Massachusetts Department of ENVIRONMENTAL PROTECTION

t e c h n i c a l u p d a t e

Assessing Risk of Harm to Benthic Invertebrates

Update to: Section 9.4 of *Guidance for Disposal Site Risk Characterization – In Support of the Massachusetts Contingency Plan* (1996)

Introduction:

ORS has adopted the practice of publishing Technical Updates as a way of revising specific parts of the *Guidance for Disposal Site Risk Characterization* (DEP 1996). Technical updates supersede those sections of the Guidance from which they depart. This Technical Update includes guidance on three interrelated topics:

(1) Defining Assessment Endpoints for Benthic Invertebrates;
(2) Measuring the Effects of Sediment Contamination on Benthic Invertebrates; and
(3) Characterizing the Risk of Harm to Benthic Invertebrates from Sediment Contamination.

MassDEP published ecological risk assessment guidelines as Chapter 9 of the *Guidance for Disposal Site Risk Characterization* in 1996. Both the 1996 Guidance and this Technical Update are consistent with EPA guidance, particularly the 1992 Framework for Ecological Risk Assessment (the Framework Document).

Parts One, Two and Three of this Technical Update correspond with the three major components of environmental risk assessment outlined in the 1992 Framework and listed below:

- The problem formulation phase is the planning process that establishes the goals, breadth and focus of the risk assessment. Assessment endpoints are developed in this phase, and the measures of effects that will be used to evaluate the assessment endpoints are selected.
- The analysis phase evaluates environmental exposures to the chemicals of concern and measures the effects of those exposures. Measures of effects are obtained during the analysis phase.
- Risk characterization combines information on exposure and effects to estimate the likelihood, severity, and or significance of adverse effects from contamination.

While this Technical Update remains consistent with the EPA Framework, some of the specific recommendations for analyzing and characterizing the risk of harm to benthic invertebrates differ somewhat from risk assessment practices that have become conventional over the years since the Framework was published.

The guidelines in this Technical Update:

- Express a preference for benthic invertebrate assessment endpoints that are formulated in terms of organism-level effects, rather than community-level effects.

- Recommend using:
    - sediment toxicity testing results to evaluate risk of harm from sediment contamination to benthic invertebrates at most sites, regardless whether or not an organism-level endpoint is used;
    - benthic invertebrate community surveys as qualitative indicators of community conditions at most sites;
    - benthic invertebrate community surveys as a quantitative measures of harm only in cases where the study design and the nature of the site make it more likely than not that adverse effects on the community will be detected where they occur.
- Recommend considering organism-level effects detected by toxicity testing as significant lines of evidence in the risk characterization, regardless of whether an organism-level or community-level assessment endpoint is used.

This Technical Update, like MassSDEP's *Guidance for Disposal Site Risk Characterization*, has been developed solely to improve the MCP risk assessment process. It is not intended for use as guidance under any other regulatory program inside or outside of MassDEP.

Massachusetts Department of Environmental Protection
One Winter Street
Boston, MA 02108-4746

Commonwealth of Massachusetts
Mitt Romney
Governor

Executive Office of Environmental Affairs
Stephen R. Pritchard
Secretary

Department of Environmental Protection
Robert W. Gollege, Jr.
Commissioner

Produced by the Office of Research & Standards, January 2006.

Part One:
Defining Assessment Endpoints for Benthic Invertebrates

Introduction:

Assessment endpoints describe the ecological resources and characteristics that will be evaluated in a risk assessment. U.S. EPA has defined assessment endpoints as "specific entities and their attributes that are at risk and that are expressions of a management goal" (U.S. EPA 2003). In *Guidance for Disposal Site Risk Characterization* (MassDEP 1996), MassDEP defined assessment endpoints as "specific effects that will be evaluated in the risk assessment" and recommended defining an assessment endpoint as an effect on a receptor. Whatever the exact form of assessment endpoint statements, they should represent the entities to be evaluated in the risk assessment and protected by risk management decisions.

Assessment endpoints should be defined so that risk management actions based on the risk assessment will be consistent with management goals. For the purpose of MCP compliance, management goals may be considered broad objectives related to MassDEP's mission and to the statutes, regulations and policies implemented by MassDEP. With respect to ecological risk assessment of surface water bodies done under the MCP, an appropriate management goal statement is: *"to restore and maintain the chemical, physical and biological integrity of the surface water body so that suitable habitat is available for sustaining a native, naturally diverse community of aquatic flora and fauna."* This goal statement is derived from:

- The stated objective of the Clean Water Act, which is "to restore and maintain the chemical, physical and biological integrity of the nation's waters", and
- The Commonwealth's Surface Water Quality regulations, which state: "the aquatic life use is supported when suitable habitat (including water quality) is available for sustaining a native, naturally diverse community of flora and fauna."

Although many waste site risk assessments do not explicitly identify management goals, the goal statement is an important guidepost in the risk assessment/risk management process. The assessment endpoint statement forms the connection between the broad statement of management goals and the specific measurements employed to evaluate risk.

The assessment endpoint statement also conveys the level of biological organization (organism, population, community or ecosystem) under consideration in the risk

assessment. Program management goals, specific regulatory requirements, and spatial scale of the contamination all influence the level of organization represented by the assessment endpoint. The following section recommends the use of an organism-level assessment endpoint. The subsequent section offers an alternative approach, employing a community-level endpoint.

Recommended Assessment Endpoint Statements for Benthic Invertebrates: Organism-level

MassDEP recommends an organism-level assessment endpoint for evaluating the risk of harm to benthic invertebrates from exposure to sediment contamination under the MCP. An example of an organism-level assessment endpoint statement offered in EPA's Generic Ecological Assessment Endpoints (EPA 2003) is: "Survival, Fecundity and Growth of Organisms." For MCP purposes, this may be adapted specifically for invertebrates to "Survival, growth, and reproduction of invertebrates".

Organism-level assessment endpoints for benthic invertebrates are preferred because:

- MassDEP considers the organism-level effects on survival, growth and reproduction that are observed in sediment toxicity tests to be indicative of a risk of harm to the invertebrate community.
- Effects on survival, growth and reproduction can be demonstrated more readily at the organism-level assessment endpoint than at a community-level assessment endpoint.

Alternative Assessment Endpoint Statements for Benthic Invertebrates: Community level

As an alternative to an organism-level assessment endpoint, the risk assessor may opt to use a community-level assessment endpoint for benthic invertebrates. Examples of community-level assessment endpoints offered in EPA guidance (EPA 2003) are "abundance of macroinvertebrate taxa and trophic groups" and "taxa richness". Community-level assessment endpoints are viewed as being more clearly linked to management goals than are organism-level endpoints.

MassDEP is not recommending the use of community-level assessment endpoints in part because community attributes are more difficult to measure directly than organism-level attributes. MassDEP recognizes, however, that some risk assessors prefer community-level assessment endpoints because they are more closely linked to management goals than are organism-level assessment endpoints. Therefore, MassDEP will accept the use of a community level assessment endpoint under the following conditions:

- Community-level measures of effects are obtained from a benthic survey that is sufficiently rigorous to produce valid data and to support a reasonably high level of confidence in the conclusion; and
- Organism-level effects (toxicity test results) are taken into account when characterizing the current and future risk of harm to the benthic community.

The subsequent sections of this Technical Update discuss in more detail the strengths and weaknesses of different measures of effects and the use of different measurements to characterize risk.

Massachusetts Department of Environmental Protection
One Winter Street
Boston, MA 02108-4746

Commonwealth of Massachusetts
Mitt Romney
Governor

Executive Office of Environmental Affairs
Stephen R. Pritchard
Secretary

Department of Environmental Protection
Robert W. Gollege, Jr.
Commissioner

Produced by the Office of Research & Standards, January 2006.

This information is available in alternate format by calling our ADA Coordinator at (617) 574-6872.

Massachusetts Department of
Environmental Protection
One Winter Street
Boston, MA 02108-4746

Commonwealth of
Massachusetts
Mitt Romney
Governor

Executive Office of
Environmental Affairs
Stephen R. Pritchard
Secretary

Department of
Environmental Protection
Robert W. Gollege, Jr.
Commissioner

Produced by the
Office of Research & Standards,
January 2006.

Part Two:
Measuring the Effects of Sediment Contamination on Benthic Invertebrates

Introduction:

Measurements of effects are used to quantify the effects of contamination on the assessment endpoints described in the previous section. These measures are obtained during the analysis phase, and the data are then integrated and evaluated in the risk characterization phase to reach a conclusion about the risk of harm from the contaminants. The practice of ecological risk assessment has evolved to encourage and support the use of multiple measurements of effects to evaluate each assessment endpoint. This practice is reflected in MassDEP's 1996 guidance.

For benthic invertebrates, U.S. EPA has long recommended the use of three different types of effects measurements: (1) determining sediment chemical concentrations and comparing those concentrations to effects-based benchmarks published in the literature; (2) performing sediment toxicity tests in the laboratory or in the field; and (3) conducting benthic community surveys, which involve collecting samples of the resident fauna to identify and count them. This combination of measurements has been referred to in the past as "the triad approach". MassDEP continues to support the use of these measures of effects in combination to characterize the risk of harm to benthic invertebrates.

Recommendations and Discussion:

MassDEP considers the use of the three measurements described in the previous paragraph generally appropriate for evaluating site conditions and the effects of sediment contamination on benthic invertebrates. Ideally, all three types of measurements should be included in each risk assessment. From this set of measurements, the risk assessor can obtain "multiple lines of evidence" related to the risk of harm from sediment contamination to benthic invertebrates. (This is not a change from current guidance.)

Each measurement approach has its own strengths and weaknesses, all of which must be considered when planning a risk assessment and interpreting the measurement data. As detailed below, each contributes a different kind of information about contaminant effects:

Comparison of Site Sediment Concentrations to Effects-based Benchmarks

The comparison of sediment concentrations to sediment benchmarks obtained from the literature should be included in all sediment risk assessments as a measure of effects on benthic invertebrates. Benchmarks provide simple, fast, inexpensive, and generally conservative (precautionary) measures of the potential for harmful effects. The weight given to benchmark exceedances varies widely among risk assessments, but in all cases benchmark comparisons serve as useful points of reference for risk managers.

Sediment benchmarks are generally derived from data bases that are not site specific. Those data bases contain data from different kinds of studies done at locations throughout the United States with different contaminant mixtures in the sediments, and in some cases the derived benchmarks are driven by the combined effects of multiple contaminants. As a consequence, the accuracy of each benchmark is uncertain. Although the level of conservatism embodied in each benchmark varies, benchmark comparisons as measures of risk of harm are widely viewed as "overly protective".

Their conservative nature notwithstanding, benchmarks can provide very useful measures of risk in certain situations:

- Sites with low levels of contamination, where benchmarks provide a straightforward way to demonstrate a condition of "no significant risk";
- Sites with very high levels of contamination where the need for remediation is a foregone conclusion, benchmark comparisons provide an efficient alternative to more complex site-specific risk assessment;
- Sites where more site-specific approaches are not feasible, or where such studies have not produced valid or conclusive results, or where site-specific studies are inconclusive, benchmark comparisons provide a practical alternative to repeating the studies.

Sediment Toxicity Tests

Sediment toxicity tests should be included as a measure of effects in all quantitative site-specific risk assessments that are done to evaluate the effects of sediment contaminants on benthic invertebrates. These tests involve evaluating the toxicity of sediment from the site on laboratory organisms using standard testing protocols, so they are more site-specific than benchmark comparisons. Standard testing protocols impose controls that ensure valid results for the test conditions.

Risk assessors have raised several concerns about how well toxicity tests represent site conditions. These include:

- The disruption of sediment during sample collection, transportation and analysis can alter its character and thus lead to some uncertainty about the extent to which toxicity tests represent actual field conditions. Such changes can increase the availability of contaminants for uptake by organisms, thus increasing the toxicity of the sediment sample.
- The test organisms may be more sensitive than native organisms. Some species in nature can adapt to contamination in the environment, and they may survive at site concentrations that are very toxic to laboratory test organisms.
- Toxicity testing results do not reflect the compensatory mechanisms by which natural populations and communities respond to and compensate for chemical stress.

Although these factors are important considerations in planning and executing the risk assessment, MassDEP does not believe that they generally diminish the value of toxicity testing results. With reference to each of the points above, balancing considerations should be taken into account:

- Contaminated sediment at sites is often subject to dynamic forces, such as waves or currents, and to disruption by human activities. Changes in the physical and chemical nature of sediment can occur in situ as well as in the laboratory.
- Considering the large number of invertebrate species indigenous to many habitats, and the high level of variation in chemical sensitivity among them, it is unlikely that the test species are more sensitive than all, or even most, of them. Further, although decreased sensitivity through adaptation to chemical exposure has been observed in some species at some sites, adaptation is a response to chemical stressors. Populations that have adapted in response to chemical stressors may have a diminished capacity to respond to other stressors. Further, many indigenous species may not experience reduced sensitivity through adaptation.
- Compensatory mechanisms such as density-dependent reproduction rates enable some populations to respond to stresses that might otherwise deplete or extirpate the population. Although such mechanisms maintain population abundance or density, they do not necessarily indicate population health or resilience. Like adaptation, density dependence is a

Massachusetts Department of
Environmental Protection
One Winter Street
Boston, MA 02108-4746

Commonwealth of
Massachusetts
Mitt Romney
Governor

Executive Office of
Environmental Affairs
Stephen R. Pritchard
Secretary

Department of
Environmental Protection
Robert W. Gollege, Jr.
Commissioner

Produced by the
Office of Research & Standards,
January 2006.

This information is available in
alternate format by calling our
ADA Coordinator at
(617) 574-6872.

response to stress, and is not necessarily an indication of "no significant risk or harm".

Overall, MassDEP believes the advantages of toxicity tests outweigh their disadvantages. Part Three of this Update presents additional discussion of this point. In general, MassDEP considers sediment toxicity tests to be a key measure of the risk of harm to benthic invertebrates for most sites.

Benthic community surveys should be included in most risk assessments. These studies involve collecting sediment samples and counting and identifying the invertebrate fauna. At a minimum, a benthic survey serves as a qualitative component of the baseline site description. This type of study provides a snapshot of the benthic invertebrate community at a specific location at a specific time.

Many risk assessors consider biological surveys to be the most reliable measure of effects on benthic organisms because the data are based on direct observation of the receptors of concern. The high level of spatial and temporal variability seen in benthic invertebrate communities, however, makes it difficult to discern effects without a large sample size for each location evaluated. MassDEP plans to publish (at a later date under separate cover) guidelines for benthic invertebrate community sampling that will address the extent and intensity of sampling needed to obtain quantitative results that can be given substantial weight in the risk characterization.

For many sites, MassDEP believes that the use of all three types of measurements will result in stronger, more conclusive risk assessments. For small sites, however, benthic community surveys may not be feasible, and may therefore be omitted.

To reach a conclusion about risk from the results of these different measures of effects, risk assessors usually apply a Weight-of-Evidence approach outlined in MassDEP's Guidance for Disposal Site Risk Characterization. In MassDEP's view, this approach, as it has been applied in MCP risk assessments, has not adequately accounted for the strengths and weaknesses of each measurement. Changes in risk characterization guidelines are discussed in more detail in the section that follows.

Part Three:
Characterizing the Risk of Harm to Benthic Invertebrates

Introduction:

In the risk characterization step, the results obtained in the analysis phase are evaluated to determine whether they support a conclusion of "no significant risk" for each endpoint. When multiple measures of effects are obtained for an assessment endpoint, the results of the measurements are used in combination to estimate the risk of harm for the endpoint receptors and attributes. Current practice has been to consider the results of benchmark comparisons, toxicity testing, and field biological surveys together in a Weight-of-Evidence evaluation, as outlined in MassDEP's *Guidance for Disposal Site Risk Characterization*. This approach involves assigning each measurement a high, medium or low weight based on various attributes, and considering the weight of each measurement in conjunction with its results to make determine whether the "weight of evidence" demonstrates a condition of "no significant risk"

When the "weight-of-evidence" approach is used, the relative weights placed on the measurement results often do not accurately reflect the strengths and weaknesses of each measurement. For example, benthic community survey results are typically given

Massachusetts Department of Environmental Protection
One Winter Street
Boston, MA 02108-4746

Commonwealth of Massachusetts
Mitt Romney
Governor

Executive Office of Environmental Affairs
Stephen R. Pritchard
Secretary

Department of Environmental Protection
Robert W. Gollege, Jr.
Commissioner

Produced by the Office of Research & Standards, January 2006.

This information is available in alternate format by calling our ADA Coordinator at (617) 574-6872.

the highest weight because they are thought to represent direct measures of community health. Benthic community survey study designs for MCP risk assessments are seldom, however, sufficiently robust to warrant a high level of confidence in the results.

This section summarizes the main characteristics of each of the effects measures listed above and how they should be prioritized when evaluating the "weight of evidence." As described below, MassDEP recommends considering toxicity test results as the primary measure of effects on invertebrates in most ecological risk assessments.

Recommendations and Discussion:

Comparison of Site Sediment Concentrations to Benchmarks

Benchmarks are generally considered the most precautionary of the three measures of effects used to evaluate risk of harm to benthic invertebrates. Many benchmarks are based at least in part on data from contaminated sediment that contains mixtures of chemicals. Adverse effects observed in such studies are associated with multiple chemicals, and data analysis cannot attribute effects to individual chemical concentrations with a high level of confidence. While sediment concentrations below a benchmark demonstrate a condition of "no significant risk" with a fairly high level of certainty, concentrations exceeding a benchmark are not considered a generally reliable indication of harm or risk of harm.

ORS recommends using benchmark comparisons for qualitative site characterization at most sites, not as quantitative measures of harm. These studies should be used as quantitative measures of effects only in exceptional cases, for example, when:

- Toxicity testing is undertaken, but the test fails (for example, as shown by high mortality in the laboratory control sample); or
- Contaminant levels are low enough so that comparing them with conservative benchmarks will demonstrate a condition of "no significant risk" of harm from sediment contamination; or
- Contamination is so severe that the need for remediation is clear, and the risk assessment will not affect the site management decision.

In general, MassDEP recommends placing lower weight on benchmark comparisons than on toxicity testing data. Benchmarks from the literature do not reflect site-specific factors that may be important in limiting exposure and uptake. They do provide simple, generally conservative measures of the potential for harmful effects. At small sites, sites where other measures are weak or unsuccessful, or sites where the need for remediation is a foregone conclusion, benchmark comparisons provide an alternative to a site-specific risk assessment. For sites where toxicity tests or benthic community survey data is available, benchmark exceedances should generally be given a relatively low weight in the risk characterization.

Toxicity Testing

Toxicity tests involve exposure of test organisms to sediment from the site of concern, and so are more site-specific than benchmarks. As a measurement, toxicity testing has a number of advantages over benthic community surveys:

- The use of laboratory organisms introduces a degree of standardization into these measurements of effects. The use of standard test organisms minimizes variability within each test, enhancing the power of the test to detect effects. This contrasts with

Massachusetts Department of Environmental Protection
One Winter Street
Boston, MA 02108-4746

Commonwealth of Massachusetts
Mitt Romney
Governor

Executive Office of Environmental Affairs
Stephen R. Pritchard
Secretary

Department of Environmental Protection
Robert W. Gollege, Jr.
Commissioner

Produced by the Office of Research & Standards, January 2006.

This information is available in alternate format by calling our ADA Coordinator at (617) 574-6872.

benthic community survey results, which are often highly variable among sub-samples.

- The use of standard laboratory organisms increases the consistency among tests of separate samples and narrows the possible reasons for toxic effects observed in toxicity tests. Differences among test organisms do not contribute significantly to measured differences in survival, growth, or reproduction. This contrasts with benthic surveys, in which some of the observed variability in community metrics may be due to differences in the viability of the organisms at different locations to begin with.

- Between site samples and reference samples, sediment chemistry is the only variable likely to affect the survival and growth of the test organisms. The identity and source of the chemical or chemicals causing toxicity may not always be clear, but it is reasonable to conclude that chemicals in the sediment, whatever their source, are responsible for toxic effects. Natural differences in sediment characteristics, organic matter content for example, may mediate exposure and toxicity, but such differences do not cause toxic effects. This contrasts with benthic community metrics, which in some cases may be affected even more strongly by physical differences in sediment than by chemical contaminant levels.

- Exposure concentrations are more controlled in toxicity tests than in benthic community surveys. In toxicity tests, exposure concentration is measured from the homogenized sediment sample to which the test organisms are exposed. In contrast, the exposure concentration corresponding to benthic community metrics are based on sediment samples co-located with the invertebrate sample. Since sediment concentrations can vary widely over a small scale, the concentration of the co-located sample may not accurately represent the concentrations to which the invertebrates are exposed.

Another advantage of sediment toxicity testing over benthic invertebrate community surveys is the option of collecting and testing sediment samples from depths below the thin biologically active zone at the sediment surface. This can enable the risk assessor to evaluate potential risk in the event that the deeper sediments become mobilized in the future, for example during a storm event.

MassDEP considers toxicity testing to be a relatively reliable measure of the risk of harm. The results of a successful, valid sediment toxicity test should be given a relatively high weight in characterizing the risk of harm to benthic invertebrates at waste sites.

Benthic Community Surveys

Benthic community surveys involve collecting samples of contaminated sediment from the site, counting and identifying the indigenous invertebrates and comparing the results to the same metrics for sediment from a reference area. Survey metrics are related to abundance and/or diversity of the benthic community. In contrast to effects observed in toxicity tests, apparent effects observed through a benthic community survey can be caused by numerous environmental factors other than chemical contamination, including sediment grain size, temperature, and surface water flow rate. These confounding factors can make it difficult or impossible to detect small but significant effects on benthic invertebrate abundance and diversity without collecting a large number of samples. In fact, it is not uncommon for the benthic community

Massachusetts Department of
Environmental Protection
One Winter Street
Boston, MA 02108-4746

Commonwealth of
Massachusetts
Mitt Romney
Governor

Executive Office of
Environmental Affairs
Stephen R. Pritchard
Secretary

Department of
Environmental Protection
Robert W. Gollege, Jr.
Commissioner

Produced by the
Office of Research & Standards,
January 2006.

This information is available in
alternate format by calling our
ADA Coordinator at
(617) 574-6872.

associated with samples from the reference area to appear more impaired than that associated with samples from the site.

Due to the effects of confounding factors and the high variability typical of benthic survey metrics, detecting adverse effects of contaminants on the benthic community requires a larger sample than is typically collected for MCP risk assessments. If the sampling effort is sufficiently intensive to discern adverse effects with a reasonable level of confidence, then the risk assessment may include benthic community metrics as quantitative measures of effects under current conditions. If such a sampling effort is not feasible, the benthic community survey results should be considered as a qualitative component of the site assessment.

Considering the strengths and weaknesses of each of the measures detailed above, MassDEP recommends considering toxicity test results as the primary measure of effects on invertebrates in most ecological risk assessments. As long as the toxicity test is valid according to testing guidelines, the results should carry more weight than benchmark comparisons, and, in most cases, more weight than benthic invertebrate field surveys.

These recommendations are consistent with the weight-of evidence guidelines (MassDEP 1996) as ORS would apply them. The weight-of-evidence terminology is not used extensively in this Technical Update because the guidelines are often interpreted and applied in a different ways by different risk assessors, and the use of the terminology associated with the guidelines can be ambiguous.

Limitations

The risk assessment practices discussed in this Update address themselves only to effects under present site conditions. They do not necessarily evaluate the potential impacts of contaminants in deeper sediment that may be released and/or transported in the future, for example by turbulence associated with a storm event. The risk assessment/risk management process must address foreseeable risk under future conditions.

Questions about this document should be directed to:

Nancy Bettinger at (617)556-1159 or nancy.bettinger@state.ma.us
Thomas Angus at (617)292-5513 or thomas.angus@state.ma.us

References:

EPA 2003. *Generic Ecological Assessment Endpoints (GEAEs) for Ecological Risk Assessment*. Risk Assessment Forum, U.S. Environmental Protection Agency, Washington D.C. EPA/30/P-02/004F. October 2003.

MASSDEP 1996. *Guidance for Disposal Site Risk Characterization in Support of the Massachusetts Contingency Plan*. Chapter 9, Method 3 Environmental Risk Characterization.

Massachusetts Department of Environmental Protection
One Winter Street
Boston, MA 02108-4746

Commonwealth of Massachusetts
Mitt Romney
Governor

Executive Office of Environmental Affairs
Stephen R. Pritchard
Secretary

Department of Environmental Protection
Robert W. Gollege, Jr.
Commissioner

Produced by the Office of Research & Standards, January 2006.

This information is available in alternate format by calling our ADA Coordinator at (617) 574-6872.