FORUM

DOI: 10.1007/s00267-001-2447-z

Quantifying Natural Resource Injuries and Ecological Service Reductions: Challenges and Opportunities

LAWRENCE W. BARNTHOUSE*
LWB Environmental Services, Inc.
105 Wesley Lane
Oak Ridge, TN 37830, USA

RALPH G. STAHL, JR.
DuPont Engineering
Corporate Remediation Group
Barley Mill Plaza # 27
Route 141 and Lancaster Pike
Wilmington, DE 19805, USA

ABSTRACT / The natural resource damage assessment (NRDA) provisions of the Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA) and the Oil Pollution Act (OPA) are complex and have been difficult to implement. The complexity and difficulty in implementation arise both from the assessment procedures specified in agency NRDA guidance and from the limited ability of ecologists to quantify impacts of hazardous substances on natural resources. This paper explores the scientific aspects of NRDA implementation, and discusses conceptual and methodological relationships between NRDA and the much broader field of ecological risk assessment (ERA). We discuss three critical components of the NRDA assessment approach: measuring natural resource injuries and reductions in resource services; evaluating causality; and establishing baseline conditions. We identify (1) specific approaches drawn from ERA practice that could improve each of these components, and (2) research needs and institutional changes that may improve the ability of the NRDA process to achieve its stated objectives. We recommend the acceleration of the ongoing dialogue among NRDA practitioners from the Trustee and PRP communities as a first step toward resolving the procedural and technical deficiencies of the NRDA process.

The natural resource damage assessment (NRDA) provisions of the Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA) and the Oil Pollution Act (OPA) provide a mechanism for restoring natural resources that have been adversely affected, or "injured," by unpermitted releases of oil or hazardous substances (West Publishing Co. 1991). In enacting this legislation, Congress believed that these resources provide valuable services to society in the form of pollution abatement, recreation, and aesthetics that are difficult to measure and may not be sufficiently considered when CERCLA remedial actions are selected (White 1990). In general, the NRDA process is an attempt to make the public "whole" for the loss of those ecological services stemming from exposure to oil or hazardous substances, before, during and after the CERCLA remediation (Sanford et al. 1998, White 1990). Developing the sound scientific bases for assessing these adverse impacts or injuries, together with the associated service loss, if any, is the challenge facing scientists, legal professionals and policy makers alike.

KEY WORDS: Natural resource damages; Injury determination; Injury quantification; Ecological risk assessment

*Author to whom correspondence should be addressed; *email:* lwb.env@worldnet.att.net

CERCLA, OPA, and other federal statutes empower federal, state, and tribal agencies (referred to as Trustees) to pursue damage claims against responsible parties for injury to, destruction of or loss of natural resources from exposure to unpermitted releases of hazardous substances and oil. The three main steps of the NRDA process—assessing and quantifying injury, determining damages, and implementing restoration—are complex and can be difficult to implement. Trustees must show a clear link between observed biological, chemical or physical effects and the exposure to a particular hazardous substance or oil, and then link that exposure to an unpermitted release from a particular responsible party or parties. Trustees must document and quantify natural resource service loss, which is not the same as documenting or measuring injury (although the two can be sometimes similar). The need to determine causality, measure injury and determine the baseline condition is presupposed within these three major steps.

The closely related, but more fully developed, process of ecological risk assessment (ERA) provides an important source of methods for improving the rigor

Table 1. Similarity of actions in ecological risk assessments and natural resource damage assessments (modified from Stahl et al. 1995)

Action	Ecological Risk Assessment	Natural Resource Damage Assessment
Step 1	Problem Formulation Phase: reviewing the issues, developing site conceptual model, determining goals.	Pre-Assessment Phase: conducting a field examination, reviewing data, examining the potential for a claim.
Step 2	Analysis Phase: ecological effects and exposure characterization, data collection and evaluation.	Assessment Phase: chemical and biological data collection, data review and evaluation.
Step 3	Risk Characterization Phase: analysis of data, integration of hazard and exposure data, estimating risk.	Injury Determination Phase: data collection and analysis.
Step 4		Injury Quantification Phase: estimating amount of injury and consequent reductions in services provided by the resource.
Step 5	Risk Management Phase: decisions on risk reduction options.	Damage Assessment Phase: establishing financial claim and/or plans and costs for restoration options.

and clarity of the NRDA process. ERAs are now routinely performed as part of the implementation of CERCLA, OPA, the Clean Water Act, and most other environmental statutes (USEPA 1997, 1998). Simply put, ERA is a process that evaluates the likelihood that adverse ecological effects may occur or are occurring as a result of exposure to one or more stressors (USEPA 1992a). These stressors may be biological, physical or chemical (hazardous substances).

The objective of an ERA performed under CERCLA is to determine whether hazardous chemicals in environmental media may have harmed or (in NRDA terminology) injured exposed organisms. In an NRDA, the Trustees are required to prove that an injury has occurred, that this injury led to a loss of ecological services and that the injury was caused by the unpermitted release of a specific hazardous substance or oil, by a specific responsible party. The Trustees must then quantify the extent of the service loss and the amount of monetary compensation required to make the public "whole." The generally parallel structures of the NRDA and ERA processes and the frequent use of the same data for two distinct purposes suggests a strong intertwining of the ERA and NRDA processes (Table 1) (Stahl et al. 1995). Yet the outcome of an ERA is not the same as the outcome of an NRDA even though similar data may be collected for both. The output of an ERA is an estimate of risk and (if necessary) a remedial action goal, while the outcome of the NRDA is the quantification of economic damages or restoration requirements for that injury. And, just as important, the requirements placed on natural resource Trustees in NRDA actions are often far more demanding than the requirements placed on USEPA and potentially responsible parties (PRPs) in CERCLA actions.

Although there are fundamental differences between ERA and NRDA, it must be recognized that ERA has been much more widely applied and has had the benefit of substantial peer review and revision, both from the regulatory and regulated communities. A number of professional societies (Society of Environmental Toxicology and Chemistry [SETAC], American Society for Testing of Materials [ASTM] and Society of Risk Analysis [SRA]), are openly involved in the debate and resolution of scientific issues common to environmental toxicology, ERA and ecological risk management (Stahl et al. 2001). The ERA process is widely accepted in the chemical industry (CMA 1997) and in regulatory agencies (USEPA 1998). Continuing refinements of the process are supported by several decades of active application in water pollution control, chemical and pesticide registration, and environmental remediation. The winnowing of unproven or unnecessary approaches from the ERA tool box over the years has been a key to its increased application in risk-based decision making.

This paper explores the scientific aspects of NRDA implementation, emphasizing the conceptual and methodological relationships between NRDA and the much broader field of ERA. The purpose of this exploration is to identify ways in which conceptual and methodological advances in ERA can help to resolve some of the difficult technical issues facing NRDA practitioners.

Linkages Between NRDA and ERA

As shown in Table 1, there are clear parallels between the two processes. ERA is not a specific collection of experimental methods, databases, or models, nor is its relevance limited to a narrow universe of regulatory applications. It is a systematic approach to organizing and analyzing data, information, assumptions, and uncertainties relating to the effects of human activities on the nonhuman environment. Although the objectives and standards of proof in NRDA differ from those employed in ERAs performed for the EPA Superfund Program and other similar regulatory programs, the common science-driven processes of data collection and evaluation provide numerous opportunities for cross-fertilization.

Although NRDA is not specifically discussed in the USEPA's Guidelines (USEPA 1998), it is discussed by USEPA in the context of CERCLA activities (USEPA 1992d, Fields 1997). Components of the NRDA process that deal with quantifying injuries and service losses clearly fall within the scope of the USEPA guidelines. Prior to initiation of a formal assessment and damage claim, natural resource Trustees must perform a Pre-assessment Screen that is a desktop review of existing information and a determination of whether or not a damage claim should be filed. Much of this step is not unlike Problem Formulation in an ERA. The Assessment Plan together with the Pre-assessment Screen closely resemble Problem Formulation in an ERA: both processes define ecological receptors at risk ("trust resources" in DOI terminology and "assessment endpoints" in EPA terminology). Both processes evaluate existing data and describe putative causal relationships between chemical releases and adverse environmental changes. Results of both are summarized in an Assessment Plan that provides a concrete guide to data collection and evaluation.

Injury determination in NRDA is roughly equivalent to exposure and effects assessment in the ERA Analysis Phase. Often the chemical and biological data used in NRDA injury determination are likely to include the same data collected for the ERA under CERCLA. The Quantification Phase of NRDA, which integrates measures of exposures and effects into units of service loss, is used to support restoration decisions and fits easily within the definition of Risk Characterization found in ERA.

Three components of NRDA seem especially appropriate for detailed evaluation: (1) the measurement of injuries, (2) the determination of causality, and (3) the establishment of baseline conditions.

Measuring Natural Resource Injuries

The terms "injury," "service loss," and "damage" are found in the NRDA lexicon, but they are not scientific terms. The unique terminology employed in NRDA regulations is a significant source of confusion and obscures the relationship of NRDA to other types of ecological assessments.

The regulations issued by the National Oceanic and Atmospheric Administration (NOAA) and the U.S. Dept. of Interior (DOI) have adopted different definitions of natural resource "injury." NOAA (1996) defines "injury" as "an observable or measurable adverse change in a natural resource or impairment of a natural resource service," while DOI (1995) defines "injury" as "measurable adverse change, either long or short-term, in the chemical or physical quality or viability of a natural resource resulting either directly or indirectly from exposure" to the released substance. Under the DOI regulations, specific criteria are provided for determining whether injury has occurred to water, geologic resources (e.g. soil), air, and biota, and the impact of such "injuries" on the natural resource services is addressed in a separate phase of the assessment.

For NRDA purposes, the appropriate definition of "injury" is a measurable adverse change in a resource such that the resource does not provide the same services as it would have in the absence of the unpermitted release of oil or a hazardous substance. The reason for this is that the public's value for a natural resource stems from the services provided by the resource, and hence that value is diminished only by a loss or reduction in such services. A "service" loss or reduction refers to the lost or reduced opportunity such as for fishing, nature viewing, hunting, or natural water treatment (removal of suspended solids by wetlands for example) due to the injury to the resource, or basic life support. In other words, it is not the resource itself, but the services it would have provided in the absence of injury that form the basis for NRD. The amount of damages awarded should thus be commensurate with the value of the services lost or impaired.

The distinction between injuries and service losses is fundamental to understanding the scientific problems encountered in implementing NRDA regulations. Injuries to individual organisms may be relatively easy to document, but are generally not as relevant ecologically as injuries sustained at the population level and above and thus generally do not affect the services provided by the resource. In most cases services are provided by populations, communities, or ecosystems, not by individual organisms. Thus, the precision achieved by measuring injuries to individuals comes at the expense of evaluating the impacts on services, which is the proper focus of NRDA.

NOAA (1996) and DOI (1995) have adopted somewhat different methods for assessing injuries and damages. The DOI "Type B" regulations, which provide methods for quantifying injuries and service losses using site-specific data, prescribe methods that are in many ways similar to methods used in site-specific ERAs. However, the DOI methodology for determining injury emphasizes documentation of effects on individuals and does not directly address service losses. As discussed more fully below, in the absence of methods that directly address population and community effects, Trustees often inappropriately equate individual-level effects with service losses. This occurs despite the provisions of the DOI regulations indicating the need to focus on population-level effects in determining biological injury. In contrast, the connection between effects on individuals and effects on populations and communities has been one of the central issues in the development of ERA methodologies.

Failure to Distinguish Between Injury Definition and Injury Quantification

Both NOAA and DOI have provided methods for the measurement of injuries to natural resources. On the one hand, NOAA permits a range of injury assessment procedures, including field or laboratory, or model or literature-based, or some combination of these. Criteria for selecting the most reliable, cost-effective, and useful procedures for a given injury assessment are not provided. The assessments can be open-ended, with few restrictions placed on the scope and rigor of the overall process. The DOI regulations, in contrast, provide specific definitions and measurement methods for determining whether injuries to surface water, groundwater, air, geologic resources (e.g., soil), or biological resources may have occurred. In addition, there are specific criteria for biological injury which must be met before the results from these measurements are deemed acceptable to demonstrate injury. Injury determination, according to DOI, must be based upon the establishment of a statistically significant difference in the biological response between samples from populations in the assessment area compared to those in the control area. A demonstration that such responses are statistically more frequent in an area contaminated by hazardous substances than in an uncontaminated reference area is, according to the DOI regulations, sufficient to demonstrate that an injury has occurred. However, differences of this type are indicators of potential effects on populations and ecosystems, not measures of actual reductions in the viability of populations or the services they provide. An incidence rate of only a few percent in an exposed population, which would probably be biologically insignificant (more substantial percentage reductions are often encountered in the environment), might satisfy the DOI acceptance criteria.

Finding an appropriate balance between (1) simple assessment procedures that can be clearly defined and widely applied, but can lead to simplistic (and wrong) answers and (2) complex assessment procedures that can provide more accurate results but can also be excessively costly or impossible to implement has been a major issue in ERA. One approach that USEPA has taken to reconciling the conflict between the need for simplicity and the need for rigor is to introduce the concept of hypothesis testing into the CERCLA ERA process (USEPA 1997). Hypothesis testing is one of several approaches that allows scientists to agree upon a precise statement of the "problem" (e.g. hypothesis: the reduction in the number of finfish in My Pond is due to mirex contamination of the pond's sediment above 10 mg/kg) and to clearly identify the information needed to test the hypothesis. Currently there appears to be no analogy to hypothesis testing, or other approach, built into the NRDA process, that provides for an appropriate balance between simplicity and rigor.

Misapplication of Threshold or Criteria Values

Some of the indicators used by Trustees are not direct measures of injury to resources. in In a large fraction of past and present NRDA cases, the potentially injured resources include invertebrates exposed to hazardous substances in water and sediment. In such cases, Trustees have used exceedences of water and sediment-quality criteria as evidence of injury (Weiss et al. 1997). Although some empirical evidence supporting this approach has been developed (Long et al. 1998), elevated chemical concentrations are not, by themselves, reliable indicators of adverse natural resource effects. For example, it is not uncommon for biological assessments of sediment (in the form of sediment toxicity tests) to be inconsistent with analytical chemistry results (ie. High concentrations of hazardous substances present in the sediment yet the sediment sample is not toxic in a sediment toxicity test). At the Newport, Delaware Superfund Site there were areas of wetland sediment that showed "elevated" levels of some metals, yet there was little or no toxicity to benthic test organisms associated with these elevated levels (Laskowski 1993). Further, the abundance and diversity of benthic invertebrates were not appreciably depressed in this same area. In sum, one cannot categorically presume that elevated levels of a particular hazardous substance equate to injuries or service losses.

Defining "Reference" and "Control" Areas

To quantify service losses, the DOI regulations direct Trustees to compare populations, ecosystems, or habitats in the "assessment" area to corresponding populations, ecosystems, or habitats in a "control" area. The only types of measurements recognized in the regulations are numerical comparisons between "assessment" and "control" areas. Estimates of frequencies of injuries, as defined by the biological response measures used in injury determination, are considered population level measures by DOI. Trustees are encouraged to compare frequencies of occurrence of these response measures in "assessment" and "control" areas. The only non survey method identified in the guidance is the *in situ* bioassay using caged fish, as a means of demonstrating that hazardous substances are present in toxic quantities.

The quantification approach prescribed by DOI oversimplifies and obscures several critical issues in applied ecology. First, the use of the term "control" is inappropriate. As noted by Hurlburt (1984), the term "control" implies an experimental design in which replicated treatments (e.g., application of a chemical) are applied at random to some fraction of a set of experimental study plots or sites. In contrast, in NRDA assessments, there are no replicates and the "treatment" is not randomly applied. Under this circumstance, statistical tests for differences between sites can show that the sites are different, but cannot show that the difference is due to the fact that hazardous substances are present at one site but not at the other. The DOI regulations provide a list of criteria for selecting "control" sites that are as similar as possible to the assessment site, but there will always be unexplained, unmeasured differences between the sites that could be responsible for any observed differences in the quantity and quality of natural resource services. The challenge is to understand what those confounding influences might be and segregate them from differences thought to be caused by unpermitted releases of hazardous substances or oil.

Differences in population abundance or age structure between exposed and unexposed populations could be caused by factors unrelated to exposure to hazardous substances. In the case of a fish population for example, frequencies of individual-level indicators such as lesions, fin rot, and chemical body burdens would be useful for establishing that exposure has occurred and that individual fish have been affected, but these indicators are not direct measures of a loss of population viability or services. It would be necessary, in addition, to (1) evaluate the effects of exposure to hazardous substances on the survival, growth, and reproduction of each life stage, (2) quantify the distribution of exposures within the population, and (3) estimate the influence of the death or impairment of individual fish on the abundance and productivity of the population as a whole. Such an assessment could draw on controlled toxicity tests as well as on field studies, and could employ the same kinds of quantitative techniques used by fisheries scientists to estimate the impacts of fishing on the abundance and viability of fish populations (Barnthouse et al. 1990).

Similarly, in the case of a benthic invertebrate community, comparisons of community metrics between exposed and unexposed sites would be insufficient to quantify the reduction in community function attributable to hazardous substance exposure. Information would also be required concerning non-chemical influences (e.g., depth, temperature, and particle size), the toxicity of the hazardous substance(s) in question to different types of invertebrates, and the relationship between the structure of the benthic community, as measured by the relative abundance of different invertebrae taxa, and the function of the community, as measured by services (e.g., food) provided to other types of organisms. As in the case of fish populations, a combination of field studies, controlled experiments, evaluation of the peer-reviewed, published literature, and mathematical modeling might be required to address all of the relevant issues.

Using Field Studies to Quantify Injuries

The utility of the assessment approach prescribed in the DOI regulations is also reduced by the inherent limitations of most measurements of population and ecosystem-level characteristics. Field sampling programs generally cannot detect between-site differences of less than about 20% in estimates of population abundance or other similar measures. For some types of organisms, regular temporal cycles in abundance make between-site comparisons difficult under short time frames, particularly when there is a need to observe populations for multiple generations.

In short, because of problems associated with the lack of true experimental controls and the inherent variability of natural systems, apparent differences between "assessment" and "control" areas can be artifacts of unexplained environmental variation or natural population fluctuations. Nevertheless, field studies can be useful in quantifying natural resource injuries and service losses, provided that their limitations are recognized. Field studies can provide direct evidence of the health of populations and communities of wildlife, which is the appropriate focus of both ERA and NRDA. If such studies are used, however, it is important to take steps to overcome the limitations associated with the interpretation of the field data, such as by combining field survey data of the types prescribed by DOI with other field and laboratory-derived indicators of exposures and effects (Stahl and Clark 1998).

ERA practice provides several alternative approaches to using field data to quantify influences of chemicals and other disturbances on populations and ecosystems. One of the best-known approaches to characterizing differences between stressed and unstressed systems is the "bioindicators" approach. In this approach, compilations of data on the species composition and population characteristics in large numbers of ecosystems are used to define "regional reference" characteristics. Systems affected (or suspected of being affected) by chemical releases, runoff, or other disturbances are compared to the regional reference. Most often the differences between disturbed and reference systems are quantified using scoring systems such as the Index of Biological Integrity (Karr et al. 1986, Karr 1991).

The utility of this approach for NRDA is limited because undisturbed ecosystems are always used to define the regional reference characteristics. Pristine conditions are not the appropriate reference for ecosystems that are already disturbed at the time hazardous chemical releases occur. Moreover, the development of regional reference systems and indices requires data collected from a relatively large number of similar systems, using similar methods, at a reasonable cost. The concept of biological indices is best developed for small streams and rivers, because these kinds of systems are common and data collection is relatively inexpensive. Very large systems such as major rivers or estuaries are much more difficult to study.

An alternative to the biotic index approach is the use of multivariate statistical techniques. These methods can be used to compare conditions in a range of ecosystems, both stressed and unstressed, without identifying any particular ecosystem type as a reference. Multivariate methods have been used to compare streams with differing disturbance histories (Boyle et al. 1984), quantify results of mesocosm experiments (Johnson 1988, Kennedy et al. 1999, Kedwards et al. 1999a, Van den Brink and Ter Braak 1999), and compare ecosystems with differing exposures to hazardous waste releases (DOE 1995, Kedwards et al. 1999b). The major challenge in using multivariate techniques is in providing biological interpretations of the results, which typically are expressed as "principal components," "Mahalanobis distances," or other abstract mathematical quantities that do not have intuitive biological meanings.

Given the above considerations, perhaps the best method is to use multiple lines of evidence in a site-specific manner and to apply them in a weight-of-evidence approach focused at the population or community level. This issue is discussed further under the topic of causality.

Determining Causality

Causality is another complex and difficult issue, both for NRDA and for ERA. A distinction must be drawn between determining the cause of an unpermitted release or exposure and the cause of observed injuries and service reductions. Chemical source identification is a technically challenging activity in its own right, but it is not the subject of this paper.

The issue of interest here is determining the causes of population or ecosystem changes associated with service losses. Were the effects caused by a hazardous substance or oil, and if so, by which hazardous substance(s) or oil? For the reasons described above in connection with measurement of effects (i.e., the absence of a true experimental design with randomly assigned treatment and control sites), observed differences (and even statistically significant differences) between exposed and unexposed sites are seldom sufficient to prove causality. Populations may be in poor condition because of habitat modification, low dissolved oxygen, or other adverse conditions unrelated to hazardous substances. Establishing that some individuals were affected by a hazardous substance is not the same as establishing that adverse population or ecosystem characteristics were caused by that substance. To justify a natural resource damage claim, Trustees must prove that injuries to population or ecosystem characteristics, not merely individual characteristics, were caused by exposure to hazardous substances.

Studies that attempt to determine the causes, magnitude, and extent of effects observed in the field are termed "ecological epidemiology" (Fox 1991, Suter 1993). Suter (1993) provides an extensive discussion of criteria that can be used to infer causality in ecological studies. Suter (1993) recommends the use of multiple lines of evidence, preferably relating to different levels of biological organization. Under CERCLA, USEPA (1997) has also defined how lines of evidence might provide a sufficient inference of risk in an ERA. Note that inference is not the same as, nor should it be a substitute for, providing proof under the requirements faced by the natural resource Trustees.

One reason for requiring multiple lines of evidence is that, in general, the diagnostic value of any indicator decreases with increasing scale of biological organization (Figure 1). Cellular and molecular responses can sometimes indicate exposure to specific classes of chemicals, but are seldom good surrogates for effects. Two such responses are included in DOI's list: cholines-

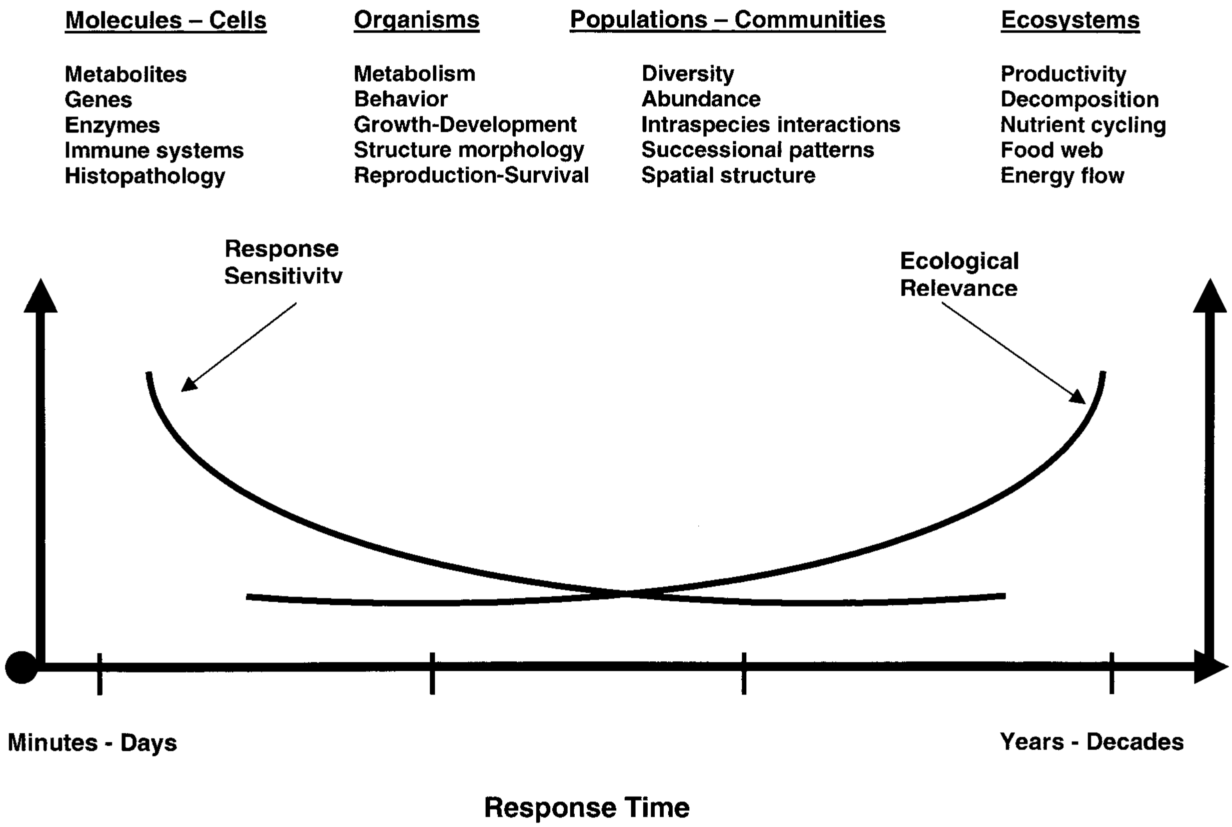

Figure 1. The competition between the ability to obtain precise and timely measurements, and the need for ecological relevance.

terase inhibition, which at least in birds is a specific indicator of exposure to organophosphate and carbamate pesticides, and ALAD inhibition, which is an indicator of exposure to lead. Tissue and whole-organism responses such as neoplasms, fin erosion, physical deformations, and behavioral anomalies may be caused by chemicals or other stressors (infections, temperature, etc.). However, all of these responses are general stress indicators that can occur in populations that are not exposed to hazardous substances.

Conversely, the ecological relevance of biological responses increases with increasing scale of biological organization. Molecular and cellular responses usually cannot be directly related to the health of the organism, because biochemical detoxification systems and other self-regulatory processes act to maintain essential physiological functions within a normal range (Dickerson et al. 1994). Tissue damage can be a better indicator of adverse effects, provided that the type of damage observed is clearly linked to the growth, survival, or reproduction of the organism. Direct measures of the growth, survival, or reproduction of organisms are better indicators of overall health, although they have a weaker connection to chemical exposure.

When evidence exists from a suite of indicators representing responses at multiple levels of organization, inferences concerning the causes of adverse population or ecosystem quality are strengthened (Adams 1990, Suter 1993). Munkittrick et al. (1994) used a weight-of-evidence approach and compared white sucker (*Catastomus commersoni*) populations downstream from 12 pulp mills to populations from five reference sites unaffected by pulp mill effluents. A variety of indicators of fish health and dioxin exposure were employed, including body size and age, organ weights, EROD induction, and sex steroid production. The use of multiple reference sites and multiple indicators of effects enabled the authors to infer, but not prove conclusively, that pulp mill effluents were responsible for reductions in the reproductive health of white sucker. However, they were not able to confirm dioxin releases as the cause of the reductions, because indicators of dioxin exposure were not correlated with the biological responses.

Several more formal approaches to establishing causality based on weight-of-evidence criteria have been developed and tested. The best-known of these is the "sediment quality triad" described by Chapman et al. (1991). The three components of the triad consist of (1) comparisons of chemical concentrations in sediment to published sediment-quality criteria or other published literature involving controlled exposures to specific chemicals of concern, (2) measures of the toxicity of the bulk sediment under controlled conditions, and (3) comparisons between benthic communities inhabiting contaminated sediment and benthic communities in unexposed sediment.

Each individual "leg" may be subject to multiple interpretations and provides at best incomplete evidence concerning effects of chemical exposure. As discussed earlier, in and of itself, comparison to published criteria or literature is not sufficient to demonstrate injury or cause, nor is it "proof" of injury or harm. Comparisons of this sort have application only in screening level efforts where individuals or groups are attempting to estimate the need for further study of a particular area.

Nevertheless, when the three legs of the triad are taken collectively, a positive response from all three elements provides strong evidence that toxic chemicals have adversely affected benthic communities. Obtaining negative responses from one or more of the elements indicates weak or unclear responses, or responses related to causes other than the hazardous substances or oil that are present in sediment.

The common thread linking all of these approaches is recognition of the fact that judgments concerning the causes of adverse ecological changes, including natural resource injuries and service losses, cannot be reduced to simple comparisons between exposed and unexposed sites. Multiple lines of evidence derived from both field studies and controlled experiments are generally required to demonstrate that (1) population or ecosystem characteristics that can be linked to service reductions have been adversely affected, and (2) the changes were due to exposure to hazardous substances rather than to confounding environmental influences or nonchemical stressors.

Establishing Baseline Conditions

According to NOAA and DOI, service losses are to be evaluated by comparing measurements of the level of service in the "assessment area" to a "baseline" level of service that would be provided but for the unpermitted release of hazardous substances or oil.

Historical data on characteristics of the assessment area prior to the unpermitted release of hazardous substances or oil may be used, if they are available. Sources of such data could include environmental impact statements, published scientific papers, computerized databases, and information obtained from local landowners or universities. Such sources can provide useful background information and can be very valuable for planning new field studies; however, only in rare cases will these kinds of data be directly applicable for determining baseline. Because historical studies most often have been conducted for reasons other than NRDA, the data are unlikely to meet the qualitative and quantitative criteria specified in the guidance. Moreover, environmental changes that occur between the time period covered by the historical data and the time of the hazardous substance or oil release would complicate the interpretation of historical data.

The DOI regulations recognize this possibility, and include a discussion of the use of samples obtained from "control" areas to quantify baseline services. The area selected for determining the baseline should have physical, chemical, and biological conditions as similar as possible to the assessment area but should not be exposed to unpermitted releases of hazardous substances or oil. The regulations provide a list of physical and biological data that should be evaluated for determining comparability. Moreover, the regulations specify that data collected from the control area should be collected "over a period sufficient to estimate normal variability in the characteristics being measured and should represent at least one full cycle normally expected in that resource."

As noted above, the term "control area" is inappropriate on methodological grounds, and the use of a single reference area does not reflect best scientific practice in ERA. Moreover, the requirement that data should be collected long enough to characterize variability and natural cycles could require years of data collection for many long-lived species.

An additional and perhaps even more serious problem is separating the effect of the hazardous substance release from the effects of confounding environmental disturbances and long-term temporal trends in environmental quality. A large fraction of ecosystems for which NRDAs are being pursued under CERCLA are in areas affected by multiple chemical releases, physical habitat alteration, nutrient enrichment, and invasions by exotic species. The DOI regulations make clear that the baseline service level should be the level provided by the existing system, including all development-related disturbances except for the specific hazardous substances or oil for which damages are being sought. "Control" areas that are affected by the same disturbances but have not been exposed to unpermitted releases of hazardous substances or oil are unlikely to be found. In most industrialized areas in the US, water quality has significantly improved over the past 20 years because of improvements in sewage treatment. The "baseline" service levels that exist today are substantially higher than those that existed in the past; injuries resulting from historic releases would have to be assessed relative to a baseline that no longer exists and may be impossible to evaluate. There is also the complicating factor of determining baseline "but for the release of hazardous substances or oil" when there are federally and state permitted discharges of the same materials, occurring across diverse habitats in the US.

One alternative to the use of a single control or reference area in estimating baseline is to use multiple reference areas that reflect a range of disturbance histories similar to that of the exposed area. Studying a series of areas provides a better estimate of the variability likely to be encountered in the assessment, and this helps the investigator bracket the range of potential responses. This approach was recently used by Clements et al. (2000) to quantify impacts of heavy metal exposures on benthic invertebrate communities in Colorado streams, and by Smith et al. (2001) to quantify impacts of contaminated sediment on macroinvertebrate communities of the Southern California mainland shelf.

In summary, the concept of a baseline, while simple in principle, can be difficult to implement in an unambiguous and rigorous way. By its very nature, the establishment of baseline will require rigorous scientific study and a case-by-case approach (site specific). Even though the term "baseline" connotes constancy and predictability, Nature, however, is inherently variable and unpredictable. Asking scientists to characterize a population or ecosystem that would have existed, except for the release of a hazardous substance or oil, is asking to characterize a state of nature that may have never existed. Without the hazardous substance release, the exposed resources would have varied in response to environmental fluctuations and endogenous biological processes that can be neither predicted nor controlled. All estimates of the services those resources would have provided are necessarily highly uncertain, and this is often the nexus at which disagreement arises between the Trustees and PRPs. Nevetheless, establishing baseline is crucial to quantifying injuries and service loss and a prime responsibility of Trustees when undertaking an NRDA. Where the estimate of baseline is supported by more or better scientific information, and where there is overlap between the values of baseline conditions from a Trustee perspective compared to a PRP perspective, resolution of this complex issue is much more likely to occur.

The Path Forward—Research Needs and Opportunities

Perhaps the greatest challenge facing Trustees and PRPs involved in NRDA proceedings is the development of an assessment process that leads to cost-effective and properly scaled restoration of natural resource services.

The planning process must, in addition to specifying data needed to confirm exposures and document injuries, consider approaches that may ultimately be used to quantify injuries and service losses. The specific data and model requirements best-suited to a particular assessment may be difficult to specify in advance. However, it should be possible to specify planning approaches that facilitate the development of cost-effective assessments.

Constraining ERAs to focus on those questions that must be answered to reach a risk management decision has been a focal point of several multi-stakeholder meetings between USEPA and the regulated community (Stahl et al. 2001, Pittinger et al. 1998), and is one area where NRDA could benefit from additional work. The only formal planning process discussed in the USEPA Guidelines is the Data Quality Objectives (DQO) approach developed by USEPA (1994c) and modified by DOE (1997). The DQO process is a strategic planning approach based on the scientific method that is used to prepare for a data collection activity. It provides a systematic procedure for defining the criteria that a sampling program design should satisfy, including when to collect samples, where to collect samples, the tolerable level of decision errors for study, and how many samples to collect. The process ties data collection to specific problems and decisions, and requires participation of both technical experts and decision makers.

In addition to the DQO approach, the inclusion of hypothesis testing into the NRDA process could also provide additional focus so that scientists and decision makers follow a clearly defined path to answering a clearly articulated question.

NRDA would certainly benefit from improved data and assessment methods; however, most of the specific types of research that would improve NRDA are the same types needed to improve ERA in general. Two specific needs are especially relevant to NRDA. One need is for better regional-scale data on environmental resource quality. Regional-scale environmental monitoring is already being implemented through EPA's Environmental Monitoring and Assessment program and NOAA's National Status and Trends Program. However, a great deal of information on regional resources is often available from state natural resource agencies, universities, and industry-sponsored monitoring programs. Synthesis of data collected by these diverse efforts can provide a much broad perspective on the level of service provided by potentially affected resources than can be obtained from restricted sampling of "assessment" and "control" areas.

A second need is for credible mathematical models that can bridge the gap between individual-level injuries and population-level service losses. Well-designed and properly tested models can integrate diverse types of information and extrapolate from readily-measured variables to variables that are more difficult to measure but are of greater management interest. Fish and wildlife managers routinely use these kinds of models to estimate impacts of harvesting on the abundance and productivity of managed populations. The feasibility of adapting models originally developed for fish and wildlife management purposes to assessments of ecological risks of chemicals has noted by Goodyear (1983) and Barnthouse et al. (1986, 1990). Recently developed techniques for "individual-based" population modeling now permit population-level consequences of behavioral (Pulliam 1994) and physiological (Jaworska et al. 1997) effects of chemicals to be directly quantified. Spatially-explicit modeling techniques (Dunning et al. 1995) permit evaluation of consequences of exposures that affect only part of a population or that affect the movement and spatial distribution of organisms. These tools would be useful for ERA as well, but for NRDA, they may be a necessity. Management applications of ecological models must, however, be supported by field and laboratory studies designed to estimate key parameters and to validate predictive capabilities. For example, site-specific data on exposure of specific resources, population-level variables (fecundity, age-class distributions, etc.) and other model inputs may be necessary. Models cannot substitute for the rigorous scientific data needed to meet the standards of proof required under NRDA regulations, but they can be used to guide data collection and to aid in interpreting the results. In addition to the development of the models themselves, standards for model parameterization and testing are needed. More importantly, it would be scientifically unsound to use any of these models to estimate biological injury without the appropriate validation. This is particularly true in those instances where important inputs to the model have been "generalized" so the model may be applicable under diverse situations, yet the site-specific situation differs substantially from the "generalized" situation used by the models' developers. Research commissioned on these types of tools would benefit both ERA and NRDA. Coupled with hypothesis testing and the DQO process, these tools could help resolve one of the more intractable problems in the NRDA process.

An improved dialogue among practitioners of NRDA from the Trustee and PRP communities could contribute to resolving the procedural and technical deficiencies of the NRDA process. As a first step, open, science-based technical workshops could be convened. Similar workshops have been highly valuable to USEPA and the regulated community in resolving both technical and policy issues in ERA. A series of these workshops could be undertaken through credible third party organizations such as the Society for Environmental Toxicology and Chemistry or RESOLVE. Such workshops would allow experts from the Trustees, PRPs and the academic communities to resolve technical problems in an open, science-based discourse. Having good solutions to the technical problems may then lead to a reexamination of and improvement in, policies and practices that drive the NRDA process.

Acknowledgments

The opinions expressed in this paper are those of the authors and do not represent official policies of any individual or organization named herein. The authors wish to thank several reviewers for their insightful and constructive comments on previous drafts of this paper.

Literature Cited

Adams, S. M. 1990. Status and use of bioindicators for evaluating effects of chronic stress on fish. *American Fisheries Society Symposium* 8:1–8

Barnthouse, L. W., J. Boreman, S. W. Christensen, C. P. Goodyear, D. S. Vaughan, and W. Van Winkle. 1984. Population biology in the courtroom. *BioScience* 34:14–19

Barnthouse, L. W., R. V. O'Neill, S. M. Bartell, and G. W. Suter II. 1986. Population and ecosystem theory in ecological risk assessment. Pages 82–96 in T. M. Poston and R. Purdy (eds.) Aquatic Toxicology and Hazard Assessment, 9th Symposium. American Society for Testing and Materials, Philadelphia, PA.

Barnthouse, L. W., G. W. Suter II, and A. E. Rosen. 1990. Risks of toxic contaminants to exploited fish populations: influence of life history, data uncertainty, and exploitation intensity. *Environmental Toxicology and Chemistry* 9:297–311

Barnthouse, L. W. 1997. Modeling ecological risks of pesticides: a review of available approaches. Pages 770–798 in G. Schüurman and B. Markert (eds.) Ecotoxicology. Swpektrum Akademischer Verlag, Berlin.

Boyle, T. P., J. Sebaugh, and E. Robinson-Wilson. 1984. A hierarchical approach to the measurement of changes in community structure induced by environmental stress. *Journal of Environmental Testing and Evaluation* 12:241–245.

Canfield, T. J., N. E. Kemble, W. T. G. Brumbaugh, F. J. Dwyer, C. G. Ingersoll, and J. F. Fairchild. 1994. Use of benthic invertebrate community structure and the sediment quality triad to evaluate metal-contaminated sediment in the upper Clark Fork River, Montana. *Environmental Toxicology and Chemistry* 13:1999–2012.

Chapman, P. M., R. N. Dexter, H. B. Anderson, and E. A. Power. 1991. Evaluation of effects associated with an oil platform, using the sediment quality criteria. *Environmental Toxicology and Chemistry* 10:407–424.

Clements, W. H., D. M. Carlisle, J. M. Lazorchak, and P. C. Johnson. 2000. Heavy metals structure benthic communities in Colorado mountain streams. *Ecological Applications* 10:626–638.

CMA. 1997. Ecological risk assessment: A tool for decision making. Chemical Manufacturers Association, Arlington, VA. 20 pp.

Department of Energy (DOE) 1995. Remedial Investigation/Feasibility Study for the Clinch River/Poplar Creek Operable Unit. DOE/OR/01-1393/V1&D1, Oak Ridge National Laboratory, Oak Ridge, TN

Dickerson, R.L., Hooper, M.J., Gard, N.W., Cobb, G.P. and Kendall, R.J. 1994. Toxicological foundations of ecological risk assessment: Biomarker development and interpretation based on laboratory and wildlife species. *Environmental Health Perspectives* 102 (Supplement 12):65–69.

Dobson, A. and Hudson, P. 1994. Assessing the impact of toxic chemicals: Temporal and spatial variation in avian survival rates. In: Kendall, R.J., and T. E. Lacher Jr., 1995. Wildlife Toxicology and Population Modeling. Lewis Publishers, Boca Raton, Florida. pp. 85–98

DOE 1997. Using the data quality objectives process during the design and conduct of ecological risk assessments. DOE/EH-0544. Office of Environmental Policy and Assistance, U.S. Department of Energy, Washington, DC

Dunning, J. B., D. J. Stewart, B. J. Danielson, B. R. Noon, T. L. Root, R. H., Lamberson, and E. E. Stevens. 1995. Spatially explicit population models: current forms and future uses. *Ecological Applications* 5:3–11.

Fields, T. 1997. CERCLA coordination with Natural Resource Trustees. OSWER Directive No. 9200-4-22A. July 31, 1997. U.S. Environmental Protection Agency, Washington, DC

Fox, G. A. 1991. Practical causal inference for ecoepidemiologists. *Journal Toxicology and Environmental Health*. 33:359–373.

Goodyear, C. P. 1983. Measuring effects of contaminant stress on fish populations. Pages 414–424 in Bishop, W. E., R. D. Cardwell, and B. B. Heidolph (eds.) Aquatic Toxicology and Hazard Assessment: Sixth Symposium. ASTM-STP 802, American Society for Testing and Materials, Philadelphia, PA

Harwell, M. J. Gentile, B. Norton, and W. Cooper. 1994. Issue paper on ecological risk assessment. Pages 2–1 through 2–49 *in* Ecological Risk Assessment issue Papers, EPA/630/R-94/009, US Environmental Protection Agency, Washington, D.C.

Jaworska, J. S., K. A. Rose, and A. L. Brenkert. 1997. Individual-based modeling of PCB effects on young-of-the-year largemouth bass in a southeastern USA reservoir. *Ecological Modelling* 99:113–135.

Johnson, A. R. 1988. Diagnostic variables as predictors of ecological risk. *Environmental Management* 12:515–523.

Karr, J. R., K. D. Fausch, P. L. Angermeier, P. R. Yant, and I. J. Schlosser. 1986. Assessing biological integrity in running waters: a method and its rationale. Illinois Natoral History Survey Special Publication No. 5. Champaign, IL.

Karr, J. R. 1991. Biological integrity: a long-neglected aspect of water resource management. *Ecological Applications* 1:66–84.

Kedwards, T. J., S. J. Maund, and P. F. Chapman. 1999a. Community-level analysis of ecotoxicological field studies. I. Biological monitoring. *Environmental Toxicology and Chemistry* 18:149–157.

Kedwards, T. J., S. J. Maund, and P. F. Chapman. 1999b. Community-level analysis of ecotoxicological field studies. II. Replicated-design studies. *Environmental Toxicology and Chemistry* 18:158–166.

Kennedy, J. H., L. P. Ammann, W. T. Waller, J. E. Warren, A. J. Hosmer, S. H. Cairns, P. C. Johnson, and R. L. Graney. 1999. Using statistical power to optimize sensitivity of analysis of variance designs for microcosms and mesocosms. *Environmental Toxicology and Chemistry* 18:113–117.

Kendall, R. J., and T. E. Lacher Jr., 1995. Wildlife Toxicology and Population Modeling. Lewis Publishers, Boca Raton, FL.

Laskowski, S. L. 1993. Record of Decision. E.I. DuPont, Newport Superfund Site, New Castle County, Delaware. USEPA Region 3. Philadelphia, PA. August, 1993.

Long, E. R., L. J. Field, and D. D. MacDonald. 1998. Predicting toxicity in marine sediments with numerical sediment quality guidelines. *Environmental Toxicology and Chemistry* 17:714–727.

Massachussets Department of Environmental Protection, 1996. Guidance on Weight-of-Evidence Evaluation.

Moore, R. H. 1978. Variations in the diversity of summer estuarine fish populations in Aransas Bay, Texas, 1966–1978. *Estuarine and Coastal Marine Science* 6:495–501.

Munkittrick, K. R., G. J. Van Der Kraak, M. E. McMaster, C. B. Portt, M. R. van den Heuvel, and M. R. Servos. 1994. Survey of receiving-water environmental impacts associated with discharges from pulp mills. 2. Gonadal size, liver size, hepatic EROC activity and plasma sex steroid levels in white sucker. *Environmental Toxicology and Chemistry* 13:1089–1103.

National Oceanic and Atmospheric Administration (NOAA). 1996. 15 CFR Part 990. Natural Resource Damage Assessments. Federal Register 61, January 5, 1996.

Pittinger, C. A., Bachman, R., Barton, A. L., Clark, J. R., deFur, P. L., Ells, S. J., Slimak, M. W., Stahl, R. G., and Wentsel, R. S. 1998. A multi-stakeholder framework for ecological risk management: Summary from a SETAC technical workshop. *Environmental Toxicology and Chemistry*, Supplement, 18(2):24.

Pulliam, H., R. 1994. Incorporating concepts from population and behavioral ecology into models of exposure to toxins and risk assessment. Pp. 13–26 in Kendall, R. J., and T. E.

Lacher, Jr. Wildlife Toxicology and Population Modeling. Lewis Publishers, Boca Raton, FL.

Roush, W. 1995. When rigor meets reality. *Science* 269:313–315.

Sanford, S. J., DiPinto, L. and Secord, A. 1998. Draft Scope for the Hudson River Natural Resource Damages Assessment Plan. The Hudson River Natural Resource Trustee Council (HRNRTC), New York State Dept. of Environmental Control, Albany, NY. September 1998.

Smith, R. W., M. Bergen, S. B. Weisberg, D. Caden, A. Dalkey, D. Montagne, J. K. Stull, and R. G. Velarde. 2001. Benthic response index for assessing infaunal communities on the southern California mainland shelf. *Ecological Applications* 11:1073–1087.

Stahl, R. G., Jr., Butler, P. B., Markarian, R. and Karmazyn, J. 1995. Ecovaluation, damage assessments and remedial options: An industrial framework. Toxic Substances in Water Environments: Assessment and control. WEF Specialty Conference Series, Water Environment Federation, Alexandria, VA, pp. 7–45 to 7–56.

Stahl, R. G., Jr., Bachman, R. A., Barton, A., Clark, J. R., deFur, P. L., Ells, S. J., Pittinger, C. A., Slimak, M. W. and Wentsel, R. S. 2001. (eds.) Risk Management: Ecological risk-based decision making. SETAC Press, Pensacola, FL, 202 pp.

Stahl, R. G., Jr. and Clark, J. R. 1998. Uncertainties in the risk assessment of endocrine modulating substances in wildlife. Chapter 18. In R. Kendall et al. (eds) Principles and Processes for Evaluating Endocrine Disruption in Wildlife. SETAC Press, Pensacola, FL, pp. 431–448.

Suter, G. W. II. 1989. Ecological endpoints. Pp. 2-1 to 2-28 in W. Warren-Hicks, B. R. Parkhurst, and S. S. Baker Jr. (eds.), Ecological assessment of hazardous waste sites: a field and laboratory reference document. EPA 600/3-89/013, U.S. Environmental Protection Agency, Corvallis, OR.

Suter, G. W. II (ed). 1993. Ecological Risk Assessment. 1993. Lewis Publishers, Boca Raton, FL.

Talmage, S. S., and B. T. Walton. 1991. Small mammals as monitors of environmental contaminants. *Reviews of Environmental Contamination and Toxicology* 119:47–145.

US Department of the Interior (DOI) 1995. 43 CFR Subtitle A, Part 11. Natural Resource Damage Assessments.

US EPA. 1991. Summary report on issues in ecological risk assessment. EPA/625/3-01/018. US Environmental Protection Agency, Washington, DC

US EPA. 1992a. Framework for Ecological Risk Assessment. EPA/630/R-02/001. US Environmental Protection Agency, Washington, DC

US EPA. 1992b. Peer review workshop report on a framework for ecological risk assessment. EPA/625/3-91/022. US Environmental Protection Agency, Washington, DC

US EPA. 1992c. Report on the ecological risk assessment guidelines strategic planning workshop. EPA/630/R-92/002. US Environmental Protection Agency, Washington, DC

US EPA. 1992d. The role of natural resource Trustees in the Superfund process. US Environmental Protection Agency, Washington, DC, ECO Update Vol. 1, No. 3, March, 1992.

US EPA. 1993. A review of ecological assessment case studies from a risk assessment perspective. EPA/630/R-92/005. US Environmental Protection Agency, Washington, DC

US EPA. 1994a. Ecological risk assessment issue papers. EPA/630/R-04/009. US Environmental Protection Agency, Washington, DC

US EPA. 1994b. Peer review workshop report on ecological risk assessment issue papers. EPA/630/R-94/008. US Environmental Protection Agency, Washington, DC

US EPA. 1994c. Guidance for the Data Quality Objectives process. EPA/QA/G-4, US Environmental Protection Agency, Washington, DC

US EPA. 1997. Ecological risk assessment guidance for Superfund: Process for designing and conducting ecological risk assessments. EPA 540-R-97-006. US Environmental Protection Agency, Washington, DC

US EPA. 1998. Guidelines for ecological risk assessment. EPA/630/R-95/001F. US Environmental Protection Agency, Risk Assessment Forum, Washington, DC, April, 1998.

Van den Brink, P. J., and C. J. F. Ter Braak. 1999. Principal response curves: analysis of time-dependent multivariate responses of biological community to stress. *Environmental Toxicology and Chemistry* 18:149–157.

Van Winkle, W., L. W. Barnthouse, and D. S. Vaughan. 1981. Analysis of the minimum detectable reduction in year-class strength of the Hudson River white perch population. *Canadian Journal of Fisheries and Aquatic Sciences* 38:627–632.

Weiss, J. C., Unsworth, R. E. and Ruder, E. 1997. Assessment plan for the natural resource damage assessment of the Grand Calumet River, Indiana Harbor Ship Canal, Indiana Harbor, and associated Lake Michigan environments. Prepared for: US Dept of Interior, State of Indiana. Industrial Economics, Inc. Cambridge, MA, 58 pp.

West Publishing Co. 1991. Selected Environmental Law Statutes. Comprehensive Environmental Response, Compensation and Liability Act of 1980, 42 U.S.C.A. 9601–9675, pp. 965–1056.

White, R. 1990. Natural resource damages: Trusting the Trustees. *San Diego Law Review* 27:407–466.