traditional risk assessment approach of using a single key study are that a wider range of studies can be used. No single study can cover all sensitive and relevant endpoints. This has, for example, been highlighted in the case of endocrine disrupting chemicals where standard studies has proven to be insufficient in identifying effects (Kortenkamp et al., 2011; United Nations Environmental Programme, 2012; Beronius et al., 2014).

In conclusion, the foundation for use of WoE evaluation and SR is laid in the majority of the investigated frameworks, but there is a need to provide more structured and detailed guidance for these processes to risk assessors. Several EU regulatory agencies and authorities are engaged in providing guidance and, based on this review, EFSA can be considered as the one taking the lead. In order to make the process of developing guidance as efficient as possible, and to ensure smooth transfer of risk assessments between EU frameworks in cases where a chemical is for example risk assessed both as an industrial chemical and as a contaminant in food, it is recommended that guidance for a WoE evaluation and SR approach is developed jointly by the EU regulatory agencies. Clear guidance should be provided for conducting and reporting all steps in the process, i.e. for formulating the problem and stating the objectives, identifying relevant evidence, evaluating and integrating evidence and arriving at a conclusion. Evidence identification and evaluation, as well as other aspects of the risk assessment process, may of course vary between different chemical groups, such as biocides and cosmetics, but the overall processes for WoE evaluation and SR are likely to be the same irrespectively of chemical framework. Better and more homogenous guidance has the potential to increase transparency and credibility of the process, and to give risk assessors the opportunity to make use of all available data of sufficient reliability and relevance to provide the scientific basis for better targeted policy decisions for chemical risk reduction.

Conflict of interest

The authors declare no conflicts of interest.

Acknowledgements

Ågerstrand is funded by Mistra (The Swedish Foundation for Strategic Environmental Research). Beronius is funded by faculty grants from Stockholm University.

References

Ågerstrand, M., Rudén, C., 2010. Evaluation of the accuracy and consistency of the Swedish environmental classification and information system for pharmaceuticals. Sci. Total Environ. 408 (11), 2327–2339. http://dx.doi.org/10.1016/j.scitotenv.2010.02.020.

Beronius, A., Hanberg, A., Zilliacus, J., Rudén, C., 2014. Bridging the gap between academic research and regulatory health risk assessment of endocrine disrupting chemicals. Curr. Opin. Pharmacol. 19 (December), 99–104. http://dx.doi.org/10.1016/j.coph.2014.08.005.

Beronius, A., Rudén, C., Håkansson, H., Hanberg, A., 2010. Risk to all or none?: a comparative analysis of controversies in the health risk assessment of bisphenol A. Reprod. Toxicol. 29 (2), 132–146. http://dx.doi.org/10.1016/j.reprotox.2009.11.007.

EC, 2002. http://ec.europa.eu/food/plant/pesticides/guidance_documents/docs/wrkdoc19_en.pdf.

EC, 2004. http://ec.europa.eu/food/plant/pesticides/guidance_documents/docs/wrkdoc20_rev_en.pdf.

EC, 2006. http://ec.europa.eu/food/plant/pesticides/guidance_documents/docs/7531_rev_10_en.pdf.

EC, 2009. http://ec.europa.eu/food/plant/pesticides/guidance_documents/docs/report_birds_mammals_guidance_doc_sanco10997_2009_31_07_09.pdf.

Eden, J., Levit, L., Berg, A., Morton, S. (Editors), 2011. Finding out What Works in Health Care. Standards for Systematic ReviewsInstitute of Medicine of the National Academies.

European Chemicals Agency, 2008. REACH Guidance Documents (Helsinki, Finland).
European Chemicals Agency, 2010. Practical Guide 2: How to Report Weight of Evidence. https://echa.europa.eu/documents/10162/13655/pg_report_weight_of_evidence_en.pdf.
European Chemicals Agency, 2011. REACH Guidance Documents (Helsinki, Finland).
European Chemicals Agency, 2014. Factsheet: Interface between REACH and Cosmetics Regulations. https://echa.europa.eu/documents/10162/13628/reach_cosmetics_factsheet_en.pdf.
European Chemicals Agency, 2015a. Guidance on the Application of the CLP Criteria. Version 4.1. https://echa.europa.eu/documents/10162/13562/clp_en.pdf.
European Chemicals Agency, 2015b. Guidance on the Biocidal Products Regulation. Volume III Human Health - Part B Risk Assessment. Version 1.1. http://echa.europa.eu/documents/10162/15623299/biocides_guidance_human_health_ra_iii_partb_en.pdf).
European Commission, 2002. Guidance Document on Aquatic Ecotoxicology. Plant Protection Products, Brussels, Belgium.
European Commission, 2011. Common Implementation Strategy for the Water Framework Directive (2000/60/EC). Guidance Document No. 27. Technical Guidance for Deriving Environmental Quality Standards (https://circabc.europa.eu/sd/a/0cc3581b-5f65-4b6f-91c6-433a1e947838/TGD-EQS%20CIS-WFD%2027%20EC%202011.pdf).
European Food Safety Authority, 2009. Risk Assessment for Birds and Mammals. http://www.efsa.europa.eu/en/efsajournal/doc/1438.pdf.
European Food Safety Authority, 2010. Application of Systematic Review Methodology to Food and Feed Safety Assessments to Support Decision Making. http://www.efsa.europa.eu/en/search/doc/1637.pdf.
European Food Safety Authority, 2011. Submission of Scientific Peer-Reviewed Open Literature for the Approval of Pesticide Active Substances Under Regulation (EC) No 1107/2009. Parma, Italy. http://www.efsa.europa.eu/it/search/doc/2092.pdf.
European Food Safety Authority, 2013. Guidance on Tiered Risk Assessment for Plant Protection Products for Aquatic Organisms in Edge-of-Field Surface Waters. http://www.efsa.europa.eu/en/efsajournal/doc/3290.pdf.
EFSA, 2012. http://ec.europa.eu/food/plant/pesticides/guidance_documents/docs/efsa_guidance_document_dermal_absorption_18042012.pdf.
EMA, 2000. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/10/WC500004394.pdf.
EMA, 2005. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/10/WC500004393.pdf.
EMA, 2006. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/10/WC500003978.pdf.
EMA, 2015. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2015/03/WC500185159.pdf.
Higgins, J.P.T., Green, S. (Editors), 2009. Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0. www.cochrane-Handbook.org.
IARC, 2006. Monographs on the Evaluation of Carcinogenic Risks to Humans. Preamble (http://monographs.iarc.fr/ENG/Preamble/CurrentPreamble.pdf).
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. http://dx.doi.org/10.1006/rtph.1996.1076.
Kortenkamp, A., Martin, O., Faust, M., Evans, R., McKinlay, R., Orton, F., Rosivatz, E., 2011. State of the Art Assessment of Endocrine Disrupters. Final Report. http://ec.europa.eu/environment/chemicals/endocrine/pdf/sota_edc_final_report.pdf.
Koustas, E., Lam, J., Sutton, P., Johnson, P.I., Atchley, D.S., Sen, S., Robinson, K.A., Axelrad, D.A., Woodruff, T.J., 2014. The navigation guide—evidence-based medicine meets environmental health: systematic review of nonhuman evidence for PFOA effects on fetal growth. Environ. Health Perspect. http://dx.doi.org/10.1289/ehp.1307177 (June).
Krimsky, S., 2005. The weight of scientific evidence in policy and law. Am. J. Public Health 95 (Suppl. 1), S129–S136. http://dx.doi.org/10.2105/AJPH.2004.044727.
Linkov, I., Loney, D., Susan, C., Kyle Satterstrom, F., Bridges, T., 2009. Weight-of-evidence evaluation in environmental assessment: review of qualitative and quantitative approaches. Sci. Total Environ. 407 (19), 5199–5205. http://dx.doi.org/10.1016/j.scitotenv.2009.05.004.
National Research Council, U.S., 2014. Review of EPA's Integrated Risk Information System (IRIS) Process. http://www.nap.edu/catalog/18764/review-of-epas-integrated-risk-information-system-iris-process.
Rhomberg, L.R., Goodman, J.E., Bailey, L.A., Prueitt, R.L., Beck, N.B., Bevan, C., Honeycutt, M., et al., 2013. A survey of frameworks for best practices in weight-of-evidence analyses. Crit. Rev. Toxicol. 43 (9), 753–784. http://dx.doi.org/10.3109/10408444.2013.832727.
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. http://dx.doi.org/10.1289/ehp.1307972 (April).
Rudén, C., 2002. From Data to Decision. A Case Study of Controversies in Cancer Risk Assessments Doctoral Thesis Karolonska Institutet, Sweden.
SCENIHR, 2012. Memorandum on the Use of the Scientific Literature for Human Health Risk Assessment Purposes — Weighing of Evidence and Expression of Uncertainty. http://ec.europa.eu/health/scientific_committees/emerging/docs/scenihr_s_001.pdf.
Schenk, L., 2010. Comparison of data used for setting occupational exposure limits. Int. J. Occup. Environ. Health 16 (3), 249–262. http://dx.doi.org/10.1179/oeh.2010.16.3.249.
Scientific Committee on Consumer Safety, 2012. The SCCS's Notes of Guidance for the Testing of Cosmetic Substances and Their Evaluation, 8th Revision. SCCS/1501/12. http://ec.europa.eu/health/scientific_committees/consumer_safety/docs/sccs_s_006.pdf.
United Nations Environmental Programme, 2012. State of the Science of Endocrine Disrupting Chemicals. http://www.unep.org/chemicalsandwaste/UNEPsWork/StateoftheScience/tabid/105913/Default.aspx.
Weed, D.L., 2005. Weight of evidence: a review of concept and methods. Risk Anal. 25 (6), 1545–1557. http://dx.doi.org/10.1111/j.1539-6924.2005.00699.x.
Whaley, P., Halsall, C., Ågerstrand, M., Benford, D., Aiassa, E., Bilotta, G., et al., 2015. Implementing Systematic Review Techniques in Chemical Risk Assessment: Challenges, Opportunities and Recommendations (submitted for publication, Paper).

Please cite this article as: Ågerstrand, M., Beronius, A., Weight of evidence evaluation and systematic review in EU chemical risk assessment: Foundation is laid but guidance is needed, Environ Int (2015), http://dx.doi.org/10.1016/j.envint.2015.10.008