Integrated Environmental Assessment and Management — Volume 2, Number 4—pp. 302–311

Review

Sequential Analysis of Lines of Evidence—An Advanced Weight-of-Evidence Approach for Ecological Risk Assessment

Ruth N Hull†* and *Stella Swanson‡*

†Cantox Environmental, 1900 Minnesota Court, Suite 130, Mississauga, Ontario L5N 3C9, Canada
‡Golder Associates, 10th Floor, 940 6th Avenue, SW, Calgary, Alberta T2P 3T1, Canada

(Received 3 August 2005; Accepted 5 December 2005)

ABSTRACT

Weight-of-evidence (WOE) approaches have been used in ecological risk assessment (ERA) for many years. The approaches integrate various types of data (e.g., from chemistry, bioassay, and field studies) to make an overall conclusion of risk. However, the current practice of WOE has several important difficulties, including a lack of transparency related to how each line of evidence is weighted or integrated into the overall weight-of-evidence conclusion. Therefore, a sequential analysis of lines of evidence (SALE) approach has been developed that advances the practice of WOE. It was developed for an ERA of chemical stressors but also can be used for nonchemical stressors and is equally applicable to the aquatic and terrestrial environments. The sequential aspect of the SALE process is a significant advancement and is based on 2 primary ideas. First, risks can be ruled out with the use of certain lines of evidence, including modeled hazard quotients (HQs) and comparisons of soil, water, or sediment quality with conservative soil, water or sediment quality guidelines. Thus, the SALE process recognizes that HQs are most useful in ruling out risk rather than predicting risk to ecological populations or communities. Second, the SALE process provides several opportunities to exit the risk assessment process, not only when risks are ruled out, but also when magnitude of effect is acceptable or when little or no evidence exists that associations between stressors and effects may be causal. Thus, the SALE approach explicitly includes interaction between assessors and managers. It illustrates to risk managers how risk management can go beyond the simple derivation of risk-based concentrations of chemicals of concern to risk management goals based on ecological metrics (e.g., species diversity). It also can be used to stimulate discussion of the limitations of the ERA science, and how scientists deal with uncertainty. It should assist risk managers by allowing their decisions to be based on a sequential, flexible, and transparent process that includes direct toxicity risks, indirect risks (via changes in habitat suitability), and the spatial and temporal factors that can influence the risk assessment.

Keywords: Weight of evidence Causality Sequential analysis of lines of evidence

INTRODUCTION

Weight-of-evidence (WOE) analysis is a risk characterization process used in ecological risk assessment (ERA) by which multiple measures are related to an assessment endpoint (Menzie et al. 1996) for a particular receptor. A recent review of WOE approaches (Burton, Chapman, et al. 2002) summarized some of the advantages and limitations of each type of approach. Many examples of use of the WOE approach come from the aquatic environment because WOE approaches are more advanced for aquatic than terrestrial environments. For example, use of the sediment quality triad (Chapman 1990) has become common practice. The WOE approach also has been used by Lowell et al. (2000) to distinguish effects of multiple stressors in river systems. Aquatic toxicity tests and quantitative aquatic field surveys (e.g., fish health studies, benthic invertebrate community analysis) have been conducted much longer than their terrestrial counterparts. These tests and surveys have standardized methods and have been conducted routinely as part of regulatory programs (e.g., to regulate effluents) for several years. Reliability of results, the definition of what constitutes an effect, and the use of several lines of evidence to identify potential cause–effect relationships have had greater use in aquatic than terrestrial assessments (e.g., the US Environmental Protection Agency [USEPA 2000b] Stressor Identification Guidance). Therefore, a greater body of experience is available to support judgments about the integration of data into a WOE risk characterization for the aquatic environment.

In contrast, terrestrial ERAs have relied more on the comparison of chemical concentrations in the environment to benchmark levels, as well as food chain modeling for wildlife. Field-based lines of evidence are used infrequently, except in the context of establishing habitat suitability according to vegetation community information and simple animal surveys (e.g., small mammal trapping for abundance measures). Recently, additional lines of evidence appropriate to various levels of complexity of a wildlife ERA have been identified (Fairbrother 2003).

Traditionally, 3 characteristics of measures have been identified as critical to the WOE process (Menzie et al. 1996): The weight assigned to each measure, the magnitude of response observed in the measure, and the concurrence among multiple measures. When evaluating the lines of evidence, consideration must also be given to the adequacy and quality of the data, the degree and type of uncertainty associated with the line of evidence, and the relationship between the evidence and the risk assessment objectives (USEPA 1998). The process of weighing the evidence amounts to determining what overall risk statement is best supported by the individual lines of evidence (Suter 1996).

The current practice of WOE (Burton, Chapman, et al. 2002; Adams 2003) suffers from several difficulties. The

* To whom correspondence may be addressed:
rhull@cantoxenvironmental.com

weight assigned to each measure in risk assessment often involves considerable professional judgment and usually incorporates criteria such as the strength of the evidence for cause–effect relationships and ecological relevance. Establishing causation is challenging, especially if the primary lines of evidence are field based, with high natural variability and many confounding variables (Burton, Batley, et al. 2002). A confident assessment of the magnitude of the response requires a valid and credible baseline or reference condition against which the response is measured. Alternatively, if one disagrees with the concept of reference condition, an understanding is needed of the role of the stressors of concern in the process of conditioning the community and thus influencing the measured trajectory (e.g., seral stage, age or size distribution over time) of the populations or communities being assessed (Matthews et al. 1996; Landis et al. 2000; Sandberg and Landis 2001). In addition, evaluation of concurrence among measures must take into account that modeled and measured lines of evidence might address uncertainty in different ways. Thus, the inherent bias within each measure will almost certainly be on different scales. These difficulties cause problems with the transparency and consistency of WOE analyses.

The USEPA (1999) states that a WOE risk characterization should include sufficient information to make a reasoned decision about causality between levels of contamination and effects; whether the observed or predicted adverse effect is of sufficient magnitude, severity, areal extent, and duration that the local populations or communities will not be able to maintain themselves in a healthy state; and whether these effects appear to exceed the natural changes in the components typical of unaffected areas. This guidance from the USEPA raises a number of issues, notably 1) how to deal with establishing cause–effect relationships in the presence of multiple stressors from several sources, 2) the knowledge required to make a credible judgment regarding how much perturbation is required to change to an unhealthy state, and 3) what constitutes natural changes in highly dynamic populations, communities, and ecosystems.

The current practice of WOE has weaknesses that limit its application, especially to landscape-scale ERA. This paper describes a WOE approach that addresses these areas of vulnerability and advances current practice.

Several advances of the WOE approach are described in this paper. The approach was developed during completion of the ecological risk assessment for the off-site area of the Teck Cominco Metals smelter in Trail, British Columbia, Canada. It grew from attempts to use more traditional approaches for risk characterization that could not sufficiently address certain issues, such as those of effects to habitat suitability or how to weigh diverse lines of evidence. This new approach includes assessment of both direct risks (in which the chemical exerts its toxicity directly on the organism) and indirect risks (in which the chemical affects the organism through toxicity to the food or plant community that makes up its habitat). It is sequential; that is, rather than presenting all lines of evidence in a matrix, it proceeds through a sequence of steps to produce an increasingly focused assessment. It incorporates the spatial and temporal context of the ERA. It includes a transparent process for evaluating the magnitude of response and causation. It is flexible and can be used for terrestrial or aquatic ERAs across a range of assessment scopes, at scales from local to landscape. It also allows risk management to go beyond the simple derivation of risk-based concentrations of chemicals of concern to risk management goals on the basis of ecological metrics (e.g., species diversity).

The primary purpose of this paper is to extend WOE practice to terrestrial systems with reference to aquatic systems (in which the practice of WOE is much more established). However, the innovations within the approach presented in this paper are equally applicable to all ecosystems and media.

SEQUENTIAL ANALYSIS OF LINES OF EVIDENCE APPROACH

The sequential analysis of lines of evidence (SALE) approach is adapted from the 3-step process suggested by Lowell et al. (2000) regarding assembling a WOE for risk characterization (i.e., establishing causality, defining acceptable change, and linking environmental components in a decision-making framework). It includes an explicit analysis of causality. It defines acceptable change via an analysis of the magnitude of response to the stressors of concern. It provides several opportunities to proceed to the evaluation of risk management options. In addition, it incorporates guidance from Fairbrother (2003), Suter (1996), the USEPA (1999, 2000b), and Landis and coworkers (Landis and Wiegers 1997; Landis 2002).

The sequential nature of the SALE process is a significant advancement on current WOE practice. The sequence inherent in SALE is based upon 2 primary ideas. First, risks can be ruled out with the use of certain lines of evidence. Second, the decision to end the risk assessment, complete additional risk assessment, or proceed to a risk management evaluation can be made at various stages of the WOE analysis. These 2 ideas, together with the 3-step process of Lowell et al. (2000), led to the development of the sequence of steps shown in Figure 1.

The approach presented in Figure 1 allows for iterations to address site-specific uncertainties as they arise, encourages common sense, and includes ruling out risk as part of the process. It is not necessary to define a priori the acceptable level of uncertainty. Often, data are insufficient to develop uncertainty benchmarks. Furthermore, qualitative information that provides useful input to a WOE is not amenable to development of traditional statistically based uncertainty benchmarks. Rigid adherence to the requirements for quantitative decision rules within the formal data quality objectives process (USEPA 2000a) is not appropriate for risk management goals that go beyond risk-based chemical concentrations. It can also prevent fruitful risk management discussions because assessors and regulators can become caught in a loop of unproductive discussions driven by the mismatch between the broader WOE approach and guidance written for a single line of evidence.

The SALE approach explicitly includes interaction between assessors and managers. It can help create greater understanding between assessors and managers by making the process of using several lines of evidence more clear. It can be used to stimulate discussion of the limitations of the ERA science and how scientists deal with uncertainty.

The SALE approach illustrates to risk managers how the decision to proceed to risk management evaluation can go beyond the simple derivation of risk-based concentrations of chemicals of concern. Rather, risk management goals can be described via ecological metrics (e.g., species diversity) and

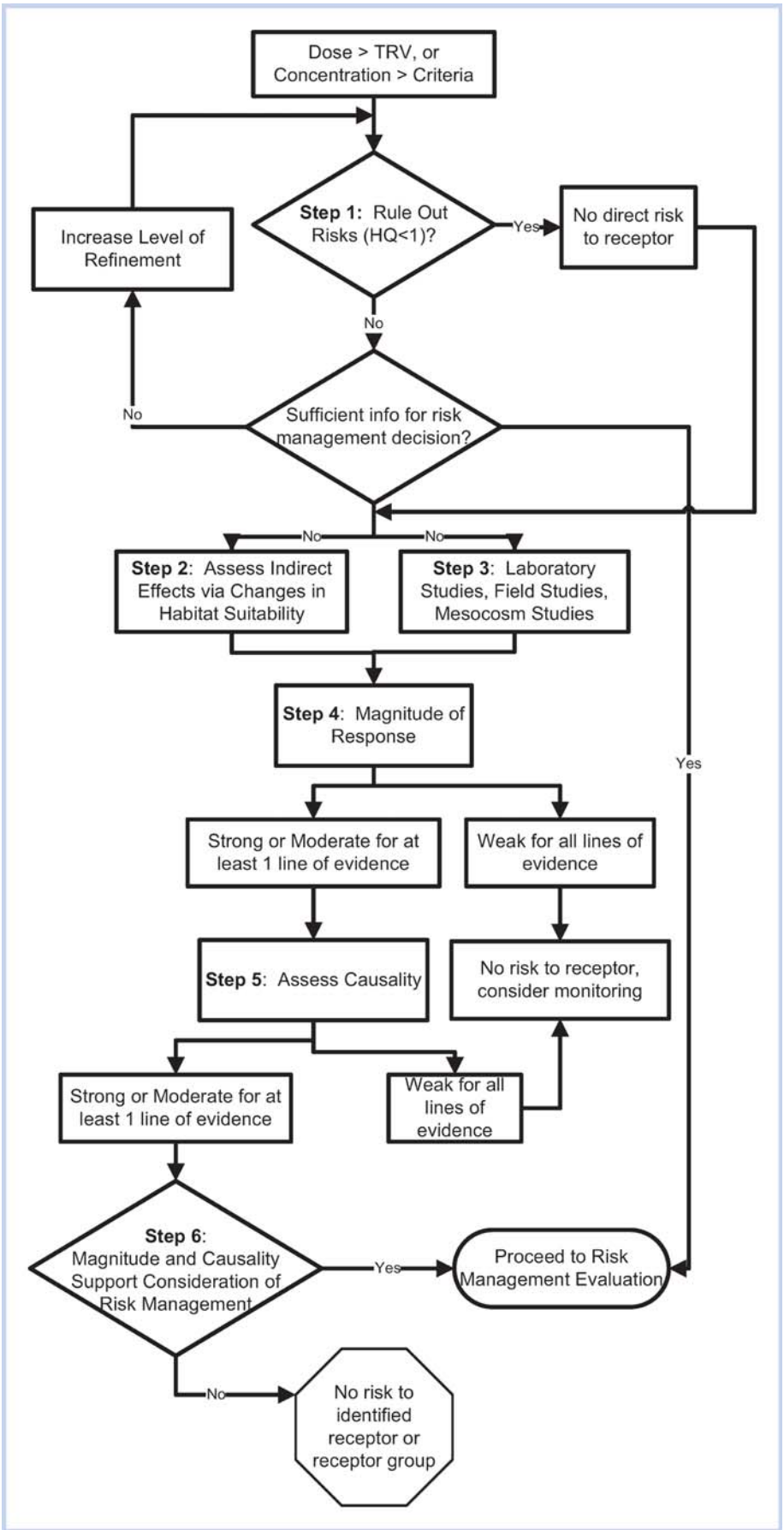

Figure 1. Sequential analysis of weight of evidence, a new system for assembling the weight of evidence. TRV= toxicity reference values; HQ = hazard quotients.

not as a chemical concentration (e.g., 500 ppm chemical in soil).

Step 1—Rule out risks through comparison of dose or concentration with effects thresholds or criteria

The 1st step in the SALE process is to identify where risks can be ruled out. This is done through modeling (e.g., food chain modeling for wildlife, body burden modeling for aquatic organisms or earthworms) or through comparisons of measured or modeled chemical concentrations in media to criteria or benchmarks for aquatic organisms, plants, and so on. Food chain modeling is conducted for specific receptors that are selected for evaluation in the ERA as representatives of particular feeding guilds and trophic levels. The predicted tissue concentrations or doses are then compared with tissue residue benchmarks (e.g., for metal concentration in fish) or toxicity reference values (TRVs).

The 1st decision point in the SALE approach occurs at this step because the hazard quotients produced by food chain modeling and comparison to conservative effects thresholds are most useful in ruling out risks rather than predicting risks (particularly to populations). Therefore, if modeling of direct toxicity results in hazard quotients (HQs) less than 1, direct toxicity risks are ruled out for chemical–receptor–area combinations. The risk assessment must use appropriate methods when completing step 1 and should consider the potential for interactions between chemicals that might act via a similar mechanism on the same target organ (e.g., some polycyclic aromatic hydrocarbons or divalent cations). In addition, the risk assessment must use methods that are conservative enough so as to ensure confidence in the process of ruling out risks. The remainder of the assessment of direct toxicity is then focused on chemical–receptor–area combinations in which potential risks are predicted on the basis of the conservative exposure and toxicity assumptions. Geographic information system mapping can be used to highlight such areas for further investigation.

Step 1 can incorporate increasingly more realistic models and site-specific data to provide a focus on those chemical–receptor–area combinations in which risks are not ruled out (Fairbrother 2003). For example, if a predicted HQ $>$ 1 was due largely to assumptions made regarding diet, then additional site-specific data can be collected on the diet composition of the receptor and chemical concentrations in diet items in the area under investigation. These site-specific data can replace generic, conservative assumptions used in earlier versions of the modeling, leading to a greater level of refinement in the food chain model. Levels of refinement can progress to probabilistic modeling or spatially explicit exposure modeling (applicable to both terrestrial and aquatic assessments). The decision to proceed to a greater level of refinement will be based on the expected contribution provided by the additional refinement to further rule out risks (Fairbrother 2003).

The validity of step 1 can be questioned if the receptor list is not regarded as a comprehensive enough representation of an ecosystem. In other words, what if direct toxicity to plants results in loss of habitat to the chosen wildlife receptors? The SALE process addresses this concern by explicitly including the potential for indirect effects on the chosen receptors because of effects on habitat, prey, or predators, for example. No risk assessment realistically can address all receptors that could potentially be affected by direct toxicity. For example, plants can be included as receptors but then what about soil microflora so important to the establishment of healthy soils for the plants? The SALE process allows a focus on fewer receptors for direct toxicity but then includes consideration of indirect effects.

The finding of no direct toxicity (e.g., from food-chain modeling) does not end the evaluation of risks to this receptor or other species for which the receptor was selected to represent. Rather, an evaluation of general wildlife or aquatic life use patterns in the study area is conducted (see detail under step 2). This includes consideration of potential indirect effects (e.g., via changes in habitat suitability) on the receptors evaluated via the modeling, as well as other similar species. For example, if all HQ $<$ 1 for black bear, then direct toxicity to black bear is no longer assessed. However, the black bear, other bear species, and other large mammals are still part of the consideration of community-level responses, such as wildlife use patterns and wildlife species diversity, particularly when habitat use and species diversity are restricted because of indirect effects on habitat suitability caused by chemicals of concern (COCs).

Step 1 offers the 1st opportunity to ask the question, "is sufficient information available to make a risk management decision?" If it is not possible to rule out all risks through the assessment of direct toxicity modeling, 3 options are available: A risk management decision is made, more advanced modeling is conducted (e.g., spatially explicit exposure models, probabilistic exposure–effects models), or 1 or more additional lines of evidence are added to the ERA for that receptor. This highlights the need for interaction between the risk assessor and risk manager because the risk manager will need to determine whether any remaining uncertainties prevent defensible decisions (i.e., whether or not enough information is available to make a decision). For example, if potential risks to receptors are ruled out for almost all of a study area and the remainder of the study area does not require further risk analysis to support common-sense risk management decisions, then no further risk assessment may be necessary. This might be the case at smaller sites. However, it is likely that further assessment will be required at larger sites.

Step 2—Assess indirect effects of changes in vegetation communities and habitat suitability

Cases in which receptors are affected not only by direct exposure to COCs but also via changes in habitat quality could occur. Habitat includes all of the environmental conditions present in the specific place occupied by an organism and is often defined to include a whole community of organisms. Thus, habitat quality can be affected by chemically induced changes in the structure of vegetation or animal communities, size of organisms, lifespan of organisms, productivity, production/biomass ratios, nutrient cycling, export of nutrients from the system, parasitism, or mutualism (Odum 1985). These effects can, in turn, produce so-called trophic cascade effects, which are chain reactions within food webs that result from changing population densities at higher trophic levels (AMNH 2002). For example, in 3-tiered food webs, an increase in the abundance of top predators can result in lower abundances of herbivore consumers and a higher abundance of plants. Habitat quality can also be changed by physical effects associated with chemical emissions (such as reduction to bare mineral soil caused by sulfur dioxide).

Furthermore, exposure of wildlife receptors to the COCs can vary according to habitat suitability; that is, animals will not use poor-quality habitat for feeding or nesting to the same extent that they use high-quality habitat. Therefore, in step 2 of the SALE process, effects of COCs on vegetation communities and other habitat features (such as soil development) and subsequent effects on habitat suitability for each wildlife or aquatic receptor are evaluated. The severity of the effect of a change in habitat will depend on the plasticity of the receptor species in terms of its requirements for food and shelter. In an aquatic ERA, step 2 would involve evaluation of effects on aquatic habitat quality via chemical effects on periphyton or macrophyte communities or physical effects such as changes in particle size distribution of sediments caused by effluent discharges.

Steps 2 and 3 may be carried out concurrently, with potential iteration back and forth between the 2 steps, particularly for landscape-scale ERAs. For example, field surveys can provide confirmation of vegetation community characteristics in relation to concentrations of chemicals in soil, thus providing supporting evidence for the relative importance of indirect effects on habitat quality.

Natural confounding variables distributed spatially within a study area, including terrain, soil, or sediment type, as well as other stressors that are spatially distributed, such as logging, linear developments, or dams, can be included in the analysis supporting step 2. Multivariate and univariate statistical analyses, along with spatial analysis with geographic information system, can assist in the understanding of the strength of association between natural and anthropogenic stressors and measures of receptor response. For example, an analysis could include the relative strength of association between terrain and the spatial distribution of plant community types versus the strength of association between soil metal concentrations and the distribution of plant community types.

The relative role of individual stressors is often difficult to distinguish because of co-occurrence spatially, temporally, or both and nonspecific responses. For example, the effect of sulfur dioxide emissions and soil metal concentrations on vegetation community characteristics can be very similar. The primary goal is to distinguish between the stressors related to the source under examination (e.g., a smelter or a pulp mill) and other confounding stressors.

Step 2 also includes consideration of the overall wildlife or aquatic community. Lines of evidence for the wildlife community can include wildlife habitat use patterns or community-level measures such as species richness. In aquatic ERA, the receptors are often entire communities (e.g., a benthic invertebrate community); therefore, a separate, community-level evaluation would not be required. However, consideration of the overall fish community might be included.

An evaluation of general wildlife use patterns in the study area might include consideration of potential indirect effects on wildlife via changes in habitat suitability caused by the chemicals of concern emitted by the source under investigation. This evaluation can be done for several broad categories of wildlife (e.g., songbirds, small mammals, ungulates). The evaluation of general wildlife use patterns is completed in a way similar to that for any individual receptor species through steps 2 through 6 of the SALE process. For some wildlife, habitat use in the study area could be observed in the field directly from monitoring programs or incidental observations during other activities. For others, the evaluation of use patterns could be completed primarily through inference on the basis of habitat suitability information (from vegetation mapping), supplemented by wildlife observations. With evidence that wildlife are affected by the source under investigation, the SALE process recommends that potential risk management options be evaluated. If not, then no further assessment is required for this general wildlife group.

Step 3—Measure effects on receptors

The scope of the studies in step 3 is based on the key sources of uncertainty in risk modeling of direct toxicity from step 1 and the specific chemical–receptor–area combinations for which risks were not ruled out. Laboratory or field studies that focus on the relationship between critical population or community-level indicators and stressor concentrations will add confidence to the assessment. For example, if the toxicity data used to derive the benchmark or TRV were based on effects to a dissimilar species, or if few dose levels were tested, or if endpoints used to derive the TRV were not relevant to persistence of receptor populations, then field population surveys might provide a better estimate of population-level effects. In aquatic ERA, step 3 often includes standardized whole effluent toxicity tests, toxicity tests with ambient media, field effects monitoring methods developed for regulation of effluent discharges, or a combination of procedures.

The type of field survey, laboratory study, or mesocosm study could vary for each receptor and might only be necessary for a small number of receptors in an ERA. Potential lines of evidence under this step can include field studies of wildlife community characteristics, such as species diversity; field studies of population responses along gradients or comparisons between upgradient and downgradient locations (e.g., relative abundance); productivity studies; evidence for stressor-related symptoms of disease; laboratory toxicity tests; mesocosm experiments; or recruitment, mortality, and emigration or immigration studies for wildlife or fish populations.

Step 4—Evaluate magnitude of response

Step 4 in the SALE process is used to classify the magnitude of the measured response to stressors in terms of ecological relevance. The emphasis on ecological relevance ensures that statistically significant differences per se are not necessarily interpreted as ecologically significant. The magnitude rankings, together with the results of all previous steps, support the detailed causal analysis in step 5.

Three categories for magnitude of response are defined (Figure 1): Strong, moderate, and weak or inconclusive. Three classes were found to be numerous enough, yet not too numerous, to allow the characterization of risks that lead to a useful risk management evaluation. If the authors of a risk assessment wanted to expand this to 4 or more, or decrease it to 2 categories, the process would still be the same. The primary requirement is a clear definition of each category.

The definitions of categories of responses will differ depending on the receptor being evaluated and the line of evidence (e.g., toxicity test vs population or community survey measures). In some cases, quantitative definitions will be possible. In other cases, qualitative definitions will be necessary because of the nature of the evidence or the lack of guidance or knowledge.

The criteria for judging whether a response is strong, moderate, or weak could be part of regulatory guidance, such as Environment Canada's (2004) guidance for interpreting monitoring measures in the national aquatic environmental effects monitoring programs. For example, strong, moderate, and weak responses in toxicity tests have been defined as reduction of more than 50% in 1 or more toxicological endpoints, reduction of more than 20% but less than 50% in 1 or more toxicological endpoints, and reduction of 20% or less in each toxicological endpoint, respectively (Environment Canada 2004). Another example of potentially applicable guidance is from the Oregon Department of Environmental Quality (ODEQ 2000), which has suggested that the median lethal dose or concentration (LD50 or LC50) be used as a population-level ecological benchmark value for risk assessment. If 20% of the individuals in a defined local population of size n has a $<$10% probability (p) of an exposure $>$ecological benchmark value, and no other observed significant adverse effects on the health or viability of the local population are identified, then risk would be defined as acceptable (ODEQ 2000). This could be a starting point for defining response categories. Interestingly, the ODEQ guidance includes a provision for assembling a weight-of-evidence risk characterization but provides no methodology, nor does it provide anything beyond the ecological benchmark value for interpretation of field observations of populations or communities. This illustrates the general lack of guidance at the field level, particularly for terrestrial receptors.

Field-based quantitative lines of evidence can be assessed against formal definitions of critical effect size, such as those defined in Canadian aquatic environmental effects monitoring programs. Methods for determination of critical effect sizes are still a subject of considerable scientific discussion; however, a common starting point is the normal range of variability seen in reference communities (WADOE 1991; Elliot 1996; Lowell 1997). This is often defined as 2 standard deviations from the reference mean (Environment Canada 2004). The ecological relevance of this type of statistical definition has not been defined as clearly for the terrestrial environment. Another benchmark is a 20% change in abundance or production of an endpoint population (Suter et al. 1995). Both the 20% benchmark and the use of critical effect size assume some ability to distinguish at-risk populations or communities according to measured differences compared with some sort of reference condition. The validity of the use of reference sites can be disputed (Landis 2002).

When scoring the magnitude of response, it also is necessary to consider the types and levels of uncertainty in measurements or studies used to evaluate receptor response. For example, a moderate response could be observed in a field measure, but with considerable uncertainty around this measure because of numerous confounding variables, poor statistical power, or the selection of inappropriate or inadequate measures. The level of uncertainty must be transparent and noted qualitatively or quantitatively in some way. If the line of evidence is quantitative, statistical criteria for definition of significant differences (α) and statistical power (β) can be incorporated into the magnitude criteria. For example, weak evidence would not meet the minimum statistical power requirements. Uncertainty in laboratory-based lines of evidence can be assessed via confidence limits for toxicity endpoints. Uncertainty in qualitative lines of evidence can be scored according to criteria such as concurrence with magnitude of response in similar cases. For example, the degree of chlorosis and bronzing of the undersides of leaves caused by sulfur (Griffiths 2003) is a qualitative measure with uncertainty that can be described by a narrative that explains how common and consistent this pattern of response is.

Quantitative, qualitative, or both types of criteria used to determine magnitude of response and the uncertainty inherent in the measures of response must be clearly explained and presented. Some published benchmarks are available for categorizing response (e.g., Suter et al. 1995; ODEQ 2000); however, the final choice of criteria for determination of weak, moderate, or strong magnitude of response must be applicable to the specific receptors, chemicals, and study area.

The assessment proceeds to step 5 (causality assessment) unless all lines of evidence for a receptor show a weak or inconclusive response (i.e., evidence is insufficient for a significant biological response to the chemicals of concern). If the assessment does show a weak or inconclusive response for all lines of evidence, then the risk assessment for the receptor ends. Consideration could be given to monitoring the receptor–area combination if uncertainty around the stressor–response relationship is high. Care also must be taken to ensure that no potential for distance effects is suspected, in which populations distant from the immediate spatial area affected by a chemical can be affected via the formation of population sinks (Landis 2002). Population sinks located in the immediate study area might not show significant responses or correlation with chemical concentrations because abundance is constantly being maintained by immigration from distant populations; however, the distant populations might decline. Thus, it is necessary to carefully define the study area for the ERA to account for this possibility and consider implications for both local and regional populations, if appropriate.

Step 5—Assess causality

The SALE process continues with an evaluation of causality for all lines of evidence (regardless of the magnitude of response, unless all responses are weak or inconclusive). Causality is evaluated with the use of a formal set of criteria based on Hill (1966), Fox (1991), Suter (1993), Beyers (1998), Culp et al. (2000), Lowell et al. (2000), and USEPA (2000b). These causal criteria assume a proportional response between exposure and effects. Landis (2002) cautions that we should not expect proportionality (i.e., clear and consistent dose–response relationships) because components of the ecosystem are linked and could be affected by changes in other ecosystem components. However, many authors have succeeded in illustrating a relationship between population- or community-level responses and the exposure to stressors (albeit primarily in aquatic ecosystems). The SALE process includes the following causal criteria.

Spatial correlation—Effects occur at the same place as exposure; effects do not occur with no exposure. In a river, effects can occur downstream of a source, but not upstream.

Temporal correlation—Effects occur with or after exposure.

Biological gradient or strength—Effects decline as exposure declines in the landscape. Similarly, effects decline as exposure declines over time (or effects increase as exposure increases over time). Evidence for cause and effect is stronger if the exposure response is monotonic with relatively high regression coefficients ($>$0.5).

Plausibility (mechanism)—It must be known how the stressor causes an effect in the affected organisms. This will

determine whether it is plausible that the observed effects are a result of the stressor. Consideration must be given to indirect mechanisms (e.g., increased nutrient levels in water cause algal blooms, which decrease oxygen levels in water, which could decrease invertebrate density).

Plausibility (stressor–response)—The magnitude of effect is expected on the basis of the level of the stressor.

Consistency of stressor–effect association—Repeated observation of effect and stressor is seen in different studies or different locations within the region being studied. In addition, information is available from other regions in which similar (analogous) stressors have caused similar effects.

Experimental verification—Effects of the stressor are observed under controlled conditions with concordance of these experimental results with field data.

Specificity of cause—The effect tends to be associated with exposure to a particular stressor. Effects should be defined as specifically as possible to increase the specificity of the association between cause and effect. In the extreme case, causation is clear when a stressor results in only 1 effect, and that effect is only related to that 1 stressor. Of course, this is rare in environmental situations.

The above list of causal criteria varies somewhat from that found in USEPA (2000b) and other sources. For example, the criterion for evidence of complete exposure pathway was removed from consideration because the causal analysis is completed only for those areas in which modeling was unable to rule out risks. Therefore, it is known that evidence of a complete exposure pathway exists. Another criterion, coherence with analogous cases, is difficult, as can be seen in the examples presented by the USEPA (2000b), for which it was not used. This criterion refers to evidence in the literature for the observation of similar effects related to exposure to a chemical that is similar to the chemical under investigation, as opposed to the same chemical. Therefore, this criterion also was not used.

At least 1 of the first 2 causal criteria (spatial or temporal correlation) is considered necessary to make a case for causality, and evidence for causality is strengthened by meeting both. That is, some spatial or temporal correlation must exist between exposure and effects. It might not be possible to meet both causal criteria because data might be inadequate to show that the stressor and response have covaried over time; or, the data might not be of sufficient quality or quantity to make definitive statements about spatial correlation (e.g., particularly when relying on statistical or geospatial techniques, such as kriging, for estimating chemical concentrations and distributions over a landscape). However, simply having this correlation in space or time is insufficient to make a strong case for causality, especially for large study areas and in situations that have experienced long-term contamination and potential confounding effects (e.g., numerous other stressors or factors also could be the cause of the observed effect).

A line of evidence should not be considered inadequate if it is not supported by all causal criteria. In fact, 2 of the causal criteria, namely the specificity of cause and experimental verification, might not be supported by lines of evidence for particular receptors. For example, endpoints of interest for wildlife receptors often are too general (e.g., abundance, population persistence) to be supported by the specificity of cause criterion. Experimental verification is relatively simple for aquatic receptors for which standard laboratory and field tests are available and of relatively short duration. However, this is complicated for avian and mammalian receptors because testing and monitoring would require significant resources (e.g., time, money), and it might be difficult to detect a change independent of other variables (e.g., weather, predation levels, etc.) when these tests are conducted in the field. Experimental verification would only be called for if a risk management decision could not be made without it. For example, 2 choices could be possible for risk management, one involving large-scale remediation of significant portions of the study area and the other involving only small, specific portions of the study area. If the uncertainty around which choice is correct was very high because of a lack of convincing cause–effect information, then the value of experimental verification data might be commensurate with the risk management decision.

The results of the examination for causality among the laboratory, mesocosm, and field lines of evidence can be summarized by applying scores that reflect the performance of each line of evidence against the causal criteria. An example of a scoring scheme is presented in Table 1, adapted from the USEPA (2000b).

The score options in Table 1 give an indication of how much weight each causal criterion can inherently assume in the overall WOE analysis. If spatial or temporal correlation receives a very low (− − −) score with confidence that distance effects are not occurring (as noted in Step 4), then this provides strong support for ruling out risk from the chemicals of concern for the particular receptor population or community. For example, if failure of recruitment of juvenile fish started to occur 50 y after a discharge of metals to a river system, but only 2 y after a dam was installed upstream, and the discharge of metals had declined by 2 orders of magnitude over time with almost no accumulation of metals in the food chain, the lack of temporal correlation with metals discharges can be taken as strong evidence to rule out metals as the cause of recruitment failure. Other causation criteria also have high weight. For example, the plausible mechanism criterion has the scores of + +, +, 0, and − − −. This indicates that if it is reasonable to assume that the COCs could cause the observed effect in the receptor, then a small amount of weight (+) is assigned to this causal criterion. However, if it is completely implausible that the COCs could act on the receptor in this way, then confidence is greater (− − −) of no link between the effects on the receptor and the COCs.

Unlike the magnitude of response step, it is not necessary to quantitatively evaluate uncertainty in the causal analysis. In essence, the uncertainty is inherent in the score. A + + + or − − − has little uncertainty around it (analogous to small error bars), whereas a 0 score has much more uncertainty (large error bars) and a single + or − would have an intermediate level of uncertainty.

The final stage of step 5 involves developing a summary of the causality information produced with the use of the criteria and scoring in Table 1. Table 2 can be used to present the results of scoring toxicity and field-based data against each causal criterion. It is then possible to credibly assign overall causation scores to each line of evidence. These overall causation scores will be identified in the Overall strength of evidence row (the final row in Table 2) and can be identified as strong, moderate, or weak or inconclusive (or representative symbols thereof). The USEPA (2000b) cautions against simply adding up the scores for each line of evidence, because

Table 1. Example of a scoring scheme for causal criteria (adapted from USEPA 2000b)

Criterion	Results	Score[a]
Spatial correlation	Strong evidence; compatible; uncertain; incompatible	++; +; 0; − − −
Temporal correlation	Strong evidence; compatible; uncertain; incompatible	++; +; 0; − − −
Biological gradient/strength	Strong and monotonic; weak or other than monotonic; none; clear association, but the more stressor, the lower the response	+++; +; −; − − −
Plausible mechanism	Actual evidence; plausible; not known; implausible	++; +; 0; − − −
Plausible stressor response	Quantitatively consistent; concordant; ambiguous; inconcordant	+++; +; 0; − − −
Consistency of association (across sites in the region)	Invariant; in many places and times; at background frequencies or many exceptions to the association	++; +; −
Experimental verification	Experimental studies: concordant; ambiguous; inconcordant	+++; 0; − − −
Specificity of cause	Only possible cause; one of a few; one of many	+++; ++; 0

[a] In addition to the scores noted, no evidence (NE) might be available relevant to the criterion or the criterion might be not applicable (NA) for the particular case.

it implies incorrectly that each causal criterion is of equal importance, but rather suggests more attention be paid to negative results, which are more likely to be decisive.

The overall causation score is determined by how consistent and strong the causation evidence is across all causal criteria. In essence, the overall causation score is a measure of the completeness of the causal pathway, which could be considered the link from stressor to response via a plausible mechanism, the plausible stressor response, showing spatial and temporal correlation, for example.

With strong or moderate evidence for causality related to COCs, the SALE process continues to step 6. With only weak evidence for each line of evidence, especially if evidence is strong for causality attributed to a different cause, the risk assessment ends (for the cause under evaluation). Monitoring the receptor or area might be appropriate.

Table 2. Scoring lines of evidence (columns) according to each causal criterion (rows)

	Line of evidence				
	Toxicity in laboratory tests		Changes in community from field measurements		Indirect effects on habitat caused by COCs
Causal criterion	Measure 1	Measure 2	Measure 1	Measure 2	
Spatial correlation					
Temporal correlation					
Biological gradient/strength					
Plausibility: mechanism					
Plausibility: stressor response					
Consistency of association					
Experimental verification					
Specificity of cause					
Overall strength of evidence					

Step 6—Evaluation of overall WOE

The final weighing of evidence involves summarizing causality scores and magnitude of response scores (factoring in the level of uncertainty) for each line of evidence of direct and indirect effect and for each site or area under evaluation. Consistency of scores across all lines of evidence provides a higher level of confidence in the recommendation to proceed or not to a risk management evaluation. A number of high uncertainty scores for a site or area would indicate that a risk management decision might benefit from more information.

Table 3 shows an example of the integration of magnitude of response, uncertainty, and causality information to produce an overall recommendation about whether or not to proceed to risk management.

Significant narrative is needed to accompany all of the tables in the SALE process, especially the final table in which the recommendation is made regarding whether or not to proceed to the consideration of risk management measures. With a moderate or strong magnitude of response (and depending on the level of uncertainty), coupled with moderate or strong support for causality, the recommendation is to proceed to an evaluation of risk management options. With weak responses and weak or inconclusive support for the source under investigation as the cause of the response, then the recommendation could be to conclude the risk assessment for this receptor. Where uncertainty is high and causation, magnitude, or both scores are inconsistent, the recommendation to proceed to risk management is also uncertain and could be flagged for further discussion or additional data gathering (e.g., site Z in Table 3).

CONCLUSION

The SALE approach recognizes that direct toxicity modeling (or a comparison of site media concentrations to criteria or benchmarks) is more effective in ruling out risk than in providing an estimation of risk. Thus, model-based lines of evidence are 1st in a sequence of steps that could lead to a recommendation to proceed to a risk management evaluation. Laboratory and field-based lines of evidence compose the remainder of the sequence, 1st to evaluate the magnitude of response and then, if the magnitude is sufficient, to evaluate causation. A recommendation to proceed to a risk management evaluation is made only when the sequential analysis has shown sufficient magnitude and causation.

The SALE approach includes consideration of indirect effects on habitat suitability as well as direct toxicity. This is an important advancement, because indirect effects are often greater than direct toxicity effects.

The SALE process makes the risk characterization transparent and facilitates the integration of ERA results into a risk management framework. It illustrates to risk managers how the decision to proceed to a risk management evaluation can go beyond the simple derivation of risk-based concentrations of chemicals of concern. Therefore, risk management goals could be described via ecological metrics (e.g., species diversity) and not as a chemical concentration (e.g., 500 ppm chemical in soil). The SALE approach can help create greater understanding between assessors and managers by making the process of using several lines of evidence more clear. It can be used to stimulate discussion of the limitations of the ERA science, and how scientists deal with uncertainty.

Table 3. Summary weight of evidence table with decision as to whether or not to proceed to a risk management evaluation[a]

Site	Toxicity bioassay						Population/community						Proceed to risk management evaluation (yes/no)[b]
	Test endpoint 1			Test endpoint 2			Measure 1			Measure 2			
	Magnitude	Uncertainty	Causation	Magnitude	Uncertainty	Causation	Magnitude	Uncertainty	Causation	Magnitude	Uncertainty	Causation	
W	○	○	○	○	○	○	○	○	○	⊙	⊙	⊙	No
X	⊙	○	●	●	○	●	●	○	●	○	⊙	●	Yes
Y	○	○	○	○	○	○	○	○	○	⊙	○	⊙	No
Z	○	⊙	⊙	○	⊙	○	○	●	○	⊙	⊙	⊙	No[c]

[a] Magnitude refers to the magnitude of response (● = strong; ⊙ = moderate; ○ = weak or inconclusive); uncertainty is related to the magnitude of response category (○ = little uncertainty; ⊙ = moderate level of uncertainty; ● = substantial uncertainty); and causation refers to the overall strength of evidence from the final row of Table 2 (● = strong; ⊙ = moderate; ○ = weak/inconclusive).

[b] The decision to proceed to risk management evaluation is either yes or no.

[c] Uncertain decision that requires further discussion or data gathering.

Acknowledgment—The authors thank Anne Fairbrother, Peter Chapman, and 3 anonymous reviewers for their helpful comments. In addition, the authors thank Steve Hilts and Bill Duncan of Teck Cominco Metals Ltd, Wayne Landis of Western Washington University, and the entire Teck Cominco Trail ERA team for many stimulating discussions on this topic. This work was funded by Teck Cominco Metals Ltd (Trail, BC, Canada, operations).

REFERENCES

Adams SM. 2003. Establishing causality between environmental stressors and effects on aquatic ecosystems. *Human Ecological Risk Assessment* 9:17–35.

[AMNH] American Museum of Natural History. 2002. Glossary compiled by American Museum of Natural History Center for Biodiversity and Conservation in support of Symposium on the Science and Policy of Marine Resource Management. March, 2002. http://research.amnh.org/biodiversity/symposia/archives/seascapes/glossary.html. Accessed 15 August 2005.

Beyers DW. 1998. Causal inference in environmental impact studies. *J N Am Benthol Soc* 17:367–373.

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

Burton GA Jr, Chapman PM, Smith EP. 2002. Weight-of-evidence approaches for assessing ecosystem impairment. *Human Ecological Risk Assessment* 8:1657–1673.

Chapman PM. 1990. The sediment quality triad approach to determining pollution-induced degradation. *Sci Total Environ* 97/98:815–825.

Culp JM, Lowell RB, Cash KJ. 2000. Integrating mesocosm experiments with field and laboratory studies to generate weight-of-evidence risk assessments for large rivers. *Environ Toxicol Chem* 19:1167–1173.

Elliot M. 1996. The derivation and value of ecological quality standards and objectives. *Mar Pollut Bull* 32:762–763.

Environment Canada. 2004. Pulp and paper technical guidance document for aquatic environmental effects monitoring. Ottawa (ON): Environment Canada. Draft, May 2004.

Fairbrother A. 2003. Lines of evidence in wildlife risk assessments. *Human Ecological Risk Assessment* 9:1475–1491.

Fox GA. 1991. Practical causal inference for ecoepidemiologists. *J Toxicol Environ Health* 33:359–379.

Griffiths, H. 2003. Effects of air pollution on agricultural crops. Toronto (ON): Ontario Ministry of Agriculture, Food, and Rural Affairs. Factsheet 85-002.

Hill, BA. 1966. The environment and disease: Association or causation? *Proc R Acad Med* 58:295.

Landis WG. 2002. Uncertainty in the extrapolation from individual effects to impacts upon landscapes. *Human Ecological Risk Assessment* 8:193–204.

Landis WG, Markiewicz AJ, Matthews RA, Matthews GB. 2000. A test of the community conditioning hypothesis: Persistence of effects in model ecological structures dosed with jet fuel JP-8. *Environ Toxicol Chem* 19:327–336.

Landis WG, Wiegers JA. 1997. Design considerations and a suggested approach for regional and comparative ecological risk assessment. *Human Ecological Risk Assessment* 3:287–297.

Lowell RB. 1997. Discussion paper on critical effect size guidelines for EEM using benthic invertebrate communities. Report to the Environmental Effects Monitoring Program. EEM/1997/8. Ottawa (ON): Environment Canada.

Lowell RB, Culp JM, Dube MG. 2000. A weight-of-evidence approach for northern river risk assessment: Integrating the effects of multiple stressors. *Environ Toxicol Chem* 19:1182–1190.

Matthews RA, Landis WG, Matthews GB. 1996. The community conditioning hypothesis and its application to environmental toxicology. *Environ Toxicol Chem* 15:597–603.

Menzie CA, Henning MH, Cura J, Finkelstein K, Gentile J, Maughan J, Mitchell D, Petron S, Potocki B, Svirsky S, Tyler P. 1996. Special report of the Massachusetts weight-of-evidence workgroup: A weight-of-evidence approach for evaluating ecological risks. *Human Ecological Risk Assessment* 2:277–304.

[ODEQ] Oregon Department of Environmental Quality. 2000. Guidance for ecological risk assessment: Levels I, II, III, IV. Portland (OR): ODEQ, Water Management and Cleanup Division.

Odum EP. 1985. Trends expected in stressed ecosystems. *Bioscience* 35:419–422.

Sandberg RS, Landis WG. 2001. Persistence of the effects of Jet-A in a microcosm with releases from the sediment. *Environ Toxicol Chem* 20:1942–1950.

Suter GW. II. 1993. Ecological risk assessment. Boca Raton (FL): Lewis. 538 p.

Suter GW. II. 1996. Risk characterization for ecological risk assessment of contaminated sites. Oak Ridge (TN): Oak Ridge National Laboratory. ES/ER/TM-200.

Suter II GW, Cornaby BW, Hadden CT, Hull RN, Stack ME, Zafran FA. 1995. An approach for balancing health and ecological risks at hazardous waste sites. *Risk Anal* 15:221–231.

[USEPA] US Environmental Protection Agency. 1998. Guidelines for ecological risk assessment. Washington DC: Risk Assessment Forum. April 1998. Final. EPA/630/R-95/002F.

[USEPA] US Environmental Protection Agency. 1999. Issuance of final guidance: Ecological risk assessment and risk management principles for superfund sites. Washington DC: Office of Solid Waste and Emergency Response. October 7, 1999. OSWER Directive 9285.7–28P.

[USEPA] US Environmental Protection Agency. 2000a. Guidance for the data quality objectives process. EPA QA/G-4. August 2000. Washington DC: Office of Environmental Information. EPA/600/R-96/055.

[USEPA] US Environmental Protection Agency. 2000b. Stressor identification guidance document. December 2000. Washington DC: Office of Water, Office of Research and Development. EPA/822/B-00/025.

[WADOE] Washington State Department of Ecology. 1991. Sediment management standards. Olympia (WA): Washington State Department of Ecology. 173-204 WAC.