ORNL/TM-13615

A GUIDE TO THE ORNL ECOTOXICOLOGICAL SCREENING BENCHMARKS: BACKGROUND, DEVELOPMENT, AND APPLICATION

Bradley E. Sample
Glenn W. Suter II
Rebecca A. Efroymson
Daniel S. Jones

Environmental Sciences Division
Publication No. 4783

May 1998
Revision 1.0

Prepared for the
U.S. Department of Energy
Office of Environmental Policy and Assistance
Air, Water, and Radiation Division

Prepared by the
OAK RIDGE NATIONAL LABORATORY
Oak Ridge, Tennessee 37831-6285
managed by
LOCKHEED MARTIN ENERGY RESEARCH CORP.
for the
U.S. DEPARTMENT OF ENERGY
under contract DE-AC05-96OR22464

ACKNOWLEDGMENTS

Funding for development of this guide was provided by the U.S. Department of Energy, Office of Environmental Policy and Assistance, Air, Water, and Radiation Division (EH-412; Stephen Domotor, Technical Project Manager). We are grateful for the multiple respondents to our request for information concerning field examples of use and acceptance of the ORNL benchmarks. This guide benefitted from reviews by Stephen Domotor (EH-412), John Bascietto (EH-413), Linda Meyers-Schöne (IT Corp., Albuquerque, NM), Elizabeth Oms (DOE Albuquerque Operations Office), and technical staff from DOE's Environmental Management Program (EM) Center for Risk Excellence.

PREFACE

The Department of Energy's (DOE) Office of Environmental Policy and Assistance (EH-41) develops environmental protection policies and guidance in response to current and emerging environmental requirements. Within these responsibilities, EH-41 provides tools and assistance on environmental risk assessment, to include ecological risk, for use within DOE Program and Operations Offices.

The development of this Guide was sponsored by EH-41 as part of it's technical assistance role in response to questions from DOE Operations Offices regarding the technical basis, appropriate application, and acceptance of ecological screening benchmarks in general and the Oak Ridge National Laboratory (ORNL) benchmarks in particular. The ORNL ecotoxicological screening benchmarks represent an assemblage of ORNL-derived screening values and screening values compiled from other sources which can be used as a screening tool in ecological risk assessments. Accordingly, EH-41 sponsored the development of this Guide to provide DOE program managers and contractors with a better understanding of what the ORNL benchmarks are, describe their basis for development, and discuss their appropriate application, strengths and weaknesses, and regulatory acceptance.

Used as recommended, ecological screening benchmarks can generally serve to identify contaminants, media, and receptors that may be at risk and that may require further investigation. Although ecological screening benchmarks have been successfully applied and accepted in many ecological risk assessment projects, EH-41 does not explicitly endorse the use of any one set of screening values (e.g., ORNL-derived compared with values from other sources). Screening benchmarks, to include ORNL-derived benchmarks, are not regulatory criteria. It is the responsibility of the user to obtain approval for the use of screening benchmarks at each site. Regulatory approval for the use of any set of screening benchmarks should be obtained as part of the Data Quality Objectives (DQO) process at the outset of any project.

TABLE OF CONTENTS

TABLES

1. OVERVIEW

Screening tools to identify chemical concentrations in environmental media that are at or below thresholds for effects are needed to evaluate the potential risks that chemical contaminants may present to ecological receptors. To facilitate the evaluation of chemical contamination on the Oak Ridge Reservation and other U.S. Department of Energy (DOE) facilities, and to remedy the limited availability or absence of approved values from regulatory agencies, the Environmental Sciences Division of Oak Ridge National Laboratory (ORNL) developed a comprehensive set of ecotoxicological screening benchmarks. The ORNL ecotoxicological screening benchmarks provide a comprehensive assembly of screening values developed by researchers at ORNL together with values developed by regulatory agencies. Since their development and dissemination, the ORNL benchmarks have become an important and sometimes controversial tool for ecological risk assessment. It should be noted that site-specific screening values exist for other sites (e.g., DOE's Rocky Flats Environmental Technology Site). These values should be considered if appropriate.

The purpose of this guide is to provide DOE program managers and contractors with a better understanding of what the ORNL benchmarks are, describe the basis for their development, and discuss their appropriate application, strengths and weaknesses, and regulatory acceptance. The first section of the guide answers general questions that are frequently asked concerning the ORNL benchmarks. The remaining four sections provide detailed descriptions of the background, development, quality assurance/quality control (QA/QC), and application and interpretation of each 'class' of benchmarks developed for specific endpoint groups. These sections are intended to address general concerns that have been raised in relation to the development, data quality, and application of specific classes of benchmarks.

1.1 What are the ORNL ecotoxicological screening benchmarks?

The ORNL ecotoxicological screening benchmarks are concentrations of chemicals in ambient media that are believed to represent acceptable concentrations with respect to selected ecological receptors. In general, if the benchmark concentrations are not exceeded, further analysis of that contaminant and ecological receptor is not warranted. The ORNL benchmarks provide a set of consistent ecotoxicological values that have been developed by experts and reviewed by users, regulators, and experts both inside and outside ORNL. Their use saves considerable time and effort that would otherwise be required to develop equivalent values for each site.

These benchmarks are presented in a series of technical manuscripts (TMs) published by ORNL. These TMs may be obtained from the ORNL Ecological Risk Assessment web site (**http://www.hsrd.ornl.gov/ecorisk/ecorisk.html)**, from the DOE Office of Environmental Policy and Assistance, Dose and Risk Assessment web site (http://tis-nt.eh.doe.gov/oepa/risk/), and through the National Technical Information Service (NTIS; U.S. Department of Commerce, 5285 Port Royal Road, Springfield, VA 22161).

While some benchmark values have been derived by ORNL staff, others have been obtained from the U.S. Environmental Protection Agency (EPA) (e.g., national ambient water quality criteria [NAWQC]) and other regulatory agencies (e.g., Ontario Ministry of

Environment sediment quality values). Benchmarks have been developed or obtained for the following types of exposure and classes of endpoint groups: exposure of aquatic biota to chemicals in water (Sec. 2); exposure of benthic biota to chemicals in sediments (Sec. 3); exposure of terrestrial plants to chemicals in soil (Sec. 4); exposure of soil invertebrates to chemicals in soil (Sec. 4); exposure of soil functional groups to chemicals in soil (Sec. 4); and exposure of wildlife to chemicals in orally ingested materials (Sec. 5). The benchmarks were derived using various methods that have been implemented by regulatory agencies, were recommended by regulatory agencies, or were consistent with past regulatory practice. When multiple methods have been implemented or recommended for benchmark development, benchmarks were derived using all available methods and included in the set.

1.2 What are appropriate and inappropriate applications of the ORNL benchmarks?

The intended purpose of the ORNL ecotoxicological benchmarks is to screen chemical concentrations in environmental media to identify chemicals that are present at a sufficiently high concentration to present a potential risk to ecological receptors. These chemicals, termed chemicals of potential ecological concern (COPECs), require further assessment to determine whether they do in fact pose significant risks. This assessment process is discussed in more detail below (Sec. 1.5).

The ORNL ecotoxicological benchmarks should not be used as remedial goals or as concentrations that pose significant risks. Rather, risk estimates for the COPECs and remedial goals for the chemicals of ecological concern (COECs - those chemicals that are estimated to pose a significant risk) should be based on a weighing of the available evidence, including any site-specific toxicity tests or biological surveys. Even if the risk estimates and remedial goals are based solely on conventional single-chemical toxicity data, the ORNL benchmarks should not be automatically adopted as thresholds for significant risk. Rather, the risk assessors should return to the primary literature to determine what studies and data are most relevant to their site and assessment endpoints.

1.3 What weaknesses and uncertainties are associated with the ORNL benchmarks?

The primary weakness of any set of generic ecotoxicological benchmarks is that there is lack of consensus on what ecological entities and properties should be protected or what level of protection should be afforded. While the EPA provides broad guidelines for the selection of assessment endpoints (EPA 1996a), they do not identify specific endpoints that are to be protected. It is therefore not possible to confidently state that any benchmark is sufficiently protective, but not overprotective, when we do not know what is to be protected and what level of effect is considered significant. Hence, the benchmarks can not be validated for all sites and situations. They can be defended only in terms of regulatory precedent.

As with the development of any effects metric used in risk analysis, the development of benchmarks requires using models to extrapolate from laboratory toxicity test endpoints to the assessment endpoints. There is no consensus among the regulatory community as to how best to perform this extrapolation for screening benchmarks. The choice of extrapolation models is based on current regulatory practice, and, when multiple types of extrapolation models have been employed, multiple benchmarks have been included.

Because of the problem listed in the prior paragraph, it is not possible to determine what extrapolation method leads to the best benchmarks.

Like any other approach to risk assessment that uses conventional, single-chemical toxicology, the benchmarks do not explicitly incorporate the combined toxic effects resulting from exposure to multiple chemicals. If multiple chemicals occur at a site at concentrations that are approaching benchmark concentrations, consideration of combined effects could lead to declaring a chemical, medium, or site to be potentially hazardous. In such cases, a default assumption, such as concentration additivity of effects, can be employed as in human health risk assessments. It is necessary in such cases to define a lower limit for significant contribution of a chemical to the combined toxicity, however. For example, one might screen out all chemicals contributing less than one percent of the estimated total toxicity.

Because the benchmarks are based on conventional laboratory toxicity data, their relationship to field exposures is uncertain. In particular, laboratory tests are generally designed to expose organisms to highly bioavailable and toxic forms of the test chemical, under conditions that maximize exposure. For example, aquatic toxicity tests are performed in filtered water with soluble forms of the test chemicals. Similarly, soil toxicity tests are usually performed with soluble forms of chemicals or chemicals with carriers, and the tests are performed without aging the soil to allow for sequestration of the chemical. In general, these considerations mean that the benchmarks are conservative, because the bioavailability and toxicity of the chemical in the field is generally lower than that in the laboratory. However, in a few cases, such as addition of metals to an acidic site water, the laboratory data may underestimate toxicity, making the benchmarks under-protective. In these cases, site-specific toxicity tests should be performed.

1.4 What is their relationship to other ecological risk guidance and tools developed by ORNL or others?

The principal use of the benchmarks is in screening assessments. The 1995 ORNL guidance document, *Guide for performing screening ecological risk assessments at DOE facilities,* (Suter 1995) explains how the benchmarks should be used in conjunction with background concentrations and waste inventories to identify COPECs. The 1996 ORNL guidance document, *Risk characterization for ecological risk assessment of contaminated sites* (Suter 1996a) explains the limited utility of the benchmarks in risk estimation for definitive assessments. In addition, the EPA's interim final guidance for ecological risk assessment for Superfund recommends the use of screening benchmarks and references the ORNL benchmarks as examples (EPA 1997).

The benchmarks were also used as a starting point for the development of Preliminary Remedial Goals (PRGs), which are concentrations of chemicals in specific ambient media that, in the absence of better evidence, are presumed to be thresholds for potential remedial actions. PRGs for most media are less conservative than benchmarks and are specific to a chemical and medium, not to a receptor, chemical, and medium. The current ORNL PRGs were published in 1997 as, *Preliminary remedial goals for ecological endpoints* (LMES et al. 1997).

1.5 How are ORNL benchmarks applied in the RI/FS process?

Screening Assessments are performed at two stages of the Remedial Investigation/Feasibility Study (RI/FS) process. During the planning of the definitive ecological risk assessment, existing data concerning contaminant concentrations are screened to determine which, if any, chemicals or media may be eliminated from further sampling and analysis. During the performance of the definitive ecological risk assessment, chemical concentrations are screened as a means of determining which chemicals should be subject to analysis in the definitive assessment. In all screening assessments, conservative estimates of the chemical concentrations to which organisms are exposed are compared to benchmark values and to background concentrations. The concentration in a medium is compared to the benchmark by calculating a hazard quotient (HQ), where HQ = ambient concentration/benchmark concentration. Chemicals that exceed relevant benchmark values (i.e., HQ>1) and relevant background concentrations are COPECs. If the waste is well-characterized, chemicals that are not part of the waste inventory may also be eliminated from the list of site-related COPECs. A chemical in soil is retained if it is not screened out with respect to any of the receptors exposed to soil. For aquatic biota and sediment biota, each of which have multiple benchmarks (Sec. 2), the chemical is retained if any benchmark is exceeded.

Definitive Assessments are performed following screening assessments to complete the baseline ecological risk assessment for the RI. They should be based on a weighing of all available lines of evidence including biological surveys, toxicity tests performed with contaminated media, and conventional single-chemical toxicity tests (EPA 1996a; Suter 1996a; EPA 1997). The role of the benchmarks in this process depends on the availability of other data to support the inference. For example, if high quality biological survey or media toxicity data are available for the site, single-chemical data may be used simply to confirm the reasonableness of those results. For that purpose, the benchmarks may be adequate. However, if more realistic and site-specific effects data are not available or give ambiguous results, it is necessary to use single-chemical toxicity data to make the best possible estimate of risks. Simply using the benchmarks is not adequate for that purpose. Rather, the assessor should return to the literature from which the benchmarks were derived and reanalyze it in a way that is relevant to the specific assessment endpoint and medium being considered. In some cases this will lead to the conclusion that one of the benchmarks provides the best estimate of the effects on the endpoint entity, property, and effects level. More often, other studies or other analyses of the data will be more appropriate. For example, although there are several alternative benchmarks for aquatic biota, definitive assessments performed at ORNL use empirical distributions of fish and invertebrate species sensitivities as the measures of effects for aquatic biota.

Remedial Goal Options are concentrations in specific media that are suggested by the risk assessors to the risk managers as possible targets for site cleanup. They should be based on the results of the definitive assessment. In particular, they should be based on the line of evidence that best represents the relationship between the occurrence of unacceptable effects on the ecological assessment endpoints and concentrations of contaminants. Therefore, they should not simply be benchmark values. For example, if media toxicity tests provide the best evidence, the remedial goal option might be the highest concentration of the primary contaminant of concern in soil which causes minimal toxicity to plants.

Examples of ecological risk assessments that used the ORNL benchmarks are available to the public on the ORNL Ecological Risk Assessment www site and to DOE employees and contractors through the OSTI Info-bridge (https://apollo.osti.gov/dds/).

1.6 When should site-specific values be developed?

If the conventional toxicity data that are available in the literature and have been used to derive the benchmarks are not applicable to a site (e.g., the site has unusual water chemistry), it is best to proceed to perform the site-specific studies needed to estimate risks for the site. Developing site-specific screening benchmarks under those circumstances would not be cost-effective. The exception, would be cases in which simple recalculation of benchmarks makes them more relevant to the site or to the concerns of regulators or stakeholders. For example, benchmarks for metals in water are normalized to 100 mg/kg hardness, which is conservative for Oak Ridge sites. Benchmarks for sites with softer water should be normalized to a hardness that is near the lower limit of observed values for the site, as explained in the aquatic benchmarks document. Another example would be cases in which local regulators or stakeholders believe that benchmarks based on EPA regulatory practice are not sufficiently conservative. In such cases, a safety factor might be applied to lower the benchmarks. This would result in more chemicals being retained by the screen as COPECs.

1.7 How do the benchmarks relate to hazard assessment?

Some confusion has occurred concerning the distinction between risk assessment and hazard assessment. Most commonly, hazard assessment refers to an approach for the assessment of new chemicals that involves iteration of testing and assessment until enough information is gathered to clearly state that the intended use of the chemical will or will not constitute an unacceptable hazard (Cairns et al. 1979). That assessment approach is less flexible and less powerful than risk assessment and is not applicable to contaminated sites (Suter 1990).

1.8 Are there examples of field application and acceptance of the ORNL benchmarks?

The EPA's interim final guidance for ecological risk assessment for Superfund recommends the use of screening benchmarks and references the ORNL benchmarks as examples (EPA 1997). The ORNL benchmarks served as a basis for EPA's threshold values (a synonym for screening benchmarks) for sediment and aquatic biota, and ORNL benchmarks were adopted when an EPA-derived value was not available (EPA 1996b). There are many examples of field applications and regulatory acceptance of the ORNL benchmarks. Examples of such applications and acceptance are presented in Table 1. Despite being based on practices employed by regulatory agencies and being accepted by various regulatory agencies, the ORNL ecotoxicological benchmarks are not regulatory criteria. Regulatory approval for the use of the ORNL benchmarks or other screening values should be obtained as part of a Data Quality Objectives (DQO) process (Bilyard et al. 1997) at the outset of any project.

There are at least two known instances where use of the ORNL benchmarks has not been fully supported by responsible regulatory agencies. First, NOAA reports that they do not allow use of ORNL benchmarks in natural resource damage assessments, because they had been used blindly as indicative of the occurrence or absence of natural resource injuries (Mike Buchman, NOAA, pers. comm.) It should be noted that this application is not consistent with the intended use of the ORNL benchmarks. Second, the State of New Mexico has not supported the use of the ORNL benchmarks at DOE sites in that State without accessing the primary data sources and verifying that the data are appropriate for site-specific use (Roger Ferenbaugh, pers. comm.).

Table 1. Selected examples of the field application of the ORNL benchmarks and their acceptance by regulators in various areas. This table is not meant to represent a comprehensive listing of the application of the ORNL benchmarks.

Agency	Application	Benchmarks used	Source of Information	Comments
U.S. EPA Regions 4 & 5, Tennessee Dept. Environment & Conservation, Ohio EPA, Kentucky Dept. Environmental Protection	Screening Assessments (>15) of DOE sites.	All	ORNL experience	
U.S. EPA Region 1	Screening assessment. Examples given for Assessments of Hanscom Air Force Base, MA	All	Chris McCarthy of CH2M HILL	
U.S. EPA Region 3	Primarily for screening. Example given for Assessment of Philadelphia Naval Yard	All	Dan Hinckley of EA Environmental Engineering, Science, and Technology	Biological Technical Assistance Group (BTAG) comments indicate that ORNL benchmarks are generally employed because they are the lowest values available.
U.S. EPA Region 9 California EPA Department of Toxic Substances Control.	Comprehensive Ecological Risk Assessment for the Mather AFB, Sacramento, CA	Wildlife, plant, aquatic biota, and sediment biota	Linda Meyers-Schone of IT Corp	Regulators accepted final document. Benchmarks were used in screening assessment.
U.S. DOE	Screening ecological risk assessment for Columbia River adjacent to the Hanford site.	Wildlife and plant benchmarks	PNNL 1998	

Table 1. (Cont.)

Agency	Application	Benchmarks used	Source of Information	Comments
California EPA Office of Environmental Health Hazard Assessment	Emergency responses for toxic spills	Wildlife	Julie Yamamoto of Cal EPA	ORNL benchmarks are one of the only wildlife resources available to develop fast, ballpark predictions of potential toxic effects
Central Valley Regional Water Quality Control Board, California	Screening Assessment of landfill site	Aquatic benchmarks	Mark Stelljes of SECOR International, Inc.	
Oregon Dept. of Environ. Quality	Screening only	Use of wildlife, plant, and certain aquatic benchmarks is allowed	Bruce K. Hope of Oregon Dept of Environ. Quality	Use of ORNL benchmarks is recommended in Oregon risk guidance documents because they are available. There are concerns about applicability, however
New York State Dept. of Environmental Conservation	Screening assessments, primarily. Example given of assessments from Brookhaven National Lab	Soil invertebrates, plants and wildlife	Paul Carella of NY DEC	Use NY Water quality standards when they exist and use sediment screening criteria developed by NY Division of Fish and Wildlife
Texas Natural Resources Conservation Commission	Screening only	Plant and soil invertebrate	Draft ecological risk assessment guidance (TNRCC 1996)	

2. BENCHMARKS for AQUATIC BIOTA

2.1 Background

The aquatic benchmarks were developed to provide values that would protect freshwater biotic communities from contaminants in water. They are different from the other benchmark sets because of the large body of data that is available for aquatic toxicity, the existence of relevant national criteria, and the large amount of research that has been done concerning the extrapolation of aquatic toxicity data. As a result, there are currently 17 different types of aquatic benchmarks included in the ORNL set (Suter and Tsao 1996). They are:

- Acute and chronic national ambient water quality criteria - these criteria are applicable regulatory standards.

- Secondary acute and chronic values - these are conservative estimates of water quality criteria for those chemicals for which available data are insufficient to derive criteria.

- Lowest chronic values for fish, daphnids, non-daphnid invertebrates, aquatic plants, and all organisms - these are the lowest acceptable chronic values (CVs) for each of the listed taxa; lowest CVs have been used in place of chronic criteria by the EPA.

- Acute and chronic OSWER threshold values - these are the EPA Office of Solid Waste and Emergency Response's screening benchmark, which are either criteria or secondary values.

- Acute and chronic Region IV values - these benchmarks, derived by the EPA's southeastern region, are criteria or test endpoints divided by 10.

- Lowest test EC20 (20% effects concentration) values for fish and daphnids - these are the highest tested concentration not causing a reduction of as much as 20% in the reproductive output of female test organisms.

- Sensitive species EC20 - these benchmarks were derived like chronic criteria except that the lowest EC20 for the chemical was used in place of the lowest CV.

- Fish population EC25 (25% effects concentration) - these are estimates of the concentration causing a 25% reduction in the recruit abundance of a population of largemouth bass.

All of the benchmarks listed under the first five bullets were derived by the EPA (water quality criteria and threshold values) or an EPA regional office (Region IV values) or have been recommended by the EPA as substitutes for water quality criteria (lowest chronic values and tier II values). Those listed under the last three bullets were derived at ORNL to determine the influence on benchmark values of alternative test endpoints or alternative extrapolation models. The benchmarks were compared and published in a refereed journal (Suter 1996b). The lowest benchmarks were those that included daphnids and those that included a safety factor (i.e., the secondary

2.2 Development history

The aquatic benchmarks were the first set developed at ORNL (Suter et al. 1992). They have been updated at two year intervals since then.

2.3 Selection of studies and other QA/QC

Quality assurance for the aquatic benchmarks included the following:

- To assure completeness, data are identified in searches of AQUIRE (AQUatic toxicity Information REtrieval), EPA water quality criteria documents, bibliographic data bases, and by hand searching of the literature.

- Data are obtained from original sources. Secondary sources were used only for bibliographic purposes.

- Data are acceptable if they meet the requirements for data used to derive the National Ambient Water Quality Criteria.

- In those cases for which there are no data that meet the NAWQC useability criteria, other data may be used if the data are judged to be a) otherwise high quality although not standard, or b) otherwise high quality although the test species is not from the waters of the U.S. In such cases, the deviation is clearly identified in the report.

- Calculation of the Tier II values were performed in a spreadsheet and the formulas were confirmed by hand calculations.

- Data and calculations were checked by a scientist who is experienced in ecotoxicology and familiar with the benchmark derivation methods.

- All data used and calculations performed have been documented in reports that are publicly available. Comments from users have served to identify questionable values in prior versions of the benchmarks.

2.4 Use and interpretation of results

Because of the large number of potential aquatic benchmarks, it is necessary to decide which ones to use in a particular assessment. At ORNL we use the entire set in order to obtain the maximum information from the screening assessment. For example, if only the benchmarks that incorporate safety factors are exceeded, we know that the chemical is a marginal hazard relative to a chemical that exceeds all or most benchmarks. Similarly, the pattern of exceedence can indicate which taxa are most likely to be affected. If, for example, the highest quotient is for the aquatic plant CV, site studies might include periphyton. The chemical is considered a COPEC if the 95% upper confidence limit on the mean concentration exceeds any benchmark.

3. BENCHMARKS for SEDIMENT-ASSOCIATED BIOTA

3.1 Background

The sediment benchmarks are chemical concentrations in whole sediment that are associated to varying degrees with adverse effects on benthic organisms. Unlike many of the other benchmark sets, the sediment benchmarks are comprised primarily of values previously calculated by other organizations and presented in government reports or the scientific literature. There are currently 18 types of sediment benchmarks included in the ORNL set (Jones et al. 1997b). They are described in Table 2 and fall into the following five general classes:

- Integrative benchmarks are derived from the distribution of concentrations observed to be toxic to benthic organisms. The toxicity values are obtained through a variety of approaches, including benthic community surveys of contaminated sites and toxicity tests of spiked-sediments and contaminated field-collected sediments. The ORNL benchmarks include the Effects Range-Low, Effects Range-Median, Threshold Effects Level, and Probable Effects Level, all of which were derived from the same database of biological effects in marine and estuarine sediments.

- Apparent effects thresholds are sediment chemical concentrations above which statistically significant biological effects always occur. They are site-specific and they may be under-protective, given that biological effects are observed at much lower chemical concentrations. The ORNL benchmarks include apparent effects thresholds for several ionic and polar organic chemicals, because other, better values are not available.

- Screening level concentrations are derived from synoptic data on sediment chemical concentrations and benthic invertebrate distributions. They are estimates of the highest concentration that can be tolerated by a specified percentage of benthic species. The ORNL benchmarks include the Ontario Ministry of the Environment Lowest and Severe Effect Levels, which were derived by this method.

- Equilibrium partitioning benchmarks are bulk sediment concentrations derived from aqueous benchmark concentrations based on the tendency of nonionic organic chemicals to partition between the sediment pore water and sediment organic carbon. The fundamental assumptions are that pore water is the principal exposure route for most benthic organisms and that the sensitivities of benthic species is similar to that of the species tested to derive the aqueous benchmarks, which are predominantly water column species. The ORNL benchmarks include the *proposed* EPA sediment quality criteria and five types of sediment benchmarks calculated from the ORNL benchmarks for aquatic biota.

- Sediment toxicity test benchmarks are derived from tests in which organisms are exposed to contaminated field-collected sediments and the observed effects are associated with the measured chemical concentrations. The ORNL benchmarks include the Threshold, Probable, and High No Effect Concentrations, which are a subset of the sediment effect concentrations calculated by Ingersoll et al. (EPA 1996b).

Table 2. Descriptions of screening benchmarks for sediment-associated biota[a]

Benchmark	Description[b]
Effects Range-Low (ER-L)	The tenth percentile of estuarine sediment concentrations reported to be associated with some level of toxic effects. These are possible-effects benchmarks.
Effects Range-Median (ER-M)	The fiftieth percentile of estuarine sediment concentrations reported to be associated with some level of toxic effects. These are probable-effects benchmarks.
Threshold Effect Level (TEL)	The geometric mean of the fifteenth percentile of reported concentrations which were associated with some level of effects and the fiftieth percentile of reported concentrations which were associated with no adverse effects. All data are for marine and estuarine sediments. These are possible-effects benchmarks.
Probable Effect Level (PEL)	The geometric mean of the fiftieth percentile of reported concentrations which were associated with some level of effects and the fiftieth percentile of reported concentrations which were associated with no adverse effects. All data are for marine and estuarine sediments. These are possible-effects benchmarks.
Ontario Ministry of the Environment (MOE) Lowest Effect Level (LEL)	Concentrations determined by the Ontario MOE to constitute thresholds for toxic effects in Ontario sediments. For most chemicals this is the concentration that can be tolerated by approximately 95% of benthic invertebrates. These are possible-effects benchmarks
Ontario Ministry of the Environment (MOE) Severe Effect Level (SEL)	Concentrations determined by the Ontario MOE to constitute thresholds for severe toxic effects in Ontario sediments. For most chemicals this is the concentration that can be tolerated by approximately 5% of benthic invertebrates. These are probable-effects benchmarks.
National Sediment Quality Criteria (EPASQC1)	Proposed sediment quality criteria based on toxicity in water expressed as chronic water quality criteria (recalculated after adding some benthic species) and partitioning of the contaminant between organic matter (1% of sediment) and pore water. In the absence of site-specific data, organic matter content is assumed to be one percent by weight. These are probable-effects benchmarks.

Table 2. (Cont.)

Benchmark	Description[b]
Equilibrium Partitioning Benchmarks (EQPAWQC1, EQPSCV1, EQPCVD1, EQPCVF1, and EQPCVI1)	Benchmarks derived in the same manner as sediment quality criteria except that the expression of aqueous toxicity is one of five benchmarks: the chronic NAWQC (EQPAWQC1), the Secondary Chronic Value (EQPSCV1), the Lowest Chronic Value for Daphnids (EQPCVD1), the Lowest Chronic Value for Fish (EQPCVF1), or the Lowest Chronic Value for Non-daphnid Invertebrates (EQPCVI1). In the absence of site-specific data, organic matter content is assumed to be one percent by weight. The EQPSCV1 is a possible-effects benchmark, all others are probable-effects benchmarks.
Threshold Effect Concentration (TEC)	The representative effect concentration selected from among the ER-Ls and TELs for *Hyalella azteca* and *Chironomus riparius* presented in EPA (1996b) based on the ranking method presented in Jones et al. (1997b). It is a concentration below which adverse effects to these organisms are not expected. The majority of the data are for freshwater sediments. These are possible-effects benchmarks.
Probable Effect Concentration (PEC)	The representative effect concentration selected from among the ER-Ms and PELs for *Hyaleaa azteca* and *Chironomus riparius* presented in EPA (1996b) based on the ranking method presented in Jones et al. (1997b). It is a concentration above which adverse effects to these organisms are likely to occur. The majority of the data are for freshwater sediments. These are probable-effects benchmarks.
High No Effect Concentration (NEC)	The representative effect concentration selected from among the high no-effect-concentrations for *Hyalella azteca* and *Chironomus riparius* presented in EPA (1996b) based on the ranking method presented in Jones et al. (1997b). It is a concentration above which adverse effects to these organisms are likely to occur. The majority of the data are for freshwater sediments. These are probable-effects benchmarks.
Region IV Benchmark (RIV)	The higher of two values, the EPA Contract Laboratory Program Practical Quantification Limit and the Effects Value, which is the lower of the ER-L and the TEL. These are possible-effects benchmarks.

Table 2. (Cont.)

Benchmark	Description[b]
EPA Office of Solid Waste and Emergency Response (OSWER) Screening Value	The lower limit of the 95% confidence interval for the proposed SQC value or, if an SQC has not been proposed, the sediment quality benchmark (SQB) calculated in the same manner as the SQC except that a Tier II Secondary Chronic Value is used. Organic matter content is assumed to be one percent by weight. The ER-L value is used if neither an SQC nor an SQB was available. These are possible-effects benchmarks.
Apparent Effect Threshold (AET)	A concentration above which toxic effects occurred at all sites in Puget Sound. These are probable-effects benchmarks.

[a] More details are presented by Jones et al. (1997b), Long et al. (1995), MacDonald (1994), and EPA Region IV (1995).

[b] Possible-effects benchmarks are conservative estimates of concentrations at which toxicity may occur. Probable-effects benchmarks are concentrations at which toxicity is likely.

Almost all of the ORNL sediment benchmarks were originally developed, calculated, and published by other agencies and organizations. The exception is the equilibrium partitioning benchmarks, which were calculated based on five types of aqueous benchmarks presented by Suter and Tsao (1996), as described in Table 2. Additionally, EPA (1996c) calculated up to 15 sediment effect concentrations for each contaminant and reported several measures of the reliability of each of these values. We selected a subset of three benchmarks (i.e., the Threshold, Probable, and High No Effect Concentrations) for each chemical based on the reported ability of a concentration to correctly classify samples as toxic or nontoxic (Jones et al. 1997b)

The ORNL benchmarks from the first two classes listed above were derived from marine and estuarine sediment data. Although data from studies of salt water sediments may not seem relevant to freshwater sediments, these data have been recommended by EPA Region IV (1995) and the EPA Office of Solid Waste and Emergency Response (OSWER 1996). Their use may be justified on the basis of the apparently small difference in the toxicity of many chemicals between the two media relative to the differences among sites within a medium. The ORNL benchmarks from the three other classes are based primarily on freshwater data (the sediment effect concentrations included a few estuarine sediments tested with 10% salinity in the overlying water).

Each of the 18 sediment benchmarks described in Table 2 is classified as either a possible-effects or probable-effects benchmark. Possible-effects benchmarks are conservative estimates of concentrations at which toxicity may occur, e.g., the tenth percentile of the sediment concentrations reported to be toxic. Probable-effects benchmarks are concentrations at which toxicity is likely, e.g., the fiftieth percentile of the sediment concentrations reported to be.

3.2 Development history

The ORNL benchmarks for sediment-associated biota have been revised four times since they were first presented in 1993 (Hull and Suter 1993 and 1994, and Jones et al. 1996, 1997a, and 1997b). The original set of benchmarks was limited to approximately 100 values for fourteen contaminants. Benchmarks of limited value (e.g., those based on background concentrations) have been removed, and high quality effects-based benchmarks have been added as they became available.

3.3 Selection of studies and other QA/QC

Unlike the other sets of ORNL benchmarks, we did not collect individual studies from the literature for calculation of the sediment benchmarks. Rather, the sediment effects data were accumulated and evaluated by the originating agencies and organizations (e.g., the National Ocean and Atmospheric Administration, the Florida Department of Environmental Protection, and the Ontario Ministry of the Environment), each of which had standardized protocols for data acceptability and quality assurance/quality control. The exception are the equilibrium partitioning benchmarks that were calculated from the ORNL aquatic benchmarks. Study selection and QA/QC for the water benchmarks is presented in Section 2.3. Additionally, data used to generate the equilibrium partitioning benchmarks were independently verified for accuracy, and a randomly selected subset of the benchmark values were checked by hand calculation. All data and calculations are documented in reports that are available to the public, enabling users to check questionable values.

3.4 Use and interpretation of results

As with the aquatic benchmarks, the large number of potential sediment benchmarks makes it necessary to decide which ones to use in a particular assessment. At ORNL, the entire suite of benchmarks is used in order to provide greater assurance of detecting all COPECs and to obtain the greatest amount of information from the screening assessment. For example, exceeding only one conservatively estimated benchmark (i.e., a possible-effects benchmark) may provide weak evidence of real effects, whereas exceeding multiple probable-effects benchmarks may provide strong evidence of real effects. These inferences can be used to refine future sampling and assessment efforts. However, a chemical is considered a COPEC if the maximum detected concentration exceeds any benchmark.

4. BENCHMARKS for PLANTS, SOIL INVERTEBRATES, and SOIL MICROBIAL PROCESSES

4.1 Background

The ORNL benchmarks for the toxicity to plants from chemical contaminants in soil were initially developed in 1993 because national regulatory criteria for soils that are intended to protect ecological receptors were not available. These thresholds for effects on growth and reproduction were derived from published toxicity studies conducted in soil or solution. The benchmarks are concentrations of chemicals that correspond to the Lowest Observed Effects Concentration (LOEC) for the 10^{th} percentile of plant species tested. Statistically significant effects thresholds are used unless a lower concentration tested corresponded with a 20% level of effects. Tests conducted in nutrient and mineral solution are assumed to be representative of exposures of plants to contaminants measured in very shallow groundwater (seeps and springs) or in aqueous extracts of soil.

The ORNL benchmarks for toxicity to soil invertebrates and heterotrophic processes from chemical contaminants in soil were initially developed in 1994. These benchmarks represent thresholds (LOECs) for statistically significant effects on growth, reproduction, or activity.

For all plants and soil organisms, the method for deriving soil benchmarks is based on the National Oceanographic and Atmospheric Administration's method for deriving the Effects Range Low (ER-L) (Long and Morgan 1991), which has been recommended as a sediment screening benchmark by the United States Environmental Protection Agency (EPA) Region IV. The ER-L is the 10th percentile of the distribution of various toxic effects thresholds for various organisms in sediments.

This approach can be justified by assuming that the toxicity of a chemical in soil is a random variate, the toxicity of contaminated soil at a particular site is drawn from the same distribution, and the assessor desires to be 90% certain of protecting plants, soil invertebrates, or heterotrophic processes on the site. The user of the benchmarks should be aware that differences in bioavailability of metal salts added to soil in laboratory tests and the multiple forms in the field contradict the assumptions above.

The toxicity benchmarks were derived by rank-ordering the LOEC values and then selecting a value that approximated the 10th percentile. If 10 or fewer values were available for a chemical, the lowest LOEC was used. If the 10th percentile fell between LOEC values, a value was chosen by interpolation. If a chemical concentration in soil represented a 50% or higher reduction in survivorship of plants, the concentration was divided by 5 to approximate the more sensitive endpoints of growth or production.

Plant toxicity benchmarks for metals are usually lower than those for soil invertebrates or microbial processes, and they are lower than most preliminary remediation goals (PRGs, LMES 1997) calculated for wildlife found on the Oak Ridge Reservation (wildlife benchmarks are not chemical concentrations in soil; see Sect. 5). Exceptions include low chromium and copper benchmarks for earthworms and low wildlife PRGs for barium, lead, mercury, selenium, and zinc. Because of a combined low sensitivity to and uptake of organic chemicals by plants, phytotoxicity benchmarks for these chemicals are typically higher than those for earthworms or PRGs for wildlife.

Threshold levels for effects on soil invertebrates and microbial processes are more uncertain than those for aquatic and sediment organisms, but probably more certain than those for wildlife. Because of the high degree of uncertainty, soil benchmarks are reported as numbers with a single significant figure. In all soil benchmarks reports, the level of confidence in the benchmarks is reported.

4.2 Development history

The ORNL plant toxicity benchmarks have been revised three times since they were first presented in 1993 (Suter et al. 1993, Will and Suter 1994a, Will and Suter 1995a); they were last updated in summer, 1997 (Efroymson et al. 1997a). The benchmarks for soil invertebrates, first presented in 1994 (Will and Suter 1994b), have been revised twice (Will and Suter 1995b, Efroymson et al. 1997b). The values for soil microbial processes that were updated in 1995 are presented unchanged in the 1997 volume. Each revision has included additional studies and benchmarks for additional chemicals.

4.3 Selection of studies and other QA/QC

References on the toxicity of selected chemicals to terrestrial plants, soil invertebrates, and heterotrophic processes were obtained from literature searches of bibliographic databases (Current Contents, BIOSIS, and POL TOX), review articles, conventional literature searches, and a numeric database for plant toxicity (PHYTOTOX).

Data for the derivation of benchmarks were generally obtained from primary sources, although secondary sources were used if a cited primary source was unavailable or if data for a particular chemical were sparse and a secondary source indicated that a lower concentration value was a more appropriate benchmark than values from the primary sources. The general criteria for inclusion of a study in the dataset used to derive soil toxicity benchmarks were:

1. The methodology was clearly stated (especially concentrations of chemicals applied to or measured in soil) and followed in the experiment.

2a. (Plants) Results were quantified as measures of growth or yield (or survival or metabolic activity if no growth or yield data were available).

2b. (Soil invertebrates) Results were quantified as measures of survivorship, growth, or reproduction.

2c. (Heterotrophic processes) Results were quantified as measures of respiration, carbon substrate or nitrogen transformation, enzyme activity, or mycelial growth.

3. Results were presented in numeric form, or graphical presentations of data were clearly interpretable.

4. An unambiguous reduction in the measured parameter existed within the range of applied concentrations of the chemical of interest.

Studies were included even if the lowest concentration of a chemical tested was associated with toxicity. Studies of plants were not excluded on the basis of plant type, although most species tested were crop species. Data were typically from greenhouse or growth chamber studies. Early versions of the benchmarks used data from plants grown in vermiculite or quartz sand (in addition to soil and solution), but these tests were determined not to be representative of field conditions and were eliminated from recent versions of the benchmarks database and reports. All studies of soil invertebrates from which benchmarks were derived were studies with earthworms; insufficient information about other soil invertebrates was available. Most studies of soil heterotrophic processes were tests of native soil microflora, with occasional tests of *Pseudomonas* sp. For nickel, data on the mycelial growth rates of numerous fungi were available.

Spreadsheet calculations and estimations of tenth percentiles of LOEC values for particular chemicals were independently verified by hand calculations. Data and calculations are documented in reports that are available to the public; thus questionable values identified by users of the benchmarks can be checked.

4.5 Application of soil benchmarks and interpretation of results

In the screening assessment, a chemical is considered a COPEC if the maximum, above-background concentration of the chemical in soil at a location exceeds a toxicity benchmark. The particular type of benchmarks used are determined by the choice of assessment endpoints; often, risk managers do not want to protect microbial processes. The maximum concentration (as opposed to the use of the 95% upper confidence limit of the mean concentration) at a location is used because of the immobility of soil organisms (i.e., some subset of the organisms is exposed to the maximum concentration).

In the definitive risk characterization, the single chemical toxicity line of evidence requires the analysis of site-specific and study-specific information, as well as toxicity benchmarks. This information includes: the relationship between bioavailability, plant or invertebrate taxon, soil type, and chemical speciation at the site and those parameters in studies from which the benchmarks were derived. A user of the benchmarks may choose to develop a set of benchmarks from a subset of the published data, if the subset better reflects the conditions at the site of concern.

5. BENCHMARKS for WILDLIFE

5.1. Background

The ORNL wildlife benchmarks were initially developed in 1993 and consist of concentrations in food or water that are equivalent to No and Lowest Observed Adverse Effects Levels (NOAELs or LOAELs) for avian and mammalian wildlife species. Wildlife NOAELs and LOAELs are estimated from experimentally derived NOAELs or LOAELs by using body weight and allometric scaling models to extrapolate the oral dose (mg/kg/d) for test species to an oral dose for wildlife species. This approach is based on the methodology used by the U.S. EPA for deriving human toxicity values from animal data (EPA 1986, 1992). These estimated NOAELs and LOAELs are combined with species-specific food and water consumption rates to generate food and water benchmarks. For wildlife species that feed primarily on aquatic organisms, a piscivore benchmark is also calculated. The piscivore benchmark combines exposure through both food and water and is calculated based on the potential of the contaminant to bioconcentrate and bioaccumulate through the food chain.

Historical basis/precedents for the approach

For wildlife, interspecies extrapolations are conventionally made by assuming that the differences in sensitivity among organisms and species are attributed to differences of physical scale. The simplest and most common example of this is the expression of doses to wildlife as dose per unit mass (mg/kg) which amounts to an assumption that toxicity is a function of the dilution of the toxicant in the mass of the organism. The model for extrapolation by scaling is, $E_a = E_t + e$, where E_a and E_t are the assessment and test endpoint species, respectively, when both the endpoint species and test species are appropriately scaled. The formal analysis of the consequences of organism size in physiology, ecology, pharmacology, and other branches of biology is termed allometry.

The most commonly used allometric model is a power function of weight, $E_x = a\,W^b$. This form has been adopted by toxicologists because various physiological processes including metabolism and excretion of drugs and other chemicals are approximated by that form (Davidson et al. 1986; Peters 1983). Exponents for various processes range from 0.6 to 0.8. Earlier versions of the ORNL wildlife benchmarks (Opresko et al. 1993) employed an exponent of 0.66, as used by the EPA (1986) in human health risk assessments. This practice is conservative for humans and mammalian wildlife in that large species such as deer are estimated to be more sensitive than the small rodents that are typically used in mammalian toxicity testing, while small wild species are estimated to be approximately equal in sensitivity to test species. More recently, the EPA has investigated the use of the less conservative 3/4 power for piscivorous wildlife (EPA 1995), and the 1996 revision of the ORNL wildlife benchmarks reflects this approach (Sample et al. 1996). Acute mammalian toxicity data sets yield exponents that are closer to 3/4 than 2/3 on average, but are consistent with either value (Goddard and Krewski 1992; Travis and Morris 1992; Watanabe et al. 1992).

Little attention has been paid to allometric models for avian toxicology. However, use of the same models for birds as mammals with the same exponents was supported by allometric models of avian physiology (Peters 1983) and pharmacology (Pokras et al. 1993). In fact, Pokras et al. (1993) present models for the extrapolation of effective doses of drugs from mammals to birds based on a common exponent of 3/4 but with a higher *a* value (intercept) for birds. In contrast, Mineau et al. (1996) performed allometric regression analyses on 37 pesticides with between six and 33 species of birds. They found that for 78% of chemicals the exponent was greater than 1 with a range of 0.63 to 1.55 and a mean of 1.1. However, because scaling factors for the majority of the chemicals evaluated (29 of 37) were not significantly different from 1, Sample et al. (1996) considered a scaling factor of 1 to be most appropriate for interspecies extrapolation among birds.

In addition to the application of allometric scaling, uncertainty factors (UF) are applied as part of the ORNL wildlife benchmarks for estimation of a NOAEL from a LOAEL and estimation of a chronic NOAEL or LOAEL from a subchronic value. In both cases, a UF of 10 is used. Use of a NOAEL-LOAEL UF of 10 is consistent with U.S. EPA guidance (EPA 1997) and is more conservative than the UF of 5 recommended in CalEPA guidance (California EPA 1996). Because EPA (1997) state that UFs of up to 10 may be used for subchronic-chronic extrapolation, the factor of 10 used in the ORNL wildlife benchmarks is also consistent with U.S. EPA guidance.

Comparison of Wildlife benchmarks to ORNL benchmarks for other taxa.

The wildlife benchmarks differ from the ORNL benchmarks for other taxa in two respects. First, the wildlife benchmarks are generally based on data from a single study. In contrast, ORNL benchmarks for other endpoint groups are based on results from multiple studies. Second, the ORNL wildlife benchmarks are species-specific, explicitly extrapolating toxic effects from test species to wildlife endpoint species. The ORNL benchmarks for other endpoint groups are based on data from multiple tests and species. While these other benchmarks may be taxa-specific (e.g., vascular plants, soil invertebrates, fish, etc.) they are not species-specific.

5.2 Development History

The ORNL wildlife benchmarks have been revised three times since they were first presented in 1993 (Table 3), with each revision expanding the database and improving the science behind the benchmarks. The most recent version of the wildlife benchmarks report is Sample et al. (1996).

Table 3. Chronology of development of wildlife benchmark values at ORNL.

Document Version	Number of Chemicals Presented	Number of Taxa Considered	Toxicity Data Presented	Interspecies Extrapolation Method
Opresko et al. 1993	55	6 mammals 8 birds	NOAELs only	body weight$^{2/3}$ for both birds and mammals
Opresko et al. 1994	76	8 mammals 9 birds	NOAELs and LOAELs	body weight$^{2/3}$ for both birds and mammals

Table 3. (Cont.)

Document Version	Number of Chemicals Presented	Number of Taxa Considered	Toxicity Data Presented	Interspecies Extrapolation Method
Opresko et al. 1995	85	8 mammals 11 birds	NOAELs and LOAELs	body weight$^{2/3}$ for both birds and mammals
Sample et al. 1996	85	9 mammals 11 birds	NOAELs and LOAELs	body weight$^{3/4}$ for mammals and body weight^{1} for birds

5.3 Selection of studies and other QA/QC

Literature Search and Selection of Studies

Data for the ORNL wildlife benchmarks were extracted from primary, published toxicity studies on birds and mammals. Studies to be considered for the benchmarks were identified by searching both computer literature databases and the reference lists of contaminant-specific review publications. The primary bibliographic and numerical databases were Current Contents and POLTOX. Review publications considered included ATSDR Toxicological Profiles, synoptic reviews of chemical hazards to fish, wildlife, and invertebrates produced by the USFWS, National Academy of Sciences reports, and WHO Environmental Criteria reports. In addition to these sources, the reference lists of each toxicity study was also evaluated for additional relevant studies.

Once studies were identified, they were evaluated by several criteria to determine if the data were suitable for benchmark development. The primary selection criteria for suitable toxicity studies were:

- Oral exposure - Only oral exposure, through food, water, or by gavage, was considered so that toxicity data would be most comparable to dietary exposures that wildlife may experience in the field.

- Chronic exposure duration - Because wildlife are resident at contaminated sites, chronic exposures were considered to be most representative of those encountered in the field. Studies were considered to be chronic in duration if they were >1 yr for mammals or > 10 wks for birds. Exceptions to this rule were made for studies of short term exposures during a critical life-stage such as gestation.

- Reproductive effects - Because reproductive effects are generally more sensitive than lethal effects and have clear population level-implications, and because most risk management decisions for wildlife are made at the population level, studies that included reproductive endpoints were emphasized (Table 4). In the absence of reproductive effects data, alternative endpoints were considered (Table 4).

- Studies with multiple dose levels. These studies provide dose-response information and allow the determination of both a NOAEL and LOAEL.

Table 4. Literature-derived toxicity test endpoints used for the ORNL wildlife benchmarks

Toxicity test endpoint	Percent of benchmarks by toxicity test endpoint		
	Mammals	Birds	All Studies
Reproduction	67.4	70.1	68.3
Mortality	11.6	27.0	15.8
Liver/kidney toxicity	7.0		5.0
Longevity	4.6		3.3
Weight loss	7.0		5.1
Growth	1.2	2.9	1.7
Blood Chemistry	1.2		0.8

QA/QC for wildlife benchmarks.

Efforts were made to make all assumptions and calculations as transparent as possible. All values and units for data extracted from studies and assumptions applied as part of dose conversions are documented for each study in Appendix A of the wildlife benchmark reports. These summaries outline the values used for benchmark development, what assumptions were made and why, how the NOAELs and LOAELs were calculated from the reported data, and references for parameter values employed. These data are presented to allow interested readers to cross-check data extracted from original sources or to employ different assumptions for NOAEL and LOAEL estimation as they feel appropriate.

In addition to transparency of data and assumptions, calculations of interspecies scaling values and the food, water, and benchmark values were performed in a spreadsheet and checked for accuracy. All species-specific parameter values (e.g., body weight, food and water ingestion rates, etc.), study-specific parameter values (e.g., experimental NOAEL and LOAEL, test species and body weight, etc.), and equation functions incorporated into the spreadsheet were checked for accuracy. Any discrepancies were resolved by referring to the original literature. In addition to verification of input data and models, hand calculation of randomly selected observations was performed to verify that calculation errors within the software were not occurring.

5.4. Application of the wildlife benchmarks and interpretation of results

Screening assessment

As has been stated previously (Sects 1.2 and 1.5), the primary, intended application of the ORNL wildlife benchmarks is for screening assessments, with the purposes being to identify COPECs, wildlife endpoints potentially at risk, media presenting the risk, and to guide future data collection. For this purpose, the NOAEL-based food, water, or piscivore benchmarks should be employed. The NOAEL-based benchmarks are used because they represent the highest dose levels at which no statistically significant effects were observed. When these benchmarks are applied, a series of conservative assumptions are made. These assumptions include:

- the diet consists of one food type that is 100% contaminated;
- wildlife reside, forage, and drink water exclusively within the contaminated area;
- the concentration of the COPEC in food or water is 100% bioavailable.

These assumptions are consistent with guidance outlined by EPA (1997).

Because wildlife are mobile and integrate contaminants spatially and temporally, the mean chemical concentration is the most appropriate estimate of chemical concentrations in environmental media to which wildlife are exposed. To be conservative, the 95% confidence limit on the mean chemical concentration in media should be compared to the benchmark value.

For most screening assessments, the only available data is chemical concentrations in soil and water. Water concentrations are appropriately compared to either the water or piscivore benchmarks. Except when direct soil ingestion by wildlife is a significant concern, food benchmarks should not be compared to soil concentrations. While comparison of the food benchmark to soil will be conservative for those chemicals with limited bioavailability that are poorly taken up by biota, this approach will grossly underestimate risk from chemicals that biomagnify or accumulate to higher levels in biota than are found in soil (e.g., organochlorines, cadmium, and mercury). Soil-to-biota uptake factors or other bioaccumulation models should therefore be employed to estimate chemical concentrations in biota. Uptake models for plants, earthworms, and small mammals can be found in Baes et al. (1984), Sample et al. (1997), and Sample et al. (1998a and 1998b).

Baseline or definitive assessment

The purpose of the baseline or definitive assessment is to provide the best estimate of risks using all available data. Literature-derived avian and mammalian toxicity data, such as those used in the ORNL wildlife benchmarks, are but one line of evidence used in a definitive assessment (see Sect. 1.5).

In a definitive assessment, more realistic and accurate exposure estimates are generated using measured chemical concentrations in soil, water, and wildlife foods (if biota data are unavailable, uptake models are used), food-chain exposure models to estimate total exposure from multiple media, and species-specific and site-specific exposure modifying factors. To account for and express the uncertainty and variability associated with the additional parameters and assumptions employed in these more detailed exposure models, Monte Carlo analyses should be performed and exposure distributions generated. Comparison of exposure distributions to literature-derived avian and mammalian toxicity data indicates the likelihood (or probability) of wildlife experiencing exposures in excess of toxicity values.

As stated in Sect 1.5, risk assessors should return to the original literature in an effort to identify those toxicity studies that are most appropriate for their specific site and wildlife endpoint characteristics as a component of the baseline assessment. The studies employed in the ORNL wildlife benchmarks may or may not be the most appropriate studies for post-screening assessments. Important parameters that should be considered include chemical form and speciation, duration of exposure, severity and nature of effects, etc. To better define the nature and severity of potential effects, the use of multiple toxicity studies with different toxicity test endpoints is recommended. Examples of baseline risk assessments for wildlife are available from the ORNL Ecological Risk Assessment web site.

6. SUMMARY

The ORNL ecotoxicological screening benchmarks provide an assemblage of screening values for ecological risk assessment. They were developed using existing data and methods employed by, or consistent with, practices of various regulatory agencies and include values promulgated by the U.S. EPA. If site-specific benchmarks exist, they should be considered, as appropriate. The ORNL benchmarks represent concentrations in environmental media that are believed to be acceptable to ecological endpoints. The intended application for the ORNL benchmarks are as part of the screening assessment; the benchmarks should not be used as remedial goals and should be used in definitive assessments only if more appropriate data are not available. Used as recommended, the ORNL benchmarks can serve to identify contaminants, media, and receptors that may be at risk and that require further investigation. Despite being based on practices employed by regulatory agencies and being accepted by various regulatory agencies, the ORNL ecotoxicological benchmarks are not regulatory criteria. It is important to emphasize that it is the responsibility of the user to obtain approval for the use of the ORNL values at each site. Regulatory approval for the use of the ORNL benchmarks or other screening values should be obtained as part of a Data Quality Objectives (DQO) process (Bilyard et al. 1997) at the outset of any project.

7. REFERENCES

Baes, C. F., R. D. Sharp, A. L. Sjoreen and R. W. Shor. 1984. A review and analysis of parameters for assessing transport of environmentally released radionuclides through agriculture. ORNL-5786. Oak Ridge National Laboratory, Oak Ridge, TN.

Bilyard, G.R., H. Beckert, J.J. Bascietto, C.W. Abrams, S.A. Dyer, and L.A. Haselow. 1997. Using the data quality objectives process during the design and conduct of ecological risk assessments. DOE/EH-0544.

Cairns, J., Jr., K. L. Dickson and A. W. Maki. 1979. Estimating the hazard of chemical substances to aquatic life. Hydrobiologia 64:157-166.

California EPA (Environmental Protection Agency). 1996, Guidance for ecological risk assessment at hazardous waste sites and permitted facilities. Part A: Overview. Department of Toxic Substances Control, Human and Ecological Risk Division.

Davidson IWF, Parker JC, Beliles RP. 1986. Biological basis for extrapolation across mammalian species. Regul. Toxicol. Pharmacol. 6:211-237.

Efroymson, R. A., M. E. Will, G. W. Suter II, and A. C. Wooten. 1997a. Toxicological benchmarks for screening potential contaminants of concern for effects on terrestrial plants: 1997 Revision. ES/ER/TM-85/R2.

Efroymson, R. A., M. E. Will, and G. W. Suter II. 1997b. Toxicological benchmarks for screening potential contaminants of concern for effects on soil and litter invertebrates and heterotrophic process: 1997 Revision. ES/ER/TM-126/R2.

EPA (U.S. Environmental Protection Agency). 1986. Guidelines for carcinogenic risk assessment. Federal Register 51:33992-34003.

EPA (U.S. Environmental Protection Agency). 1992. Draft Report: A cross-species scaling factor for carcinogen risk assessment based on equivalence of $mg/kg^{3/4}/day$; Notice. Federal Register. 57(109)24152-24173.

EPA (U.S. Environmental Protection Agency). 1995. Great Lakes Water Quality Technical Support Document for Wildlife Criteria. Washington, D.C.: Office of Water, U.S. Environmental Protection Agency. EPA-820-B-95-009.

EPA (U.S. Environmental Protection Agency). 1996a. Proposed guidelines for ecological risk assessment. Fed. Regist. 61(175):47552-47631.

EPA (U.S. Environmental Protection Agency). 1996b. Ecotox thresholds. EPA 540/F-95/038. Office of Solid Waste and Emergency Response, U.S. Environmental Protection Agency, Washington, D.C.

EPA (U.S. Environmental Protection Agency). 1996c. Calculation and evaluation of sediment effect concentrations for the amphipod *Hyalella azteca* and the midge *Chironimus riparius.* EPA 905/R96/008. Great Lakes National Program Office, Chicago, IL.

EPA (U.S. Environmental Protection Agency). 1997. Ecological risk assessment guidance for superfund: process for designing and conducting ecological risk assessment, interim final. U.S. Environmental Protection Agency, Environmental Response Team, Edison, NJ.

Goddard MJ, Krewski D. 1992. Interspecies extrapolation of toxicity data. Risk Analysis 12:315-317.

Hull, R. N., and G. W. Suter II 1993. Toxicological Benchmarks for Screening Potential Contaminants of Concern for Effects on Sediment-Associated Biota. ES/ER/TM-95, Oak Ridge National Laboratory, Oak Ridge, Tenn.

Hull, R. N., and G. W. Suter II 1994. Toxicological Benchmarks for Screening Potential Contaminants of Concern for Effects on Sediment-Associated Biota: 1994 Revision, ES/ER/TM-95/R1, Oak Ridge National Laboratory, Oak Ridge, Tenn.

Jones, D. S., R. N. Hull, and G. W. Suter II 1996. Toxicological Benchmarks for Screening Potential Contaminants of Concern for Effects on Sediment-Associated Biota: 1996 Revision. ES/ER/TM-95/R2, Oak Ridge National Laboratory, Oak Ridge, Tenn.

Jones, D. S., G. W. Suter II , and R. N. Hull 1997a. Toxicological Benchmarks for Screening Potential Contaminants of Concern for Effects on Sediment-Associated Biota: 1997 Revision. ES/ER/TM-95/R3, Oak Ridge National Laboratory, Oak Ridge, Tenn.

Jones, D. S., G. W. Suter II , and R. N. Hull 1997b. Toxicological Benchmarks for Screening Potential Contaminants of Concern for Effects on Sediment-Associated Biota: 1997 Revision. ES/ER/TM-95/R4, Oak Ridge National Laboratory, Oak Ridge, Tenn.

LMES (Lockheed Martin Energy Systems, Inc.) 1997. Preliminary remediation goals for ecological endpoints. ES/ER/TM-162/R2.

Long, E. R,. and L. G. Morgan 1991. The Potential for Biological Effects of Sediment-Sorbed Contaminants Tested in the National Status and Trends Program, NOAA Technical Memorandum NOS OMA 52, National Oceanic and Atmospheric Administration.

Long, E. R., D. D. MacDonald, S. L. Smith, and F. D. Calder 1995. “Incidence of Adverse Biological Effects within Ranges of Chemical Concentrations in Marine and Estuarine Sediments,” Environmental Management 19(1), 81–97.

MacDonald, D. D. 1994. Approach to the Assessment of Sediment Quality in Florida Coastal Waters, Florida Department of Environmental Protection, Tallahassee, Florida.

Mineau, P., B.T. Collins, and A. Baril. 1996. On the use of scaling factors to improve interspecies extrapolation of acute toxicity in birds. Reg. Toxicol. and Pharmacol. 24: 24-29.

Opresko, D.M., B.E. Sample, and G.W. Suter, II. 1995. Toxicological benchmarks for wildlife: 1995 Revision. Oak Ridge National Laboratory, Oak Ridge, TN. ES/ER/TM-86/R2.

Opresko, D.M., B.E. Sample, and G.W. Suter, II. 1994. Toxicological benchmarks for wildlife: 1994 Revision. Oak Ridge National Laboratory, Oak Ridge, TN. ES/ER/TM-86/R1.

Opresko, D.M., B.E. Sample, and G.W. Suter, II. 1993. Toxicological benchmarks for wildlife. Oak Ridge National Laboratory, Oak Ridge, TN. ES/ER/TM-86.

OSWER (Office of Solid Waste and Emergency Response) 1996. "Ecotox thresholds," ECO Update 3(2):1–12.

Peters RH. 1983. The Ecological Implications of Body Size. Cambridge: Cambridge University Press.

Pokras MA, Karas AM, Kirkwood JK, Sedgewick CJ. 1993. An introduction to allometric scaling and its uses in raptor medicine. In: Redig PT, Cooper JE, Remple JD, Hunter DB, editors. Raptor Biomedicine. Minneapolis: U. Minnesota Press. p 211-224.

PNNL (Pacific Northwest National Laboratory). 1998. Screening assessment and requirements for a comprehensive assessment: Columbia River comprehensive impact assessment. DOE/RL-96-16.

Region IV (U.S. Environmental Protection Agency Region IV) 1995. Ecological screening values, Ecological Risk Assessment Bulletin No. 2, Waste Management Division, U.S. Environmental Protection Agency Region IV, Atlanta, Ga.

Sample, B.E., D.M. Opresko and G.W. Suter II. 1996. Toxicological Benchmarks for Wildlife: 1996 Revision. ES/ER/TM-86/R3. Oak Ridge National Laboratory, Oak Ridge, Tennessee.

Sample, B. E., M. S. Aplin, R. E. Efroymson, G. W. Suter, II and C. J. E. Welsh. 1997. Methods and tools for estimation of the exposure of terrestrial wildlife to contaminants. Oak Ridge National Laboratory, Oak Ridge, TN. ORNL/TM-13391.

Sample, B.E., J. Beauchamp, R. Efroymson, G.W. Suter, II, and T.L. Ashwood. 1998a. Development and validation of literature-based bioaccumulation models for small mammals. ES/ER/TM-219. Oak Ridge National Laboratory, Oak Ridge, TN.

Sample, B.E., J. Beauchamp, R. Efroymson, G.W. Suter, II, and T.L. Ashwood. 1998b. Development and validation of literature-based bioaccumulation models for earthworms. ES/ER/TM-220. Oak Ridge National Laboratory, Oak Ridge, TN.

Suter, G. W., II. 1990. Environmental risk assessment/environmental hazard assessment: similarities and differences: pp. 5-15. W. G. Landis and W. H. v. d. Schalie (eds.), Aquatic Toxicology and Risk Assessment: Thirteenth Volume. ASTM, Philadelphia, Pennsylvania.

Suter, G. W. II, M. E. Will, and C. Evans. 1993. Toxicological benchmarks for screening potential contaminants of concern for effects on terrestrial plants. ES/ER/TM-85.

Suter, G.W., II. 1995. Guide for performing screening ecological risk assessments at DOE facilities. Oak Ridge National Laboratory, Oak Ridge TN. ES/ER/TM-153.

Suter, G. W., II. 1996a. Risk characterization for ecological risk assessment of contaminated sites. ES/ER/TM-200. Oak Ridge National Laboratory, Oak Ridge, TN.

Suter, G. W., II. 1996b. Toxicological benchmarks for screening contaminants of potential concern for effects on freshwater biota. Environ. Toxicol. Chem. 15:1232-1241.

Suter, G. W., II, M. A. Futrell and J. A. Kerchner. 1992. Toxicological benchmarks for screening potential contaminants of concern for effects on aquatic biota on the Oak Ridge Reservation. ORNL/ER-139. Oak Ridge National Laboratory, Oak Ridge, TN.

Suter, G. W. II, M. E. Will, and C. Evans. 1993. Toxicological benchmarks for screening potential contaminants of concern for effects on terrestrial plants. ES/ER/TM-85.

Suter, G. W., II and C. L. Tsao. 1996. Toxicological benchmarks for screening potential contaminants of concern for effects on aquatic biota: 1996 revision. ES/ER/TM-96/R2. Oak Ridge National Laboratory, Oak Ridge, TN.

TNRCC (Texas Natural Resources Conservation Commission). 1996. Guidance for conducting ecological risk assessments under the Texas Risk Reduction Program. TNRCC, Office of Waste Management. RG-263. DRAFT.

Travis CC, and J.M. Morris. 1992. On the use of 0.75 as an interspecies scaling factor. Risk Analysis 12:311-313.

Watanabe K, Bois FY, Zeise L. 1992. Interspecies extrapolation: a reexamination of acute toxicity data. Risk Analysis 12:301-310.

Will, M. E. and G. W. Suter II. 1994a. Toxicological benchmarks for screening potential contaminants of concern for effects on terrestrial plants: 1994 Revision. ES/ER/TM-85/R1.

Will, M. E. and G. W. Suter II. 1995a. Toxicological benchmarks for screening potential contaminants of concern for effects on terrestrial plants: 1995 Revision. ES/ER/TM-85/R2.

Will, M. E. and G. W. Suter II. 1994b. Toxicological benchmarks for potential contaminants of concern for effects on soil and litter invertebrates and heterotrophic processes. ES/ER/TM-126.

Will, M. E. and G. W. Suter II. 1995b. Toxicological benchmarks for potential contaminants of concern for effects on soil and litter invertebrates and heterotrophic processes. ES/ER/TM-126/R1.