Science of the Total Environment 409 (2011) 1406–1417

Contents lists available at ScienceDirect

Science of the Total Environment

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

Why and how to combine evidence in environmental assessments: Weighing evidence and building cases

Glenn W. Suter II *, Susan M. Cormier

National Center for Environmental Assessment, U.S. Environmental Protection Agency, 26 W. Martin L. King Drive, Cincinnati, OH, 45268, USA

ARTICLE INFO

Article history:
Received 13 September 2010
Received in revised form 15 December 2010
Accepted 21 December 2010
Available online 31 January 2011

Keywords:
Weight of evidence
Risk assessment
Ecological risk assessment
Environmental epidemiology
Kit fox
Lines of evidence

ABSTRACT

All types of environmental decisions benefit from assessments that assemble and analyze diverse evidence. The diversity of that evidence creates complexities that can be managed using an explicit, well-designed process. We suggest two adaptations from the legal lexicon, weight of evidence and building a case. When weighing evidence, weights are assigned to each piece of evidence, and then the body of evidence is weighed in favor of each hypothesis by amassing the weights. Finally, the total weights of evidence for the alternative hypotheses are compared to determine which alternative has the preponderance of evidence in its favor. When building a case, pieces of evidence are organized to show relationships among multiple hypotheses or complex interactions among agents, events, or processes. We provide processes for weighing evidence and building a case and illustrate both approaches in a case study involving the decline of a kit fox population. The general approach presented here is flexible, transparent, and defensible. During its development, it has been applied to risk assessments for contaminated sites and to causal assessments in aquatic and terrestrial systems. It is intended to balance the need for rigor and discipline with the need for sufficient flexibility to accept all relevant evidence and generate creative solutions to difficult environmental problems.

Published by Elsevier B.V.

1. Introduction

Environmental assessors regularly find that multiple pieces of evidence are relevant to an inference. These may include multiple estimates of a parameter, multiple models of a relationship, or multiple types of evidence (e.g., a laboratory test, a field experiment, and an observational field study). Assessors may simply choose one piece of evidence and ignore the others, or they may consider all relevant evidence in a "weight-of-evidence process." Arguments for and against combining evidence are presented in Appendices A and B.

The phrase, weight of evidence (WoE), is used commonly but inconsistently and often vaguely. Dale et al. (2008) concluded that "An approach to interpreting lines of evidence and weight of evidence is critically needed for complex assessments, and it would be useful to develop case studies and/or standards of practice for interpreting lines of evidence." Similarly, Stahl (1998) complained that the U.S. Environmental Protection Agency's guidelines for ecological risk assessment lack guidance on weight-of-evidence approaches. Existing reviews (most notably, Weed (2005); Krimsky (2005), and Linkov et al. (2009)) have described existing practices but have not provided methodological guidance.

We believe that the lack of consensus concerning WoE is largely due to a lack of agreement about what it is and what it does. The term is used for approaches that range from genuine weighing of commensurable pieces of evidence to a process of interrelating heterogeneous evidence that we call "building a case." In this paper, we attempt to provide a useful theory and practice of combining evidence for environmental assessment. We do this by presenting the two metaphors (weighing and building), and then provide a simple general framework, alternative methods for implementing it, and an approach that we find to be generally useful. However, we do not intend to prescribe a particular methodology. We believe that it would be impractical for two reasons. First, the diversity of applications is too great. Second, most assessors develop their own methods that fit their preferences and those of their stakeholders and decision makers. Rather, our intent is to provide assistance to assessors as they decide how to combine evidence to solve problems.

We present alternative methods for combining evidence and recognize that there are potentially more methods. More than one method may be applied in a case to different evidence or in different stages of the assessment process. However, the most important advice is to use a formal method and to be explicit about what method you are using (Suter and Cormier, 2010).

1.1. Two legal metaphors

The concepts of weight of evidence (WoE) and building a case (BaC) can both be traced to jurisprudence. WoE is represented by the scales of justice that balance the weight of the evidence for guilt against that for innocence or for one party against another. Pieces of

* Corresponding author. Tel.: +1 513 569 7808; fax: +1 513 569 7475.
E-mail addresses: suter.glenn@epa.gov (G.W. Suter), cormier.susan@epa.gov (S.M. Cormier).

0048-9697/$ – see front matter. Published by Elsevier B.V.
doi:10.1016/j.scitotenv.2010.12.029

evidence of different weights are placed in the appropriate pan of Justice's scales. The side that is lowest in the end prevails.

If one party bears the burden of proof, a standard weight (i.e., sufficient WoE) is placed in one pan. Evidence exceeding the sufficient weight must be loaded in the other pan to confirm that party's position. There may be multiple standard weights if there are multiple possible outcomes (e.g., human carcinogen, probable human carcinogen …not a carcinogen).

In adapting this judicial model to environmental assessment, we weight each piece of evidence, then we weigh the body of evidence in favor of each hypothesis by amassing the weights, and finally we compare the combined WoE for the alternative hypotheses to determine which has a preponderance of evidence in its favor. In environmental assessment, this weighing is typically done implicitly by professional judgment (Weed, 2005).

This judicial model fits cases that are genuinely analogous to a trial in that the assessment process must reach a dichotomous decision. Is the chemical a carcinogen, did this effluent cause the impairment, will the pesticide cause bird kills, etc.? However, many assessment questions involve magnitude, probability, or frequency, so the weighing of evidence must combine estimates as well as weights. What is the slope of the dose–response model, what is the probability that cadmium was the cause, how frequently will this pesticide cause bird kills, etc.? Hence, the weighing of environmental evidence must include processes for combining information as well as weights.

In sum, weighing evidence is a synthetic process that combines the information content of multiple weighted pieces of evidence. The information may be dichotomous (supports or not), quantitative values (e.g., an exposure or risk estimate), qualitative properties (e.g., large, medium or small), or a model. The weights that are applied to the information may express various properties that affect its credibility or importance and the weights themselves may be qualitative or quantitative. The combining of evidence may be a simple quantitative operation (e.g., weighted averages of concentration estimates) but more often involves difficult qualitative judgments.

The metaphor of building a case implies a very different process. Pieces of evidence are not simply equivalent discrete masses, but rather multiple parts of a structure or device. The constructed arguments may show relationships among multiple hypotheses. Fans of courtroom dramas are familiar with instances of this metaphor in which a case appears weak until a few critical pieces of evidence are provided that make the case fit together and explain how the crime occurred. Hence, under this metaphor, an assessor should be concerned about how the evidence might be logically combined rather than with which hypothesis has the weightiest body of evidence.

We believe that both metaphors are potentially useful. Some instances of combining evidence are simply a matter of weighing, some of building a case, and many require both.

2. Background to weighing evidence

2.1. Reasons for weighting and weighing evidence

Many assessment methods that are called weight of evidence combine evidence without explicitly assigning weights to pieces of evidence. This practice implies that all evidence is equally strong and of equal quality. That presumption is improbable. Even if all evidence was generated using high quality methods by people who never make mistakes, it is unlikely that all pieces and types of evidence provide equally strong or clear information. For example, observational and experimental data differ inherently. If you do not explicitly weight the evidence, you must either ignore those differences or consider them implicitly. Implicit weighting is not transparent to reviewers and stakeholders and may be subject to unconscious biases or incomplete logic.

2.2. Weighing evidence in different types of assessments

Although frameworks for performing "Weight of Evidence Assessments" have been proposed (Burton et al., 2002), we suggest that weighing evidence is not a particular type of assessment, but rather a method for planning, analyzing, or synthesizing information in various types of assessments. Specifically, WoE can be applied to each type of environmental assessment in the fully integrated framework (Fig. 1) (Cormier and Suter, 2008).

Condition assessments analyze monitoring data to determine whether environmental goals are being achieved that protect human health and ecosystems. WoE is applied when more than one measure of condition is available. For example, sport fishing records and data from electrofishing, seining, or snorkeling may be combined to determine whether a trout fishery is impaired (Wiseman et al., 2010).

Causal assessments use different pieces and types of evidence to determine whether an apparent association of cause and effect is actually causal (Suter et al., 2002) (http://www.epa.gov/caddis).

Weighing evidence has been a standard approach to causal assessment since the U.S. Surgeon General's Commission and A.B. Hill used it to demonstrate that smoking causes lung cancer.

Predictive assessments include risk and management assessments. Risk assessments estimate the nature, magnitude, and probability of effects for alternative policies or management actions. Management assessments may identify a preferred management action by weighing multiple types of evidence concerning benefits, costs, risks, public preferences, technical feasibility, and other considerations. The weighing is usually performed informally, but multi-criteria decision analysis, cost-benefit analysis, net benefit analysis, or other formal methods may be employed to weigh the evidence.

Outcome assessments determine whether a management action has succeeded. Although they seldom weigh evidence, outcome assessments could benefit from multiple categories of evidence, particularly in difficult or controversial cases. For example, to determine the outcome of a remedial action for contaminated sediment, one might apply all three components of the sediment quality triad (Chapman, 1990).

2.3. Steps in the assessment process involving weighing evidence

Environmental assessments of all types have three steps: planning, analysis, and synthesis (Fig. 2). The steps have different names in different types of assessments and different contexts. For example, for

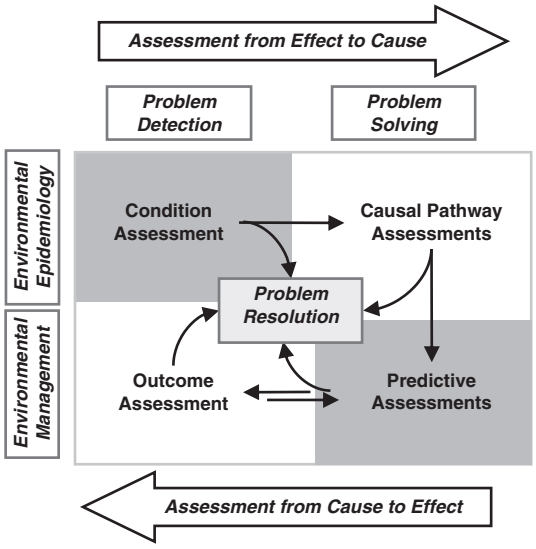

Fig. 1. The basic structure of an integrated framework for environmental assessment (Cormier and Suter, 2008).