Guide for Benthic Invertebrate Studies in Support of Natural Resource Damage Assessment and Restoration

Guide for Benthic Invertebrate Studies in Support of Natural Resource Damage Assessment and Restoration

By David J. Soucek, Aïda M. Farag, John M. Besser, and Jeffery A. Steevens

Open-File Report 2022–1110

U.S. Geological Survey, Reston, Virginia: 2023

Contents

Figures

Tables

Conversion Factors

U.S. customary units to International System of Units

Multiply	By	To obtain
Length		
inch (in.)	25,400	micrometer (μm)

International System of Units to U.S. customary units

Multiply	By	To obtain
Length		
micrometer (μm)	0.00003937	inch (in.)

Abbreviations

ANOVA	analysis of variance
BMI	benthic macroinvertebrate
DOI	U.S. Department of Interior
FFG	functional feeding group
NRDAR	Natural Resource Damage Assessment and Restoration

Guide for Benthic Invertebrate Studies in Support of Natural Resource Damage Assessment and Restoration

By David J. Soucek, Aïda M. Farag, John M. Besser, and Jeffery A. Steevens

Introduction

This guide is intended to assist with characterizing injury to freshwater benthic macroinvertebrates (BMIs) in Natural Resource Damage Assessment and Restoration (NRDAR) cases. The contents are narrowly focused on insects, crustaceans, snails, and other invertebrate fauna that are typically considered part of BMI communities and are not intended to address studies of injury to larger benthic taxa such as freshwater mussels, crayfish, or benthic fishes or amphibians. Although some percentage of the community functions as predators, BMIs are predominantly primary consumers (for example, scrapers, shredders, and filterer/gatherer feeding groups [detailed in the "Identification and Metrics" section]) that play an essential role in converting carbon and nitrogen from plant tissues into animal biomass for higher-order consumers, especially in flowing waters. Aquatic contaminants can disrupt the quantity and quality of energy transferred (ecosystem function) by reducing invertebrate biomass and diversity. Additionally, the accumulation of toxic residues in invertebrate tissues may be a source of exposure leading to adverse effects in higher trophic levels. The goal of NRDAR BMI assessments is to establish direct linkages of contaminant exposure to injuries reflected by changes in community structure (for example, reduced density and taxa richness) or by effects at the individual population level (for example, survival, growth, and reproduction). BMIs are infrequently the U.S. Department of Interior (DOI)-managed resource in a NRDAR case, with managed resources more frequently including migratory birds, fish, or other insectivorous vertebrates. Therefore, it is critical to have clearly defined objectives for evaluating BMIs and an understanding of how invertebrate data relate to the quantification of injuries to the DOI-managed resource; some of these pathways are shown in figure 1. This guide is intended to assist decisions on whether or not to proceed with BMI studies, use of existing information and data for screening purposes, and what types of studies can support a BMI-injury determination. Examples of the latter are illustrated in figure 2, and these broader categories of study types form the basis for the sections that follow. This document is intended to provide general considerations and best practices for assessing BMIs. Relevant guidance and references are listed throughout the report as sources for specific methods and analysis.

The following components of (BMI) studies are covered in the following guide:The "Characterizing Chemical Exposure" section discusses the following:

- Exposure may be by way of water, sediment, and (or) diet.

- For chemical of interest, considerations include concentration, potential toxicity, mobility, persistence, fate, and factors that affect bioavailability.

- Contaminants in invertebrate tissue may lead to adverse effects in higher trophic levels.

The "Benthic Community Surveys" section discusses the following:

- Study objectives and habitat influence methods for sampling, processing, and identification.

- Abundance and taxonomic composition are the frequent basis for benthic assessment.

- Documentation of habitat quality is required to clarify contaminant effects.

The "Toxicity Testing" section discusses the following:

- Effects may be quantified at individual or population levels while controlling noncontaminant factors.

- Standardized test methods rely on sensitive, easily cultured test organisms.

- In situ toxicity tests are a compromise between causal inference of laboratory toxicity tests and site specificity of field assessments.

The "Data Analysis" section discusses the following:

- Relations between exposure and effects should be demonstrated, and water quality and habitat variation should be incorporated as influential and (or) confounding factors.

- Responses of invertebrates exposed to contaminants are compared to those under reference conditions.

- Thresholds for injury are estimated based on concentration-response models.
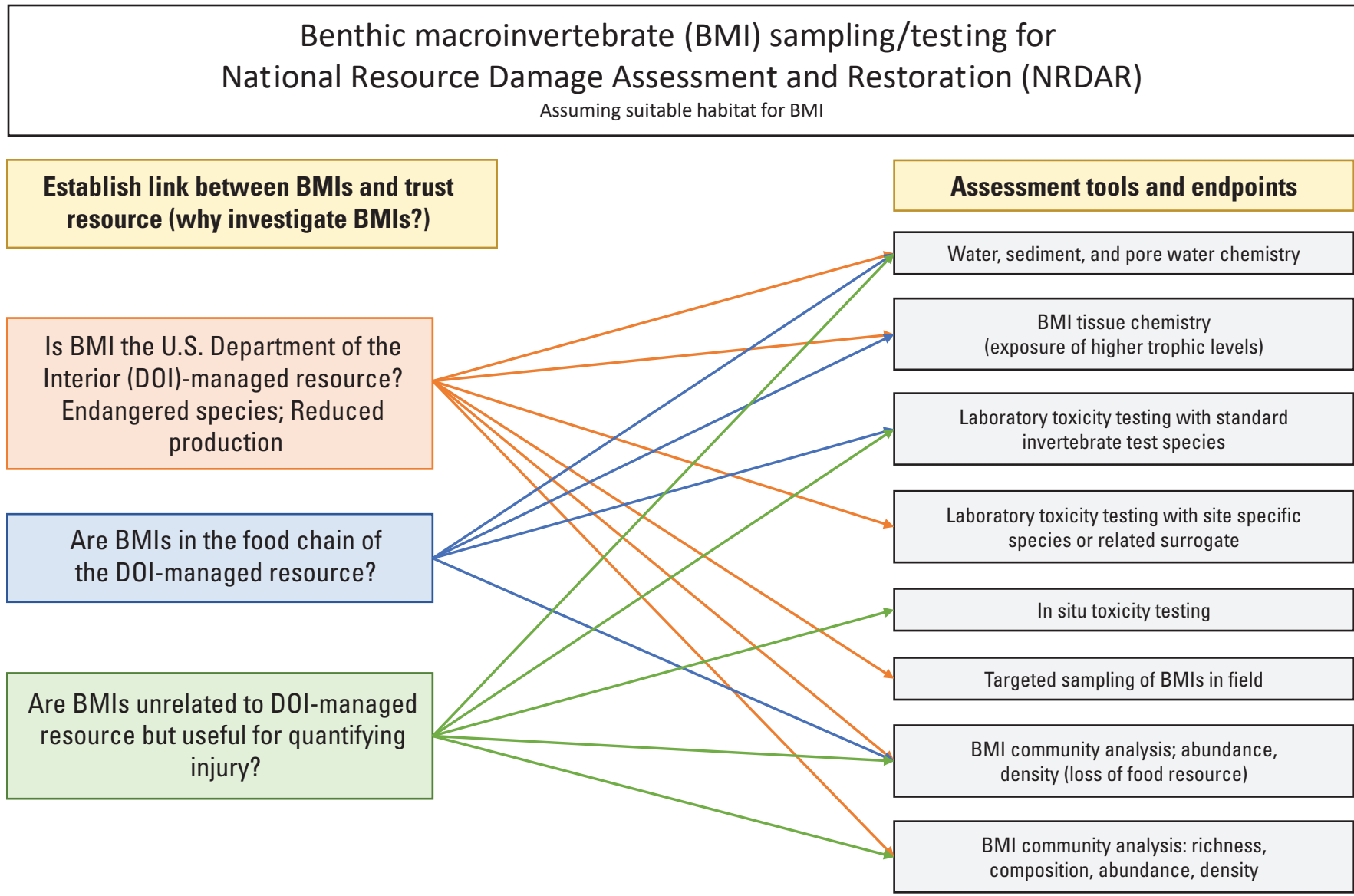
Figure 1.   Benthic macroinvertebrate sampling and testing tools for Natural Resource Damage Assessment and Restoration.
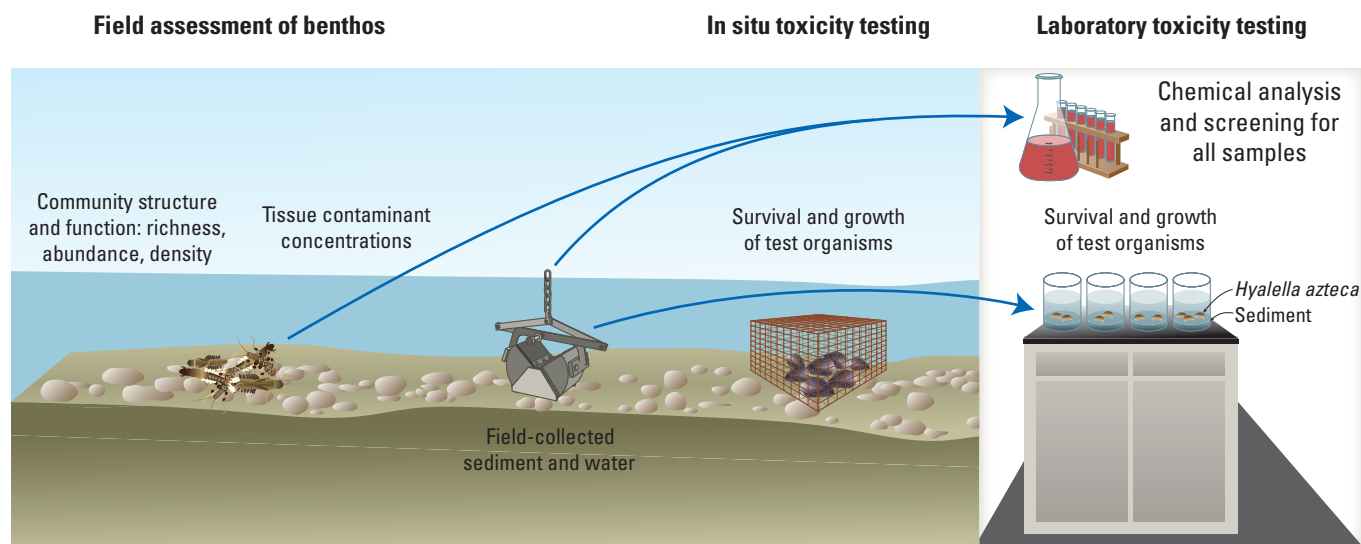
Figure 2. Overview of the various types of studies that can support benthic macroinvertebrate injury determination.

The "Monitoring Restoration Success" section discusses the following:

- Methodologies for assessing restoration are similar to methods for assessment.

- Long-term monitoring can be used to support adaptive management.

Characterizing Chemical Exposure

Characterizing injury to an aquatic ecosystem requires understanding of the nature of chemical contaminants present and their concentrations in environmental media. This information is important whether the BMI community (or a particular taxon) is the receptor of interest or is viewed as a potential source of contaminants to organisms at higher trophic levels. Armed with an initial understanding of the range of chemicals present in water or sediment, it is important to identify which chemicals and exposure levels pose substantial hazard of toxicity and (or) bioaccumulation. It is also important to understand factors like mobility and persistence of these chemicals for efficient targeting of samples. For example, toxic effects of ammonia are often related to acute effects of dissolved ammonia; however, toxic effects of persistent pollutants like polychlorinated biphenyls and metals are more likely a result of chronic exposure by way of water, diet, and sediments. In addition to understanding the source of contaminant exposures, knowledge of chemical cofactors that influence contaminant bioavailability may be necessary. Bioavailability is the amount of a chemical that is taken up from the environment and able to elicit an adverse effect. These cofactors may

include, but are not limited to, sediment organic matter, acid-volatile sulfide, pH, dissolved organic carbon, alkalinity, and hardness.

Although measuring contaminants in environmental media is essential for estimating injury, measuring contaminant concentrations in BMI tissue may also be necessary for demonstrating exposure of the invertebrate. Measuring contaminant concentrations in invertebrates can be difficult because of the large dry mass of tissue required to achieve acceptable analytical detection limits; often, large numbers of organisms must be composited to produce a single sample. If the primary interest is in food chain transfer from invertebrates to higher trophic levels (for example, fish or amphibians), and the diet of the higher levels is well characterized, it may be more efficient to focus on concentrations in the higher trophic level receptors. However, in some cases where tissue concentrations are measured, there may be sufficient information to relate tissue concentrations in invertebrates to expected effect levels (Jarvinen and Ankley, 1999). Additionally, some persistent contaminants including polychlorinated biphenyls, mercury, and metals can remain in the tissues of emergent adult insects, which can transfer aquatic contaminants into terrestrial ecosystems (Kraus and others, 2014).

Characterizing chemical exposure with a limited scope can also be a useful reconnaissance tool for practitioners. Reconnaissance is a preliminary survey of the area and can help a practitioner determine the potential next steps. A preliminary sample collection during reconnaissance may be done in-house with limited resources, and it does not need to necessarily meet the rigor reserved for the NRDAR sample collection and preparation. However, data generated during reconnaissance could assist in directing future sampling efforts should they be deemed useful.

Benthic Community Surveys

The process of sampling and evaluating of BMI communities is roughly divided into three major activities: field sampling, sorting and counting of the BMI, and identification of BMI and calculation of metrics. Issues related to these major activities are discussed further in the following sections.

Field Sampling

Sampling methodology for the BMI community analysis is extremely varied, and the selection of methods depends on the objectives of the study and other factors such as water current velocity and habitat type. Broad categories of habitat types include lentic (still water) and lotic (flowing water) habitats. The lotic habitats have been further divided, with respect to sampling gear, into shallower streams or rivers that tend to be higher gradient, and larger, deeper, lower gradient rivers. Furthermore, different sampling devices may be selected depending on the nature of bottom substrata in the study area, which may range from fine sediment in pools to cobbles in riffles. Qualitative and quantitative sampling methods are available for all habitats, and these methods provide different information. Qualitative sampling focuses on the number of taxa present and their relative abundances (semiquantitative), whereas quantitative sampling determines absolute abundance, density, or even biomass/secondary production for a defined physical area. Merritt and others (2019a) provide an extensive list of options for sampling methods in lotic and lentic habitats.

Most State agencies tasked with addressing requirements of the Clean Water Act (33 U.S.C. 1251 et seq.) sample BMI communities with a kick-net device like a D-frame net, most commonly with a 500-micrometer (µm) mesh-size (Carter and Resh 2001; 2013). These programs frequently follow protocols similar to those described in the U.S. Environmental Protection Agency's "Rapid Bioassessment Protocols for Use in Streams and Wadeable Rivers—Periphyton, Benthic Macroinvertebrates and Fish"(Barbour and others, 1999). For trustees, it may be prudent to follow protocols established by the State in which the NRDAR is located. Additionally, coordination or cooperative sampling with the State is recommended to benefit from additional expertise and potentially comparable data.

A critical component of understanding differences in the structure and function of a BMI community is assessment of the physical habitat. When possible, it is desirable to sample BMIs from habitats of equal quality at reference and contaminated sites, so that chemical stressors are not confounded. Detailed and time-intensive quantitative techniques can be used to assess physical habitats like pebble counts (Wolman, 1954), particle size distributions, and grain classifications (Blair and McPherson, 1999), and the propensity of a streambed to change or shift, otherwise known as relative bed stability (Kaufmann and others, 1999). Alternatively, Barbour and others (1999) provide detailed and widely used protocols for visually assessing physical habitats by assigning sites into well-defined categories (ranging from poor to optimal) for 10 different measures of physical habitats that are supportive of BMI communities. This process allows for less investment of time and results in a single numerical value that represents the quality of the habitat based on the scores for the individual metrics. Similar to macroinvertebrate sampling, it may be prudent to follow protocols for habitat assessment that are utilized by the State in which the NRDAR is located.

Sorting and Counting Benthic Macroinvertebrates

After sampling in the field, preserved samples are typically returned to the laboratory for separation, identification, and the counting of invertebrates collected. Out of the multiple methods available for counting, the primary choice is between the subsampling method and the full enumeration method. Qualitative and semiquantitative sampling methods typically involve subsampling, which is when a fixed count of organisms, often with an allowable range (for example, plus or minus 10 percent) is separated from litter material. Fixed counts can be limited to between 100 to 550 organisms (Carter and Resh, 2001). Although full enumeration of every organism collected in a sample is time consuming, there are multiple benefits to this approach. These benefits include the following:

- The identification of a greater number of taxa, and, in particular, a greater number of rare taxa (Pence and others, 2021).

- The ability to estimate total abundance and, if a fixed area was sampled, density of organisms per unit area.

In full enumeration studies, the greater number of taxa detected can result in a greater ability to detect differences in community metrics between moderately polluted streams and reference streams (Pence and others, 2021). Another consideration is that the effort required to fully enumerate a contaminated site may be small if the invertebrates are limited, and defining this condition can assist in injury determinations. Conversely, if the physical habitat is somewhat poor at the site of discharge/release and the available nearby reference sites, using the minimum fixed counts of invertebrates may prevent a later need for censoring data at a site where too few macroinvertebrates were collected. Whichever method is selected, it is important to use the same method of sampling at both reference and contaminated sites.

Identification and Metrics

Once invertebrates have been sorted and enumerated, organisms are identified to some level of taxonomic specificity. When possible, identifications can be made by an experienced or certified taxonomist; for example, the Society

for Freshwater Science conducts a taxonomy certification program. As an alternative, some digital photographs of various types of invertebrates that were collected and identified for verification could be maintained at a later time as needed. Identifications are usually made at least to family level, but preferably to "lowest practical taxonomic level," which for many taxa is genus. Identification to species level has some benefits, but it typically requires a high degree of specific expertise, as well as often rearing organisms to adult stage. This information is used to calculate various metrics or multimetric indices. A list of metrics that have been tested repeatedly for responsiveness to environmental perturbations, including the predicted directional response to perturbation, is provided in table 7–1 of Barbour and others (1999). These metrics fall into several categories, including metrics based on the following:

- Counts of taxa present (richness).
- Composition (for example, the percent of the community that is in the order Ephemeroptera [mayflies]).
- Tolerance measures.
- Functional feeding groups (FFGs).
- Habit measures, related to how organisms move or stay stationary within their environment.

Multimetric indices combine several metrics into a single value that is calibrated and known to be useful for delineating references from impacted conditions. Many of these indices were developed for a specific State, region, or habitat type, and they should only be used in locations or ecoregions for which they were developed and (or) calibrated.

For purposes of NRDAR, useful metrics include those based on the number of taxa, abundance, density, or percent of a community that belongs to three sensitive insect orders: Ephemeroptera (mayflies), Plecoptera (stoneflies), and Trichoptera (caddisflies), otherwise known as the "EPT" taxa (Barbour and others, 1999). These metrics are frequently useful for demonstrating the departure from the baseline. The number of Ephemeroptera taxa and percent Ephemeroptera are both particularly sensitive to certain types of contaminants like metals and major ions (Pond and others, 2008).

Merritt and others (2019a) provide extensive tables assigning individual insect genera to FFGs, which are broad categorizations of how invertebrates obtain nutrition. Metrics based on FFGs may be sensitive to certain contaminants but also provide information about ecosystem function. For example, if a contaminant causes a particular taxon to be absent from a site, evaluating FFG metrics can demonstrate whether other taxa provide functional redundancy or if there is a loss of the ecosystem service they provide. Also, FFG metrics may be especially useful in combination with quantitative sampling and full enumeration, allowing estimates of the abundance and (or) density of taxa present. Reduced density of certain FFGs could also be an indication of a reduction in ecosystem function when a richness metric might not indicate this loss. For example, the areal density of grazing mayflies could be even more indicative of injury than simply how many kinds of mayflies are present. Although potentially useful for assessing ecosystem services, the relevance of FFGs to an injury claim should be clearly defined.

Results of BMI surveys may be confounded by temporal variation of the target community. Natural variation can be driven by thermal patterns, flow patterns, or simply by the different phenologies of life cycles for different taxa. These factors can affect which taxa are present at the site and the size class of the present organisms. Size is not only correlated with sensitivity, but it also affects the ability to detect and identify invertebrates. Observed differences in community metrics may be influenced by this natural variation due to differences in when samples are collected (Boehme and others, 2016). For this reason, methods, manuals, and State agencies typically define an "index period(s)", during which time samples should be collected. These frequently span through the summer months, such as June–September, or when a State has two index periods, such as March–June and August–December. Again, the number and duration of index periods vary and depend on the location of interest.

A final point on BMI community assessment is to be cognizant of other sources of data or information related to the site of interest that has not been necessarily collected as part of the NRDAR process. Evaluation and incorporation of these data may be assisted by considering points detailed in this section. The "Key Considerations in Selecting Methods for BMI Community Sampling and Analysis" sidebar highlights key considerations for sampling and the assessment of BMI communities.

Key Considerations in Selecting Methods for BMI Community Sampling and Analysis*:

- What is the broad habitat category (lentic, shallow lotic, deep lotic), and the fine-scale habitat of interest (riffle, run, pool)?

- Should sampling for macroinvertebrates target multiple habitats or a single habitat type?

- Is injury likely to reflect direct impact on benthic community structure or function, or indirect effects by way of contaminant uptake and trophic transfer?

- Will qualitative methods be sufficient or is quantitative analysis required?

- How will injury be evaluated (for example, comparing multimetric index reference values, or statistical comparison of metrics among sites)?

- When sorting and identifying organisms, will subsampling suffice, or will full enumeration be beneficial?

- How will the sampling season potentially impact conclusions?

*Case managers should consult with the assigned DOI economist and attorney in determining the answers to these questions. The sampling design should align with NRDAR injury quantification and scaling methods.

Toxicity Testing

Assessment of injury to benthic communities based on field surveys alone has some limitations, notably the possibility that habitat degradation not associated with oil or hazardous substance releases may modify invertebrate communities. Associations between exposure to chemical pollutants and community responses can be further documented with controlled laboratory toxicity tests. These tests may focus on documenting contaminant effects in field-collected water or sediment compared to reference sites, or they may focus on characterizing site-specific toxicity thresholds for contaminants of concern.

Laboratory toxicity tests typically rely on standardized methods that provide a high level of confidence in testing results. As discussed in the "Characterizing Chemical Exposure" section, the environmental fate of the oil or hazardous substance of interest will determine exposure route and the type of toxicity test (that is, surface water, pore water, or sediment) that should be performed. Successful laboratory tests can provide rigorous proof of cause-and-effect relations between chemical exposure and responses of invertebrates. The most widely applicable strategy for evaluating toxic effects on benthic invertebrates is laboratory tests with field-collected sediments (American Society for Testing and Materials, 2020). Tests for toxicity of effluents, which are also applicable to testing ambient waters, are designed to test fish and planktonic invertebrates; however, they may be adapted for testing benthic invertebrates as well. Standard test methods can be used for repeatedly testing sediment from sites undergoing remediation to quantify the recovery of invertebrate populations (Steevens and others, 2020). This repeated laboratory testing is most valuable if it is performed in coordination with ongoing biomonitoring of invertebrate communities.

Laboratory tests have limitations as well, such as difficulty of maintaining exposure conditions (for example, water quality, food quality, and dietary exposure) comparable to conditions in the field. Comparisons of laboratory and field responses may be hampered by the difficulty of calibrating individual responses (survival, growth, and fecundity) from toxicity tests with aggregate responses (species richness, age/size structure, and abundance) from community surveys. One approach that may reduce this laboratory-field disconnect is the modification of standard methods to allow testing with resident species from the study area or closely related surrogates.

Studies with test organisms exposed to chemicals in situ have been recommended to avoid some of the limitations of both laboratory toxicity tests (for example, lack of site specificity) and field surveys (for example, limited evidence of causality) (Burton and others, 2005). Test organisms deployed in the field presumably experience more realistic exposure conditions, making their individual-level responses more directly comparable with responses of resident invertebrate communities. In situ assessments are especially well-suited for documenting effects of episodic or transient pollution events. However, the rarity of in situ tests reflects the limitations of the approach because cages may create unfavorable holding conditions that obscure site effects. In situ tests are subject to greater risks from stochastic events such as vandalism, sedimentation, and high- or low-flow events. The biggest limitation of in situ testing with invertebrates likely is the lack of standard methods and the difficulty of finding laboratories or investigators with the ability to perform in situ tests.

Data Analysis

In all situations regarding collecting data for NRDAR, data must meet quality-assurance and -control guidelines to be legally appropriate for the NRDAR case. Barbour and others (1999) and U.S. Environmental Protection Agency (2016) among others provide guidance on appropriate quality-assurance procedures for activities ranging from sample labelling to taxonomic identification. Other data-quality questions include the following:

- Are the sampling methods appropriate for the ecosystem of interest?

- Do laboratory toxicity tests meet test acceptability criteria (for example, table A2.1 *in* American Society for Testing and Materials, 2020)?

- Were appropriate quality-control samples incorporated in analytical chemistry?

Data from field and laboratory studies with benthic invertebrates may be analyzed with a variety of statistical tools, depending on the objectives of the study. Multivariate statistical methods, such as ordination and cluster analysis, are useful as exploratory data analyses to look for spatial or temporal patterns in community data and associations of these data with environmental conditions. Results of these analyses may be helpful for identifying chemicals of concern and sensitive taxa for toxicity studies.

Significant observed toxic effects are often determined by the analysis of variance (ANOVA) or related techniques, and the responses in contaminated field sites or higher laboratory exposure levels can be compared to baseline responses. The nature of these comparisons differs between laboratory and field studies. In laboratory toxicity tests, the baseline response is estimated by a negative control that receives no added chemical but is known to support survival, growth, or reproduction of the test organisms. Controls are not available in field studies, so investigators may select one of the following:

- A site-specific reference, either upstream of contamination within the same watershed, or in an adjacent but unaffected watershed.

- A regional reference condition, sometimes termed as "reference envelope," wherein baseline responses are drawn from a suite of relatively unimpaired sites.

Because it is normal for baseline responses to vary among low-chemical reference sites, determinations of impairment are typically made by comparison to the range of responses from multiple reference samples. For example, a common approach is to define a deviation from reference condition as a high or low quantile (for example, 10th percentile) of the reference distribution (Hawkins and others, 2010). Specifically, low quantiles are used for metrics that decline with increased perturbation and high quantiles are used for those that increase with increased perturbation. Further background on the reference condition and considerations useful in selecting reference sites can be found in Hughes (1995), Reynoldson and others (1997), Barbour and others (1999), Stoddard and others (2006), and the U.S. Environmental Protection Agency (2006, 2016). Another potential approach for showing departure from the baseline is the Before-After-Control-Impact model, which requires temporal baseline data prior to the discharge or release of oil or a hazardous substance (Smith and others, 1993).

If ANOVA detects significant toxicity in a laboratory or field study, the next step in the data analysis is often to resort to concentration-response modeling that utilizes nonlinear regression to estimate threshold concentrations for toxic effects. Based on ANOVA, the concordance of spatial patterns of toxicity with exceedance of toxicity thresholds, which are estimated by concentration-response models, provides strong evidence of exposure and effects. The estimation of thresholds by regression methods to estimate a specified level of effect is a widely accepted approach for demonstrating departure from control response(s). Thresholds for different levels of effect may be appropriate to meet the objectives of NRDAR studies. If conditions for laboratory tests are comparable to those in the field, modeled effect concentrations may serve as benchmarks for assessing toxicity in sites that have not been tested directly. Pathways for appropriate assessment tools and endpoints for BMIs, given the relation between the BMI and the DOI-managed resource, are provided in figure 1. Table 1 expands on figure 1 with more details on the tools, endpoints, and references that are potentially useful in injury quantification and claim development.

Table 1.   Examples of benthic invertebrate assessment tools and endpoints for Natural Resource Damage Assessment and Restoration claim development.

[EPT, Ephemeroptera (mayflies), Plecoptera (stoneflies), and Trichoptera (caddisflies)]

Tool	Endpoint	Key References
Water chemistry	Ambient water quality criteria	Ambient water quality criteria (https://www.epa.gov/wqc/national-recommended-water-quality-criteria-aquatic-life-criteria-table)
Sediment and (or) pore water chemistry	Sediment quality guidelines	MacDonald and others, 2000
	Equilibrium partitioning sediment benchmarks	U.S. Environmental Protection Agency, 2003, 2005, 2008; Cleveland and others, 2017
Tissue chemistry	Tissue effect thresholds, diet effects thresholds	Jarvinen and Ankley, 1999; Adams and others, 2011; McElroy and others, 2011; Steevens and others, 2005; Kraus and others, 2020
Water or sediment toxicity testing	Survival, growth, reproduction, biomass	American Society for Testing and Materials 2016; U.S. Environmental Protection Agency, 2002a, b; American Society for Testing and Materials, 2020; U.S. Environmental Protection Agency, 2000
Benthic macroinvertebrate community sampling	Abundance/density: overall, EPT, sensitive taxa, functional feeding groups	Barbour and others, 1999; Carter and Resh, 2013; Hughes, 1995; Kaufmann and others, 1999; Merritt and others, 2019b; Moulton and others, 2000, 2002; U.S. Environmental Protection Agency, 2006, 2013b, 2016

Monitoring Restoration Success

A similar suite of techniques used for assessing baseline and injury can be used to monitor remediation or restoration efforts and document movement along the restoration trajectory. For BMIs, careful consideration must be given to points made in the "Key Considerations in Selecting Methods for BMI Community Sampling and Analysis" sidebar, particularly the seasonality of sampling during restoration monitoring. For example, after remediation or restoration, it would be most common to do the following:

- Compare contaminant concentrations in water and sediment against known thresholds.

- Test water and (or) sediment toxicity in the laboratory after cleanup.

- Assess the suitability of remediated and restored physical habitat.

- Compare BMI communities to upstream sites or regional reference conditions.

In some instances, when an active restoration effort is not considered advantageous for the system, managers may prefer to monitor over time (months, years, or decades) to assess recovery rates to support adaptive management decisions (Magar and others, 2009). Frequency of monitoring after restoration is case-specific; however, take note of the following considerations: (1) in some cases, remedial efforts could initially make things worse for invertebrate communities; (2) if remediation of the release or discharge is successful and physical habitat is rehabilitated, invertebrate communities may respond quickly; and (3) stochastic events like drought(s) or flooding may necessitate changes in a monitoring plan. A higher frequency of sampling may be necessary until the spatial and temporal variations in data are characterized.

Selected References

Adams, W.J., Blust, R., Borgmann, U., Brix, K.V., DeForest, D.K., Green, A.S., Meyer, J.S., McGeer, J.C., Paquin, P.R., Rainbow, P.S., and Wood, C.M., 2011, Utility of tissue residues for predicting effects of metals on aquatic organisms: Integrated Environmental Assessment and Management, v. 7, no. 1, p. 75–98. [Also available at https://doi.org/10.1002/ieam.108.]

American Society for Testing and Materials, 2016, Standard guide for conducting acute toxicity testing on test materials with fishes, macroinvertebrates, and amphibians: West Conshohocken, Pa., ASTM International, p. E729–E796.

American Society for Testing and Materials, 2020, Standard test method for measuring the toxicity of sediment-associated contaminants with freshwater invertebrates, chap. 13 *of* Experimental design and statistical analysis; Annexes A1-A5, Detailed test methods for selected invertebrate species: West Conshohocken, Pa., ASTM International, p. E1706–E1720.

Barbour, M.T., Gerritsen, J., Snyder, B.D., and Stribling, J.B., 1999, Rapid bioassessment protocols for use in streams and wadeable rivers—Periphyton, benthic macroinvertebrates and fish (2d ed.): Washington, D.C., U.S. Environmental Protection Agency, Office of Water, EPA 841–B–99–002, 337 p.

Besser, J.M., Ivey, C.D., Steevens, J.A., Cleveland, D., Soucek, D.J., Dickinson, A., Van Genderen, E.J., Ryan, A.C., Schlekat, C.E., Garman, E., Middleton, E., and Santore, R., 2021, Modeling the bioavailability of nickel and zinc to *Ceriodaphnia dubia* and *Neocloeon triangulifer* in toxicity tests with natural waters: Environmental Toxicology and Chemistry, v. 40, no. 11, p. 3049–3062. [Also available at https://doi.org/10.1002/etc.5178.]

Blair, T.C., and McPherson, J.G., 1999, Grain-size and textural classification of coarse sedimentary particles: Journal of Sedimentary Research, v. 69, no. 1, p. 6–19. [Also available at https://doi.org/10.2110/jsr.69.6.]

Boehme, E.A., Zipper, C.E., Schoenholtz, S.H., Soucek, D.J., and Timpano, A.J., 2016, Temporal dynamics of benthic macroinvertebrate communities and their response to elevated specific conductance in Appalachian coalfield headwater streams: Ecological Indicators, v. 64, p. 171–180. [Also available at https://doi.org/10.1016/j.ecolind.2015.12.020.]

Bruce, J.F., Roberts, J.J., and Zuellig, R.E., 2018, Comparability among four invertebrate sampling methods and two multimetric indexes, Fountain Creek Basin, Colorado, 2010–2012: U.S. Geological Survey Scientific Investigations Report 2018–5061, 11 p., accessed April 2022 at https://doi.org/10.3133/sir20185061.

Burton, G.A., Jr., Greenberg, M.S., Rowland, C.D., Irvine, C.A., Lavoie, D.R., Brooker, J.A., Moore, L., Raymer, D.F.N., and McWilliam, R.A., 2005, In situ exposures using caged organisms—A multi-compartment approach to detect aquatic toxicity and bioaccumulation: Environmental Pollution, v. 134, no. 1, p. 133–144. [Also available at https://doi.org/10.1016/j.envpol.2004.07.008.]

Cain, D., Croteau, M.N., and Luoma, S., 2011, Bioaccumulation dynamics and exposure routes of Cd and Cu among species of aquatic mayflies: Environmental Toxicology and Chemistry, v. 30, no. 11, p. 2532–2541. [Also available at https://doi.org/10.1002/etc.663.]

Carter, J.L., and Resh, V.H., 2001, After site selection and before data analysis—Sampling, sorting, and laboratory procedures used in stream benthic macroinvertebrate monitoring programs by USA state agencies: Journal of the North American Benthological Society, v. 20, no. 4, p. 658–682. [Also available at https://doi.org/10.2307/1468095.]

Carter, J.L., and Resh, V.H., 2013, Analytical approaches used in stream benthic macroinvertebrate biomonitoring programs of State agencies in the United States: U.S. Geological Survey Open-File Report 2013–1129, 50 p., accessed April 2022 at https://doi.org/10.3133/ofr20131129.

Chen, K., Hughes, R.M., and Wang, B., 2015, Effects of fixed-count size on macroinvertebrate richness, site separation, and bioassessment of Chinese monsoonal streams: Ecological Indicators, v. 53, p. 162–170. [Also available at https://doi.org/10.1016/j.ecolind.2015.01.011.]

Cleveland, D., Brumbaugh, W.G., and MacDonald, D.D., 2017, A comparison of four porewater sampling methods for metal mixtures and dissolved organic carbon and the implications for sediment toxicity evaluations: Environmental Toxicology and Chemistry, v. 36, no. 11, p. 2906–2915. [Also available at https://doi.org/10.1002/etc.3884.]

Davies, A., 2001, The use and limits of various methods of sampling and interpretation of benthic macro-invertebrates: Journal of Limnology, v. 60, 1s, p. 1–6. [Also available at https://doi.org/10.4081/jlimnol.2001.s1.1.]

Haag, W.R., Culp, J.J., McGregor, M.A., Bringolf, R., and Stoeckel, J.A., 2019, Growth and survival of juvenile freshwater mussels in streams—Implications for understanding enigmatic mussel declines: Freshwater Science, v. 38, no. 4, p. 753–770. [Also available at https://doi.org/10.1086/705919.]

Hawkins, C.P., Olson, J.R., and Hill, R.A., 2010, The reference condition—Predicting benchmarks for ecological and water-quality assessments: Journal of the North American Benthological Society, v. 29, no. 1, p. 312–343. [Also available at https://doi.org/10.1899/09-092.1.]

Hughes, R.M., 1995, Defining acceptable biological status by comparing with reference conditions, *in* Davis, W.S., and Simon, T.P., eds., Biological assessment and criteria—Tools for water resource planning and decision making: Boca Raton, Fla., CRC Press, p. 31–47.

Hunt, J.W., Anderson, B.S., Phillips, B.M., Newman, J., Tjeerdema, R.S., Fairey, R., Puckett, H.M., Stephenson, M., Smith, R.W., Wilson, C.J., and Taberski, K.M., 2001, Evaluation and use of sediment toxicity reference sites for statistical comparisons in regional assessments: Environmental Toxicology and Chemistry, v. 20, no. 6, p. 1266–1275. [Also available at https://doi.org/10.1002/etc.5620200615.]

Jarvinen, A.W., and Ankley, G.T., 1999, Linkage of effects to tissue residues—Development of a comprehensive database for aquatic organisms exposed to inorganic and organic chemicals: Pensacola, Fla., Society of Environmental Toxicology and Chemistry, 364 p.

Kaufmann, P.R., Levine, P., Peck, D.V., Robison, E.G., and Seeliger, C., 1999, Quantifying physical habitat in wadeable streams: Corvallis, Oreg., U.S. Environmental Protection Agency, Regional Ecology Branch, Western Ecology Division, National Health and Environmental Effects Research Laboratory, 102 p.

Kraus, J.M., Schmidt, T.S., Walters, D.M., Wanty, R.B., Zuellig, R.E., and Wolf, R.E., 2014, Cross-ecosystem impacts of stream pollution reduce resource and contaminant flux to riparian food webs: Ecological Applications, v. 24, no. 2, p. 235–243. [Also available at https://doi.org/10.1890/13-0252.1.]

Kraus, J.M., Walters, D.M., and Mills, M.A., 2020, Contaminants and ecological subsidies— The land-water interface: Switzerland, Springer, 384 p. [Also available at https://doi.org/10.1007/978-3-030-49480-3.]

MacDonald, D.D., Ingersoll, C.G., and Berger, T.A., 2000, Development and evaluation of consensus-based sediment quality guidelines for freshwater ecosystems: Archives of Environmental Contamination and Toxicology, v. 39, no. 1, p. 20–31. [Also available at https://doi.org/10.1007/s002440010075.]

Magar, V.S., Chadwick, D.B., Bridges, T.S., Fuchsman, P.C., Conder, J.M., Dekker, T.J., Steevens, J.A., Gustavson, K.E., and Mills, M.A., 2009, Technical guide—Monitored natural recovery at contaminated sediment sites: U.S. Department of Defense, Environmental Security Technology Certification Program Project ER-0622, 277 p.

McElroy, A.E., Barron, M.G., Beckvar, N., Driscoll, S.B.K., Meador, J.P., Parkerton, T.F., Preuss, T.G., and Steevens, J.A., 2011, A review of the tissue residue approach for organic and organometallic compounds in aquatic organisms: Integrated Environmental Assessment and Management, v. 7, no. 1, p. 50–74. [Also available at https://doi.org/10.1002/ieam.132.]

Merritt, R.W., Cummins, K.W., Resh, V.H., and Batzer, D.P., 2019a, Sampling aquatic insects—Collection devices, statistical considerations, and rearing procedures, *in* Merritt, R.W., Cummins, K.W., and Berg, M.B., eds., An introduction to aquatic insects of North America (4th ed.): Dubuque, Iowa, Kendall Hunt, p. 143–195.

Merritt, R.W., Cummins, K.W., and Berg, M.B., eds., 2019b, An introduction to aquatic insects of North America (4th ed.): Dubuque, Iowa, Kendall Hunt, 1158 p.

Moulton, S.R., II, Carter, J.L., Grotheer, S.A., Cuffney, T.F., and Short, T.M., 2000, Methods of Analysis by the U.S. Geological Survey National Water Quality Laboratory—Processing, taxonomy, and quality control of benthic macroinvertebrate samples: U.S. Geological Survey Open File Report 00–212, 49 p., accessed April 2022 at https://doi.org/10.3133/ofr00212.

Moulton, S.R., II, Kennen, J.G., Goldstein, R.M., and Hambrook, J.A., 2002, Revised protocols for sampling algal, invertebrate, and fish communities as part of the National Water-Quality Assessment Program: U.S. Geological Survey Open-File Report 02–150, 75 p, accessed April 2022 at https://doi.org/10.3133/ofr2002150.]

Pence, R.A., Cianciolo, T.R., Drover, D.R., McLaughlin, D.L., Soucek, D.J., Timpano, A.J., Zipper, C.E., and Schoenholtz, S.H., 2021, Comparison of benthic macroinvertebrate assessment methods along a salinity gradient in headwater streams: Environmental Monitoring and Assessment, v. 193, no. 765, 16 p.

Pond, G.J., Passmore, M.E., Borsuk, F.A., Reynolds, L., and Rose, C.J., 2008, Downstream effects of mountaintop coal mining—Comparing biological conditions using family- and genus-level macroinvertebrate bioassessment tools: Journal of the North American Benthological Society, v. 27, no. 3, p. 717–737. [Also available at https://doi.org/10.1899/08-015.1.]

Reynoldson, T.B., Norris, R.H., Resh, V.H., Day, K.E., and Rosenberg, D.M., 1997, The reference condition—A comparison of multimetric and multivariate approaches to assess water-quality impairment using benthic macroinvertebrates: Journal of the North American Benthological Society, v. 16, no. 4, p. 833–852. [Also available at https://doi.org/10.2307/1468175.]

Smith, E.P., Orvos, D.R., and Cairns, J., Jr., 1993, Impact assessment using the Before-After-Control-Impact (BACI) model—Concerns and comments: Canadian Journal of Fisheries and Aquatic Sciences, v. 50, no. 3, p. 626–637. [Also available at https://doi.org/10.1139/f93-072.]

Steevens, J.A., Besser, J.M., Dorman, R.A., and Sparks, D.W., 2020, Influence of remediation on sediment toxicity within the Grand Calumet River, Indiana, USA: Chemosphere, v. 249, no. 126056, p. 1–16. [Also available at https://doi.org/10.1016/j.chemosphere.2020.126056.]

Steevens, J.A., Reiss, M.R., and Pawlisz, A.V., 2005, A methodology for deriving tissue residue benchmarks for aquatic biota—A case study for fish exposed to 2,3,7,8-tetrachlorodibenzo-p-dioxin and equivalents: Integrated Environmental Assessment and Management, v. 1, no. 2, p. 142–151. [Also available at https://doi.org/10.1897/IEAM_2004a-014.1.]

Stoddard, J.L., Larsen, D.P., Hawkins, C.P., Johnson, R.K., and Norris, R.H., 2006, Setting expectations for the ecological condition of streams—The concept of reference condition: Ecological Applications, v. 16, no. 4, p. 1267–1276. [Also available at https://doi.org/10.1890/1051-0761(2006)016[1267:SEFTEC]2.0.CO;2.

Stoddard, J.L., Herlihy, A.T., Peck, D.V., Hughes, R.M., Whittier, T.R., and Tarquinio, E., 2008, A process for creating multimetric indices for large-scale aquatic surveys: Journal of the North American Benthological Society, v. 27, no. 4, p. 878–891. [Also available at https://doi.org/10.1899/08-053.1.]

U.S. Environmental Protection Agency, 2000, Methods for measuring the toxicity and bioaccumulation of sediment-associated contaminants with freshwater invertebrates (2d ed.): Duluth, Minn., Office of Research and Development, U.S. Environmental Protection Agency-600-R-99-064; Washington, D.C., Office of Science and Technology, Office of Water, 192 p.

U.S. Environmental Protection Agency, 2002a, Methods for measuring the acute toxicity of effluents and receiving waters to freshwater and marine organisms (5th ed.): Washington, D.C., EPA-821-R-02-012, 30 p

U.S. Environmental Protection Agency, 2002b, Short-term methods for estimating the chronic toxicity of effluents and receiving waters to freshwater organisms (4th ed.): Washington, D.C., U.S. Environmental Protection Agency, EPA-821-R-02-013, 335 p.

U.S. Environmental Protection Agency, 2003, Procedures for the derivation of equilibrium partitioning sediment benchmarks (ESBs) for the protection of benthic organisms—PAH mixtures: Washington, DC, U.S. Environmental Protection Agency-600-R-02-013, Office of Research and Development, 176 p.

U.S. Environmental Protection Agency, 2005, Procedures for the derivation of equilibrium partitioning sediment benchmarks (ESBs) for the protection of benthic organisms—Metal mixtures (cadmium, copper, lead, nickel, silver, and zinc): Washington, D.C., U.S. Environmental Protection Agency, , Office of Research and Development, EPA-600-R-O2-011, 121 p.

U.S. Environmental Protection Agency, 2006, Best Practices for Identifying Reference Condition in Mid-Atlantic Streams: Washington, D.C., U.S. Environmental Protection Agency 260-F-06-002, Office of Environmental Information, 8 p.

U.S. Environmental Protection Agency, 2008, Procedures for the derivation of equilibrium partitioning sediment benchmarks (ESBs) for the protection of benthic organisms—Compendium of tier 2 values for non-ionic organics: Washington, D.C., U.S. Environmental Protection Agency, Office of Research and Development, EPA/600/R-02/016, 76 p.

U.S. Environmental Protection Agency, 2013, National Rivers and Stream Assessment 2013–2014—Field Operations Manual Wadeable (Version 1.0): Washington, D.C., U.S. Environmental Protection Agency-841-B-12-010, Office of Water, Office of Environmental Information, 177 p., accessed April 2022 at https://www.epa.gov/nationalaquatic-resource-surveys/national-rivers-and-streams-assessment-2008-2009-technical-report.

U.S. Environmental Protection Agency, 2016, National Rivers and Streams Assessment 2008-2009 Technical Report: Washington D.C., Office of Water, Office of Research and Development, 140 p., accessed April 2022 at https://www.epa.gov/sites/default/files/2016-03/documents/nrsa_08_09_technical_appendix_03082016.pdf.

Wolman, M.G., 1954, A method of sampling coarse river-bed material: Eos, Transactions American Geophysical Union, v. 35, no.°6, p. 951–956. [Also available at https://doi.org/10.1029/TR035i006p00951.]