SAB REVIEW OF USEPA's ECOLOGICAL ASSESSMENT ACTION PLAN: CHARGE QUESTION 3

Use of the weight of evidence approach in ecological risk assessment

Peter Chapman, Allen Burton, Loveday Conquest, Richard Di Giulio

This Charge Question notes that, although ecological risk assessments (ERAs) often involve multiple lines of evidence (LOE), no guidance exists on how to weight those LOE to make inferences. The Action Plan proposes that USEPA develop such guidance. We agree. The SAB Ecological Processes and Effects Committee (EPEC) is asked to comment on the scientific merit and limitations of using a weight of evidence (WOE) approach in decision making and offer any guidance on weighing ERA LOE.

WOE has Scientific Merit

The scientific merit of WOE is clear and evidenced by the large number of scientific publications and its consistent and continuing use in ERA. For instance, in 2002 a series of articles on WOE were published in the journal *Human and Ecological Risk Assessment*. **Numerous articles on WOE have** since **been published** in a wide variety of journals and books. Reviews of WOE approaches (e.g., Burton et al. 2002a; Weed 2005; Linkov et al. 2009) uniformly recommend its usage, particularly in ERA, but also note the **need for transparency and guidance in its usage**.

USEPA recognizes that "today's environmental challenges are increasingly subtle and complex", and that research must not be just inter-disciplinary but in fact trans-disciplinary, "combining perspectives to form entirely new concepts and reach new levels of scientific understanding" (Anastas 2012). **WOE clearly has scientific merit both inside and outside of USEPA; this merit has been affirmed by EPEC in previous advice to the Agency (USEPA SAB 2007).**

In order to **have a scientifically rigorous WOE process it must rely less on best professional judgment (BPJ) and more on statically based decision points**. This will not be possible without USEPA providing WOE guidance that is program specific, ideally structured decision-making frameworks. **Specific recommendations are required regarding the use of WOE.**

As we gain better understanding of how to relate scale in time and place to stressor intensity and as we develop better understanding of baseline ecological conditions, the use of WOE should not be as necessary. Presently, there can be much argument over how much weight to give certain LOE (see below). As a better understanding arises concerning multiple and complex stressors, and as these are related to life history parameters (See Charge Question 7), we would expect less argument over the strongest LOE.

Limitations to WOE

The term weight of evidence‖ (WOE) appears to have a variety of interpretations in the context of risk assessment. It begins with the general idea that more than a single line of inquiry is desirable when assessing risk. At issue is just how to integrate and synthesize evidence from different studies. The

studies might not all measure the same thing (e.g., chemical responses, individual organism responses, community responses).

Rothman and Greeland (2005), based on the classic paper by Hill (1965) on causes of occupational diseases, listed the following causality criteria:

1. Strength of the association (stronger associations support the notion of causality);
2. Consistency (more studies find similar results);
3. Specificity (specific exposures exert specific effects and, at the same time, certain exposures can lead to multiple effects);
4. Temporality (exposure should precede the effect);
5. Biological gradient (a dose-response relationship lends evidence to causality);
6. Plausibility (knowledge of biological processes involved lends evidence to causality);
7. Coherence (other observed biological effects lend evidence to a causal association);
8. Experimental evidence (e.g., does the amount of a toxicant in a body of water decrease following changes in practice by an industrial plant?); and,
9. Analogy (does a similar agent exert similar effects?).

The above were originally formulated in the context of potential carcinogens and human disease, and can also be interpreted in terms of ERA. Simply contemplating all the items on this list serves as a reminder that there is a lot involved in trying to quantify the process involved in the idea of exposure to something (e.g., a toxicant, a management practice) and a resulting effect.

USEPA (2005) has used a WOE approach in the context of carcinogens and toxicology. However, it seems unlikely that ERAs are as amenable to formalization as, say, human cancer risk assessments, which are not totally straight-forward either. **Many ERAs are inherently unique, and a high degree of flexibility to address the nuances associated with a particular assessment will remain desirable for the foreseeable future.**

A uniform definition for WOE does not exist. The definition of Burton et al. (2002b) is likely the best definition at present because it does not unduly limit the concept: **"a process used in environmental assessment to evaluate multiple lines-of-evidence concerning ecological condition".** The EPEC has previously (USEPA SAB 2007) also described ERA "as a process, not just a technique."

The challenges inherent in using WOE for decision-making are well known (Batley et al. 2002; Burton et al. 2002a,b; Wenning et al., 2005). **We agree that more instructive and consistent guidance is needed for using WOE approaches**. WOE approaches are most often based on BPJ, and vary widely in their scientific rigor and statistical credibility (Burton et al. 2002a). Consequently, they may not reduce uncertainty as they are meant to and may confound effective decision-making.

WOE depends to a certain extent on BPJ, which varies depending on the professionals making those judgments (e.g., Bay et al. 2007; Thompson et al. 2012). USEPA (2010) has identified BPJ as a source of uncertainty. Lack of agreement among experts extends beyond the environmental sciences (e.g., Large

and Nielssen 2008). **Bay et al. (2007)** suggest that uncertainty related to the use of BPJ must be recognized in ERA, and will be less important at the extremes (e.g., sites that are clearly contaminated and toxic, and those that are clearly not) than between the extremes. They **recommend three steps to reduce uncertainty in the integration and interpretation of multiple LOE:**

1. **Key elements of the assessment strategy (e.g., relative weight of each LOE, how multiple LOE will be combined [e.g., scores, ranks, logic frameworks], criteria for determining the ERA conclusion) should be determined during the Problem Formulation phase of the ERA.**
2. **Guidance is required on the specific methodology/methodologies for measuring and assessing each LOE.**
3. **Training, including guidance documents, is required for individuals interpreting both individual LOE and the overall WOE.**

The **EPEC has previously (USEPA SAB 2007) similarly recommended, and continues to recommend: development of a consistent approach in ERA to interpreting LOE and WOE, both to reduce uncertainty and to assist in decision-making based on ERA; exploration of the use of such methods as Bayesian analysis and causal argumentation to develop hypotheses or risk questions focused on causal relationships and WOE; and, "case studies and/or standards of practice for interpreting lines of evidence and weight of evidence with an emphasis on application in decision making." We agree with USEPA (2003) that case studies should be developed and should in particular focus on whether some LOE carried more weight than others or whether they were ignored or too difficult to interpret or use. This information will assist in future weighting of ERA LOE.**

<u>Guidance on Weighting ERA LOE</u>

WOE is an approach to evaluating and integrating multiple sources of evidence, rather than a single technique. As such, WOE should follow certain principles, but not a particular recipe nor algorithm. Any effort that claims to use WOE to reach conclusions should be completely transparent regarding the different sources of evidence considered and any qualitative (e.g., expert opinion) or quantitative weighting schemes used (see Swaen and van Amelsvoort 2009, albeit in an epidemiology/carcinogen setting). Also meriting consideration are data quality and reliability of different studies. Weed (2005) points out that applying an **arbitrary weighting scheme without a solid theoretical foundation to integrate different lines of evidence into a single risk score may not actually improve decision making.**

The most specific WOE approach is meta-analysis, used when different studies have provided estimates of the same effect. The estimated effects coming out of the different studies are weighted (inversely proportional to the variance associated with the effect) and combined together to form a weighted average effect (with a weighted variance). In this manner, for instance, the presence of many studies with almost statistically significant|| results can lead to an overall, statistically significant, result.

A well developed WOE framework would be able to assign quantitative weights to results from different studies (with associated estimates of uncertainty), and to combine them into an assessment of a defined

risk. Thus far, this has been largely done in epidemiological contexts. In the area of ERA, having quantitative results from adaptive management experiments based on sound principles of statistical design would make it easier to construct WOE arguments on ecological risk.

Useful statistically-based WOE approaches have been reported that address many of the weaknesses of qualitative based approaches (Bailer et al. 2002; Burton et al. 2002b; Grapentine et al. 2002; Kapo and Burton 2006; Kapo et al. 2008; Reynoldson et al. 2002). Examples such as these provide a solid basis for USEPA guidance that can be structured towards unique program needs.

The WOE process should be described in the Problem Formulation stage and ensure credible stakeholder input and a transparent understanding of what constitutes reference condition, restoration goals, remedy objectives, and/or ecological impairments in the context of site spatial and temporal variations. This point was highlighted within the SAB report on improving the ERA process, which resulted in this current RAF process (Dale et al. 2008).

In terms of weighting ERA LOE, we have three specific recommendations:

1. Building on the recommendations of Wenning et al. (2005) we propose that, **in general, chemical LOE receive less weight than biological LOE, and that LOE that involve individual organisms and species receive less weight than LOE that involve resident natural communities and populations of organisms.**
2. We **counsel against arbitrary numerical weightings** as site- and situation-specific considerations will affect weightings. WOE assessments need to be "flexible, transparent and defensible…[with] sufficient flexibility to accept all relevant evidence and generate creative solutions to difficult problems" (Suter and Cormier 2011). We agree with USEPA (2010) that "weighing of evidence should be considered during each problem formulation, and a method for weighing evidence should be included, as appropriate, in the analysis plan."
3. We **suggest further investigation of multicriteria decision analyses (MCDA)** as recommended by Linkov et al. (2011): "Each WOE method is based on a unique rationale and capable of considering a different scope of LOEs. Thus, each method has specific benefits and drawbacks. The different nature of methods means that one cannot *a priori* determine the superior method for a particular application. One must consider the method employed in addition to the evidence."

Beyond the weighting of different LOE, **WOE has progressed over the years but as it is currently practiced is a qualitative tool without a probabilistic basis**. A WOE is essentially a Bayesian approach without a realization of the calculation. Evidence should be taken that can differentiate between alternative hypotheses. As discussed by Newman et al. (2007) there are ways to perform specific calculations and to use Bayesian networks to improve analyses conducted within risk assessments. For example, Bayesian networks can be tied directly to the cause-effect conceptual model that should be generated for every risk assessment**. WOE is probably best in deciding between alternative hypotheses, using Bayesian approaches.**

We recommend development of specific quantitative guidance with associated case studies that can be used by risk assessors and risk managers. Such would clearly be a valuable addition to the risk assessment toolbox.

References

Anastas PT. 2012. Fundamental changes to EPA's research enterprise: The path forward. Environ Sci Technol 46: 580-586.

Bailer AJ, Hughes MR, See K, Noble R, Schaefer R. 2002. A pooled response strategy for combining multiple lines of evidence to quantitatively estimate impact. Human Ecol Risk Assess 8: 1597-1611.

Batley GE, Burton GA Jr, Chapman PM, Forbes VE. 2002. Uncertainties in sediment quality weight of evidence (WOE) assessments. Human Ecol Risk Assess 8: 1517-1547.

Bay S, Berry W, Chapman PM, Fairey R, Gries T, Long E, MacDonald D, Weisberg SB. 2007. Evaluating consistency of best professional judgment in the application of a multiple lines of evidence Sediment Quality Triad. Integr Environ Assess Manage 3: 491–497.

Burton GAJr, Chapman PM, Smith ES. 2002a. Weight-of-evidence approaches for assessing ecosystem impairment. Human Ecol Risk Assess 8: 1657-1673.

Burton GAJr, Batley GE, Chapman PM, Forbes VE, Smith EP, Reynoldson T, Schlekat CE, den Besten PJ, Bailer AJ, Green AS, Dwyer RL. 2002b. A weight-of-evidence framework for assessing sediment (or other) contamination: improving certainty in the decision-making process. Human Ecol Risk Assess 8: 1675-1696.

Dale VH, Biddinger GR, Newman MC, Oris JT, Suter GW II, Thompson T, Armitage TM, Meyer JL, Allen-King RM, Benfield EF, Burton GA, Chapman PM, Conquest LL, Fernandez IJ, Landis WG, Master LL, Mitsch WJ, Mueller TC, Rabeni CF, Rodewald AD, Sanders JG, van Heerden IL. 2008. Enhancing the ecological risk assessment process. Integr Environ Assess Man 4: 306-313.

Grapentine L, Anderson J, Burton GA Jr, Chapman PM, DeBarros C, Gaudet C, Johnson G, Marvin C, Milani D, Painter S, Pascoe T, Reynoldson T, Richman L, Solomon K, Turner R. 2002. Development of a decision-making framework for sediment assessment. Human Ecol Risk Assess 8: 1641-1655.

Hill AB. 1965. The environment and disease: Association or causation? J Royal Soc Medicine 58: 295-300.

Kapo K, Burton GA Jr. 2006. A geographic information system based weights-of-evidence approach for diagnosing aquatic ecosystem impairment. Environ Toxicol Chem 25: 2237-2249.

Kapo KE, Burton GA Jr, De Zwart D, Posthuma L, Dyer SD. 2008. Quantitative lines of evidence for screening-level diagnostic assessment of regional fish community impacts: a comparison of spatial database evaluation methods. Environ Sci Technol 42: 9412–9418.

Large MM, Nielssen O. 2008. Factors associated with agreement between experts in evidence about psychiatric injury. J Am Acad Psychiatry Law 36: 515-521.

Linkov I, Loney D, Cormier S, Satterstrom FK, Bridges T. 2009. Weight-of-evidence evaluation in environmental assessment: Review of qualitative and quantitative approaches. Sci Tot Environ 407: 5199-5205.

Linkov I, Welle P, Loney D, Tkachuk A, Canis L, Kim JB, Bridges T. 2011. Use of multicriteria decision analysis to support weight of evidence evaluation. Risk Anal 31: 1211-1225.

Newman MC, Zhao Y, Carriger JF. 2007. Coastal and estuarine ecological risk assessment: the need for a more formal approach to stressor identification. Hydrobiologia 577: 31–40.

Reynoldson TB, Smith EP, Bailer AJ. 2002. A comparison of three weight-of-evidence approaches for integrating sediment contamination within and across lines of evidence. Human Ecol Risk Assess 8: 1613-1624.

Rothman KJ, Greenland S. 2005. Causation and causal inference in epidemiology. Am J Public Health 95 (Suppl 1): S144-150.

Suter GW II, Cormier SM. 2011. Why and how to combine evidence in environmental assessments: Weighing evidence and building cases. Sci Tot Environ 409: 1407-1417.

Swaen G, van Amelsvoort L. 2009. A weight of evidence approach to causal inference. J Clinical Epidemiol 62: 270-277.

Thompson B, Weisberg SB, Melwani A, Lowe S, Ranasinghe JA, Cadien DB, Dauer DM, Diaz RJ, Fields W, Kellogg M, Montagne DE, Ode PR, Reish DJ, Slattery PN. 2012. Low levels of agreement among experts using best professional judgment to assess benthic condition in the San Francisco Estuary and Delta. Ecol Indic 12: 167-173.

USEPA (U.S. Environmental Protection Agency). 2003. Generic Ecological Assessment Endpoints (GEAEs) for Ecological Risk Assessment. EPA/630/P-02/004F. Risk Assessment Forum, Washington, DC, USA.

USEPA. 2005. Guidelines for Carcinogen Risk Assessment. EPA/630/P-03/001F. Washington, DC, USA.

USEPA. 2010. Integrating Ecological Assessment and Decision-Making at EPA: A Path Forward. Results of a Colloquium in Response to Science Advisory Board and National Research Council Recommendations. EPA/100/R-10/004. Risk Assessment Forum, Washington, DC, USA.

USEPA SAB (Science Advisory Board). 2007. Advice to EPA on Advancing the Science and Application of Ecological Risk Assessment in Environmental Decision Making. EPA-SAB-08-002. Washington, DC, USA.

Weed DL. 2005. Weight of evidence: A review of concept and methods. Risk Anal 25: 1545-1557.

Wenning R, Ingersoll C, Batley G, Moore M (eds.). 2005. Use of Sediment Quality Guidelines (SQGs) and Related Tools for the Assessment of Contaminated Sediments. SETAC Press, Pensacola, FL, USA.