

**Risk Characterization for Ecological  
Risk Assessment  
of Contaminated Sites**

This document has been approved by the  
K-25 Site Technical Information Office  
for release to the public. Date: \_\_\_\_\_

**Risk Characterization for Ecological  
Risk Assessment  
of Contaminated Sites**

G. W. Suter II

Date Issued—October 1996

Prepared by the  
Environmental Restoration Risk Assessment Program  
Lockheed Martin Energy Systems, Inc.  
Oak Ridge, Tennessee 37831

Prepared for the  
U.S. Department of Energy  
Office of Environmental Management  
under budget and reporting code EW 20

LOCKHEED MARTIN ENERGY SYSTEMS, INC.  
managing the  
Environmental Management Activities at the  
Oak Ridge K-25 Site Paducah Gaseous Diffusion Plant  
Oak Ridge Y-12 Plant Portsmouth Gaseous Diffusion Plant  
Oak Ridge National Laboratory  
under contract DE-AC05-84OR21400  
for the  
U.S. DEPARTMENT OF ENERGY

## **PREFACE**

Risk characterization combines information concerning exposure to chemicals with information regarding effects of chemicals to estimate risks. This document describes the approach for estimating risks based on individual lines of evidence and then combining them through a process of weighing the evidence. The lines of evidence are integrated independently so that the implications of each are explicitly presented. This makes the logic of the assessment clear and allows independent weighing of the evidence by risk managers and stakeholders. This work was performed under Work Breakdown Structure 1.4.12.2.3.04.05.02 (Activity Data Sheet 8304). Publication of this document meets an Environmental Restoration Risk Assessment Program milestone for FY 96. The general approach to risk characterization was described in the strategy for ecological risk assessment on the Oak Ridge Reservation (Suter et al. 1995). This document expands that guidance by providing more specific information on how ecological risk characterization should be performed.

# CONTENTS

PREFACE .....	iii
FIGURES .....	vii
TABLES .....	ix
ACRONYMS .....	xi
EXECUTIVE SUMMARY .....	xiii
1. INTRODUCTION .....	1
2. SINGLE CHEMICAL TOXICITY .....	1
2.1 FISH .....	5
2.1.1 Aqueous Exposure .....	5
2.1.2 Fish Body Burdens .....	8
2.2 BENTHIC INVERTEBRATES .....	8
2.3 TERRESTRIAL PLANTS, SOIL INVERTEBRATES, AND MICROBIAL PROCESSES .....	12
2.4 SOURCES OF VARIANCE IN EFFECTS DISTRIBUTIONS .....	12
2.5 WILDLIFE .....	13
3. AMBIENT MEDIA TOXICITY TESTS .....	16
4. BIOLOGICAL SURVEYS .....	20
5. BIOMARKERS .....	22
6. WEIGHT OF EVIDENCE .....	24
7. FUTURE RISKS .....	30
8. REFERENCES .....	31

## FIGURES

1.	Risk characterization based on chemical analyses and single chemical toxicity	2
2.	Plot of concentrations of COPECs normalized to toxic units (TU)	4
3.	Empirical distribution functions for acute toxicity (LC50 and EC50 values) and chronic toxicity (Chronic Values) of copper to fish and aquatic invertebrates and for individual measurements of copper in surface water from two reaches	7
4.	Empirical distribution functions for acute toxicity and chronic toxicity of copper to fish and aquatic invertebrates and for individual measurements of copper in sediment pore water from five reaches	10
5.	Empirical distribution functions for toxicity of copper in sediment to sediment-associated organisms and for individual measurements of copper in sediment	11
6.	Distributions of estimated mercury exposures for rough-winged swallows in Poplar Creek subreaches.	14
7.	Distributions of estimated mercury exposures for osprey in Poplar Creek subreach 3.02	15
8.	Risk characterization based on toxicity testing of ambient media	17
9.	Mean proportional reduction in survival of medaka eggs in ambient water relative to controls and proportion of tests in which survival was significantly reduced (i.e., either reduced at least 20% or statistically significantly reduced) in numbered reaches and subreaches	19
10.	Risk characterization based on biological survey data	21
11.	Risk characterization based on biomarker data	23
12.	Risk characterization based on weighing of multiple lines of evidence	25

## TABLES

1.	Summary of integration of single chemical toxicity . . . . .	5
2.	Why contaminant concentrations in ambient media may not be correlated with toxicity of those media . . . . .	18
3.	Summary table for integration of ambient media toxicity test results . . . . .	20
4.	Summary table for integration of biological survey results . . . . .	22
5.	Summary table for integration of biomarker results . . . . .	24
6.	A hypothetical summary of a risk characterization by weight of evidence for a soil invertebrate community in contaminated soil . . . . .	27
7.	Example of a table summarizing the risk characterization for the fish community in a stream at a waste site . . . . .	28
8.	Summary of weight-of-evidence analyses for reaches exposed to contaminants in the Clinch River/Poplar Creek Operable Unit . . . . .	29

## ACRONYMS

AEC	ambient exposure concentration
COPEC	chemical of potential ecological concern
CV	chronic value
EPA	U.S. Environmental Protection Agency
ERA	ecological risk assessment
ER-L	effects range low
ER-M	effects range median
HQ	hazard quotient
LOAEL	lowest observed adverse effects level
LOEC	lowest observed effects concentration
LOEL	lowest observed effects level
PAH	polyaromatic hydrocarbon
PCB	polychlorinated biphenyl
NAWQC	National Ambient Water Quality Criteria
NOAA	National Oceanographic and Atmospheric Administration
NOAEL	no observed adverse effects level
NOEC	no observed effects concentration
NOEL	no observed effects level
SCV	secondary chronic values
TEC	toxicologically effective concentration
TU	toxic unit
$\Sigma$ TU	sum of toxic units

## EXECUTIVE SUMMARY

Risk characterization for ecological risk assessments (ERAs) is performed by weight of evidence (Risk Assessment Forum 1992). That is, rather than simply modeling risks, ecological risk assessors examine all available data from chemical analyses, toxicity tests, biological surveys, and biomarkers to estimate the likelihood that significant effects are occurring or will occur and describe the nature, magnitude, and extent of effects on the designated assessment endpoints. This document describes the approach for estimating risks based on individual lines of evidence and then combining them through a process of weighing the evidence. The lines of evidence are integrated independently so that the implications of each are explicitly presented. This makes the logic of the assessment clear and allows independent weighing of the evidence by risk managers and stakeholders. For each line of evidence, it is necessary to evaluate the relationship of the measurement endpoint to the assessment endpoint, the quality of the data, and the relationship of the exposure metrics in the exposure-response data to the exposure metrics for the site. The general approach was described in the strategy for ERA on the Oak Ridge Reservation (Suter et al. 1995). This document expands that guidance by providing more specific information on how ecological risk characterization should be performed.

The single chemical toxicity line of evidence uses analyses of individual chemicals in individual media to estimate exposure and uses literature values for effects of individual chemicals to estimate effects. They are combined in two steps. First, the chemicals are screened against ecotoxicological benchmarks, against background exposures, and, where possible, against characteristics of the source to determine which are chemicals of potential ecological concern (COPECs). This may have been done previously in screening assessments for earlier phases in the remedial process such as the Remedial Investigation work plan, but it should be repeated for each new assessment. Second, a more definitive characterization is performed by comparing the distributions of exposure and effects for each COPEC.

Risk characterization for the ambient media toxicity line of evidence begins by determining whether the tests show significant toxicity. Toxicity is not significant if the effects relative to controls are less than 20% (e.g., less than 20% mortality) and the effects are not statistically significantly different from controls. That is, effects are considered significant if (1) the hypothesis of no difference between responses in contaminated media and in either reference media or control media is rejected with 95% confidence (i.e., statistical significance) or (2) an effect of 20% or greater in survival, growth, or reproduction relative to either reference media or control media is observed (i.e., biological significance).

If biological survey data are available for an endpoint species or community, then the first question to be answered is whether the data suggest that significant effects are occurring. For some groups, notably fish and benthic invertebrates, there are abundant data from reference streams for comparison. For most other endpoint groups, references must be established *ad hoc* and the lack of temporal or spatial replication may make inference tenuous. For some taxa such as most birds, traditional survey data are not useful for estimating risks from wastes because mobility, territoriality, or other factors obscure demographic effects. However, survey results may be more reliable if efforts are made to control extraneous variance such as by setting out nest boxes on contaminated and reference sites to monitor reproductive success.



Biomarkers are seldom useful for estimating risks by themselves, but they can be used to support other lines of inference. The inference begins by asking if the levels of the biomarkers significantly differ from those at reference sites. If they do, then it is necessary to determine whether they are diagnostic or at least characteristic of any of the COPECs or of any of the habitat factors that are thought to affect the endpoint biota. If the biomarkers are characteristic of contaminant exposures, then the distribution and frequency of elevated levels must be compared to the distributions and concentrations of contaminants. Finally, to the extent that the biomarkers are known to be related to overt effects such as reductions in growth, fecundity, or mortality, the implications of the observed biomarker levels for populations or communities should be estimated.

The weighing of evidence begins by summarizing the available lines of evidence for each endpoint. Given that one has estimated risks based on each line of evidence, the process of weighing the evidence amounts to determining what estimate of risks is most consistent with the results for all lines of evidence. If the assessment endpoint is defined in terms of some threshold for significance, then the process can be conducted in two steps. First, for each line of evidence determine whether it is consistent with exceedence of the threshold, inconsistent with exceedence, or ambiguous. Second, determine whether the results as a whole indicate that it is likely or unlikely that the threshold is exceeded. If the results for all lines of evidence are consistent or inconsistent, the result of the weighing of evidence is clear. Assuming that there is no consistent bias in the assessment, agreement among multiple lines of evidence is strong evidence to support a conclusion. However, if there are inconsistencies, the true weighing of evidence must occur.

## 1. INTRODUCTION

Risk characterization combines information concerning exposure to chemicals with information concerning effects of chemicals to estimate risks. Risk characterization for ecological risk assessments (ERAs) is performed by weight of evidence (Risk Assessment Forum 1992). That is, rather than simply modeling risks, ecological risk assessors examine all available data from chemical analyses, toxicity tests, biological surveys, and biomarkers to estimate the likelihood that significant effects are occurring or will occur and describe the nature, magnitude, and extent of effects on the designated assessment endpoints. This document describes the approach for estimating risks based on individual lines of evidence and then combining them through a process of weighing the evidence. The lines of evidence are integrated independently so that the implications of each are explicitly presented. This makes the logic of the assessment clear and allows independent weighing of the evidence by risk managers and stakeholders.

For each line of evidence, it is necessary to evaluate the relationship of the measurement endpoint to the assessment endpoint, the quality of the data, and the relationship of the exposure metrics in the exposure-response data to the exposure metrics for the site. The general approach was described in the strategy for ERA on the Oak Ridge Reservation (Suter et al. 1995). This document expands that guidance by providing more specific information on how ecological risk characterization should be performed.

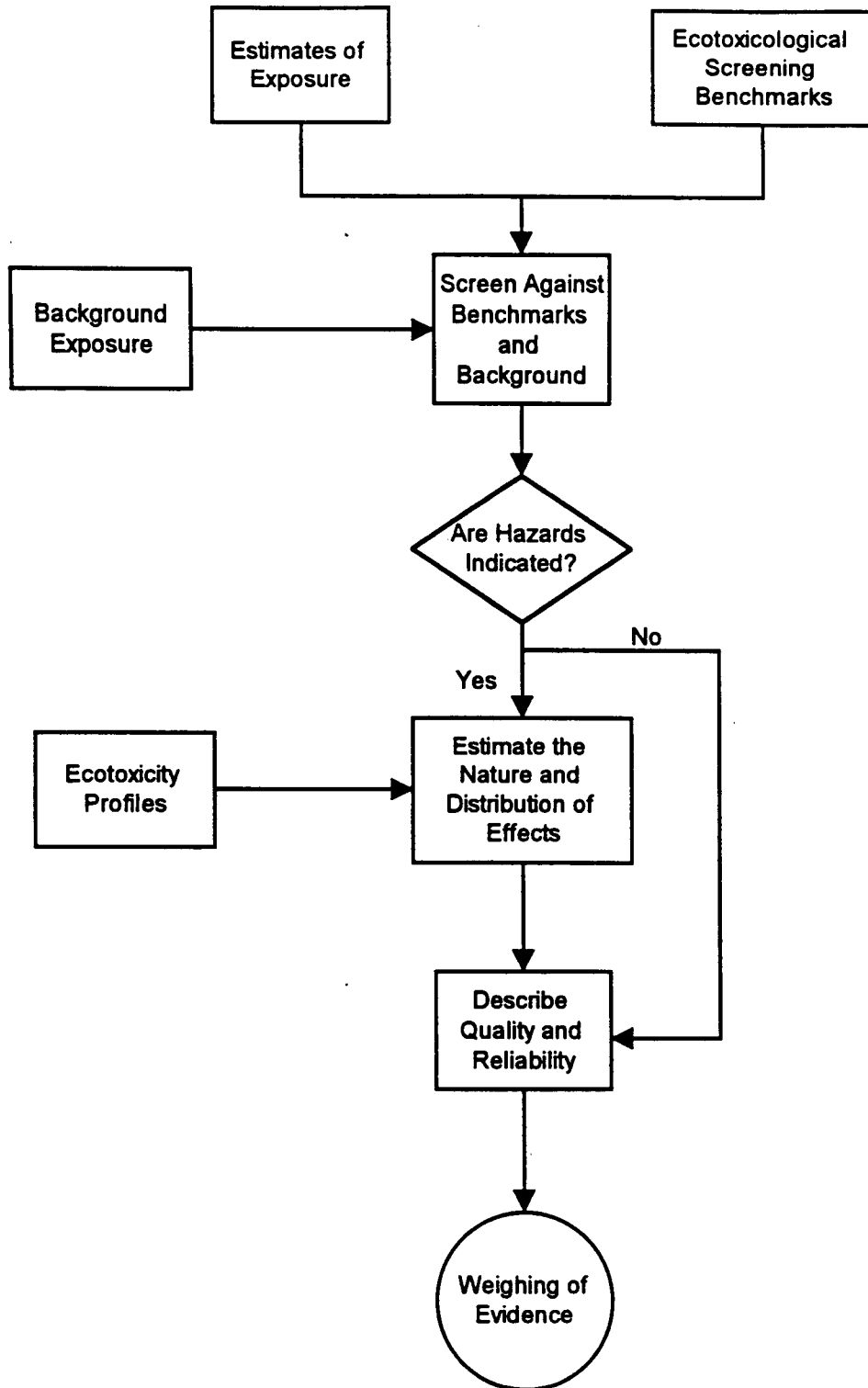
## 2. SINGLE CHEMICAL TOXICITY

This line of evidence uses analyses of individual chemicals in individual media to estimate exposure and uses literature values for effects of individual chemicals to estimate effects (Fig. 1). They are combined in two steps. First, the chemicals are screened against ecotoxicological benchmarks, against background exposures, and, where possible, against characteristics of the source to determine which are chemicals of potential ecological concern (COPECs). This may have been done previously in screening assessments for earlier phases in the remedial process such as the Remedial Investigation work plan, but it should be repeated for each new assessment. Methods for this portion of the assessment are presented in prior guidance (Suter 1995). The results of the screening assessment should be presented as a table listing all of the chemicals that exceeded benchmarks, indicating which are COPECs and the reasons for acceptance or rejection.

The integration of exposure with single chemical toxicity data is minimally expressed as a quotient of the ambient exposure concentration (AEC) divided by the toxicologically effective concentration (TEC):

$$HQ = AEC/TEC.$$

The TEC may be a test endpoint, a test endpoint corrected by a factor or other extrapolation model, or a regulatory criterion or other benchmark value; this type of analysis is used for risk characterization in screening assessments. In that case, conservative AEC values are used and a hazard quotient (HQ) greater than one is treated as evidence that the chemical is worthy of concern. For definitive assessments, more realistic exposure estimates are used for the AEC and effects are expressed as test endpoints that are closely related to the assessment endpoint or to regulatory standards.



**Fig. 1. Risk characterization based on chemical analyses and single chemical toxicity.**

If numerous chemicals occur at potentially toxic concentrations, it is useful to calculate an index of total toxicity, the sum of toxic units ( $\Sigma$ TU). This permits the assessor and reviewers to compare the COPECs to each other and examine their distributions across reaches or areas within a site. Since the relative importance of COPECs is a function of their potential toxicity rather than their concentration, toxicity normalized concentrations or toxic units (TUs) are calculated. This is a common technique for dealing with exposures to multiple chemicals by expressing concentration relative to a standard test endpoint (Finney 1971).

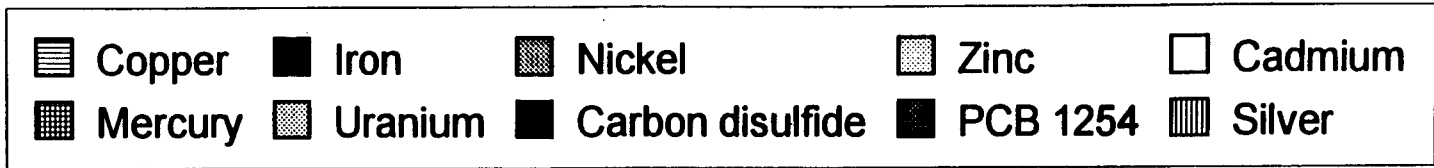
TUs are quotients of the concentration of a chemical in a medium divided by the standard test endpoint concentration for that chemical. They are similar to HQs except that a common test endpoint is used rather than conservative benchmarks because TUs are used for comparative purposes rather than to draw conclusions. The expression of concentration and the test endpoint vary among media; for water they are the upper 95% confidence limit concentration and the 48-hour EC50 for *Daphnia magna* (the most common aquatic test endpoint). If the TU for a chemical equals one, the interpretation is that the aquatic community in that reach is exposed to a conservatively estimated average concentration sufficient to kill or immobilize *Daphnia* within 48 hours.

The chemicals that constitute a major component of toxicity (i.e., TUs > 0.01) should be plotted for each reach or area for water, sediment, soil, and wildlife intake (e.g., Fig. 2). The choice of a cutoff for inclusion is based on the fact that acute values are used in calculating the TUs, and chronic effects can occur at concentrations as much as two orders of magnitude below acute values. Other values may be used if specific circumstances warrant.

The height of the plot at each subreach is the sum of toxic units ( $\Sigma$ TU) for that medium and subreach. This value can be conservatively interpreted as the total toxicity-normalized concentration and therefore as a relative indication of the toxicity of the medium in that subreach. In addition, the  $\Sigma$ TU is commonly assumed to estimate the absolute toxicity of the medium. That is, if the  $\Sigma$ TU equals one, then the endpoint effect (e.g., *Daphnia* acute lethality) will occur. This will be the case if all of the chemicals have the same mode of action. For heterogeneous chemical mixtures it is likely to be a conservative assumption because combined toxic effects of chemicals in environmental samples have been found to be additive or less than additive, not superadditive (Alabaster and Lloyd 1982). Because the test endpoints are chosen for their consistency rather than their relationship to an assessment endpoint, the plots of TUs are heuristic, providing an indication of the relative toxicity of sites and the relative contributions of chemicals to that toxicity.

For all COPECs for each endpoint, exposures must be compared to the full toxicity profile of the chemical to characterize risk. For example, the distribution of concentrations in water would be compared to the distribution of concentrations of thresholds for chronic toxicity across fish species and across prey species, the nature of the chronic effects would be described, and the exposure durations needed to achieve effects in the laboratory would be compared to temporal dynamics of concentrations in the field. Characteristics of the chemicals that are relevant to risks are also examined such as the influence of metal speciation on toxicity, tendency of the chemical to accumulate in prey species, etc.

Inferences about the risk posed by the COPECs is based on the distribution of concentrations relative to the distribution of effects. Distributions provide a better basis for inference than point estimates because they allow consideration of variance in concentration over space or time and of sensitivity across species, measures of effects, media properties, or chemical forms. In all cases, risk is a function of the overlap between the exposure and effects distributions, but the interpretation



Sum TU

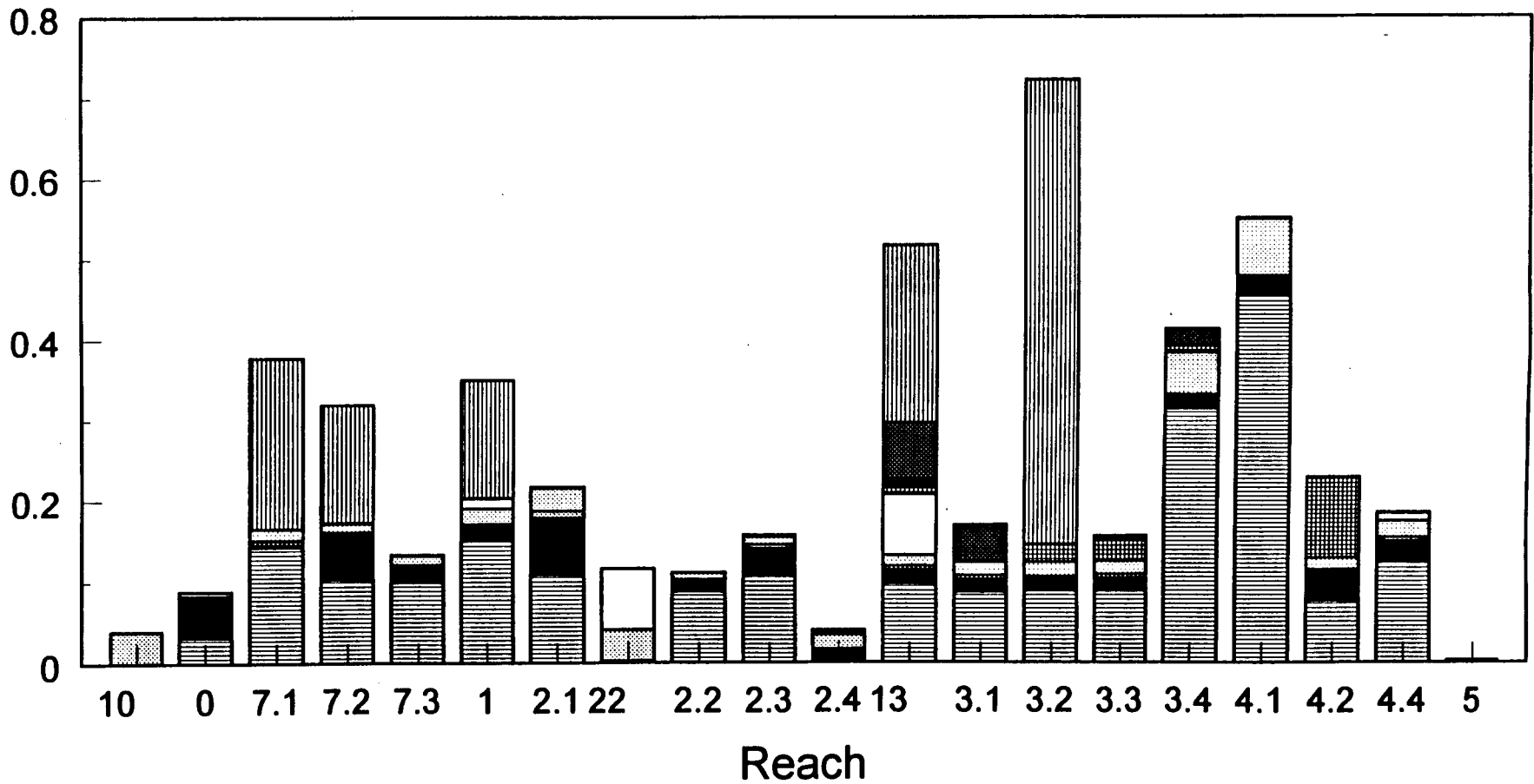


Fig. 2. Plot of concentrations of COPECs normalized to toxic units (TU). The level of the tops of the bars is the  $\Sigma$ TU for the reach or subreach.

depends on the data that are used. These interpretations are explained in the following subsections for the different classes of endpoints.

For all endpoints the risk characterization ultimately depends on weighing of all of the lines of evidence. To facilitate the weight-of-evidence analysis and to make the bases clear to the reader, it may be useful to summarize the results of this integration for each endpoint in each reach or area where potentially toxic concentrations were found using the following table.

**Table 1. Summary of integration of single chemical toxicity**

<b>Issue</b>	<b>Result</b>
Taxa affected at ambient concentrations	<i>List species or higher taxa and life stages and proportion of tested species</i>
Severity of effects at ambient concentrations	<i>List types and magnitudes of effects</i>
Spatial extent of toxic concentrations	<i>Meters of stream, square meters of land, etc.</i>
Frequency of toxic concentrations	<i>Proportion of time or number of distinct episodes</i>
Association with source	<i>Spatial and temporal relationships to hypothesized sources</i>
Estimated effect	<i>Summarize the expected nature and extent of effects and credible upper bounds effects</i>
Confidence in results	<i>Rating and supporting comments</i>

## **2.1 FISH**

### **2.1.1 Aqueous Exposure**

Fish are exposed primarily to contaminants in water. Contaminants in water may come from upstream aqueous sources including waste sites, other anthropogenic sources, and background; exchange of materials between the surface water and contaminated sediments; or exchange of contaminants between the biota and the water column. The consensus of the scientific community and of the U.S. Environmental Protection Agency (EPA) Office of Water is that aquatic biota should be assumed to be exposed to the dissolved fraction of the chemicals in water because that is the bioavailable form (HECD 1992, Prothro 1993). However, EPA Region IV prefers to use total concentrations as conservative estimates of the exposure concentration.

The reader should note that use of total concentrations is not always conservative. First, because the high levels of acid extractable metals may cause analytical interferences, the limits of detection may be greater for total concentrations, and toxic concentrations of metals may not be detected. This apparently occurred in the Bear Creek Remedial Investigation where copper was a COPEC in filtered samples but was not detected in total samples. Second, when comparing to background, if the dissolved concentration is small relative to total, there may be a significant increase in dissolved concentrations relative to background but no significant increase for total. Therefore, dissolved phase concentrations of metals are used to provide a best estimate of risk and total concentrations are used to satisfy regional regulators.

Because water in a reach is likely to be more variable in time than space, due to the rapid replacement of water in flowing systems and the lack of spatial gradients in the ponds that occur on waste sites, the mean water concentration within a reach or subreach is an appropriate estimate of the chronic exposure experienced by fishes. The upper 95% confidence bound on the mean is usually an appropriately conservative estimate of this exposure for use in the contaminant screening. Note, however, that if episodes of high exposures are known to occur, those episodic concentrations should be screened as well. In any case, the full distribution of observed concentrations is used to estimate risks.

Some fish spend most of their lives near the sediment and the eggs and larvae of some fish (particularly sunfish and black bass) develop at the sediment-water interface. These epibenthic species and life stages may be more highly exposed to contaminants than is suggested by analysis of samples from the water column. If available, water samples collected just above the sediments provide an estimate of this exposure. Alternately, the estimated or measured sediment pore water concentrations may be used as a conservative estimate of this exposure.

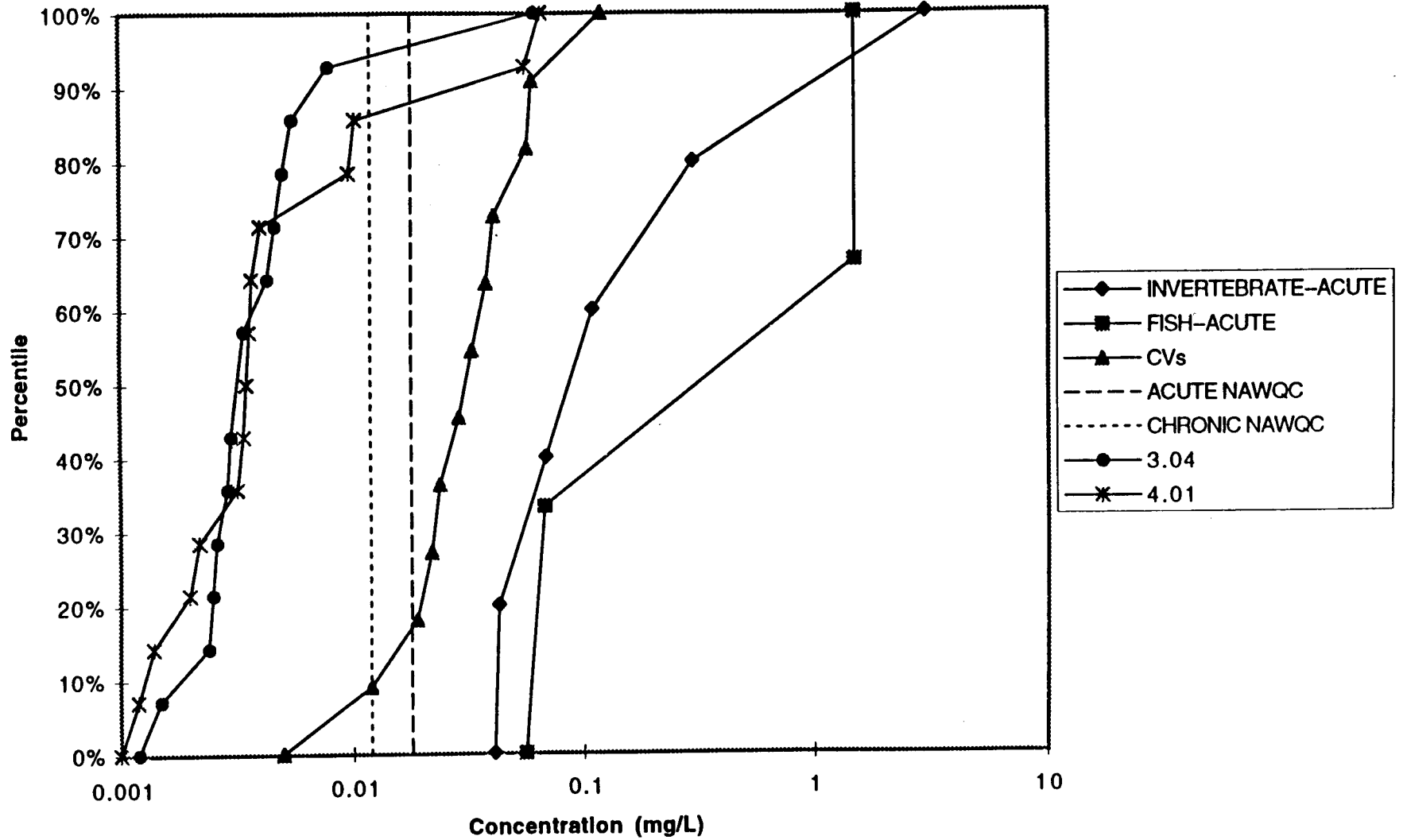
The screening benchmarks for aquatic biota are taken from Suter and Tsao (1996). Sets of alternative benchmarks were calculated for each chemical. The benchmark preferred by the regulatory agencies is the chronic National Ambient Water Quality Criteria (NAWQC), but they are available for relatively few industrial chemicals. Secondary chronic values (SCV), which are conservative estimates of chronic NAWQC, were calculated for chemicals that do not have NAWQC. Other benchmarks are included to provide greater assurance of detecting all COPECs.

NAWQC that are functions of water hardness must be corrected for site-specific conditions. For purposes of screening, choose conditions that would constitute reasonable maximum toxicity, defined as conditions that would persist for 7 days. For the Clinch River and Poplar Creek, this is a hardness of approximately 100 mg/L. However, for other waters such as Bear Creek, other hardness levels are appropriate.

Toxicity profiles must be prepared and presented for all COPECs. Toxicity profiles summarize the existing toxicity information for each chemical including concentrations causing acute lethality and chronic lethal and sublethal effects and physical-chemical conditions that modify toxicity. For chemicals with hardness-dependent criteria, test endpoints must be corrected as described previously. Note that, for two reasons, toxicity to aquatic invertebrates as well as fish should be included in the toxicity profile and in the risk characterization. First, invertebrates are the primary food for fish in freshwater communities, so an effect on invertebrates would be expected to result in indirect effects on fish. Second, fish and invertebrates are assumed by the EPA to have approximately the same distribution of sensitivity to toxic effects. Therefore, some fish species are likely to experience toxic effects at any concentration that affects an invertebrate.

The aqueous toxicity data from the toxicity profiles and the aqueous chemical concentrations should be used to present distributions of exposure and effects. For exposure of fish and other aquatic organisms to chemicals in water, the exposure distributions are allocations of aqueous concentrations over time and the effects distributions are allocations of sensitivities of species to acutely lethal effects (e.g., LC50s) and chronically lethal or sublethal effects (CVs). If the water samples were collected in an temporally unbiased design (preferably stratified random or random), overlap of these two distributions indicates the approximate proportion of the time when aqueous concentrations of the chemical are acutely or chronically toxic to a particular proportion of aquatic species. For example, 10% of the time copper concentrations in Reach 4.01 are at levels chronically toxic to approximately half of aquatic animals (Fig. 3). Interpretation of this result depends on

## Copper in Surface Water



**Fig. 3. Empirical distribution functions for acute toxicity (LC50 and EC50 values) and chronic toxicity (Chronic Values) of copper to fish and aquatic invertebrates and for individual measurements of copper in surface water from two reaches. Vertical lines are acute and chronic national ambient water quality criteria.**



knowledge of the actual temporal dynamics of the exposures and effects. For example, 10% of a year is 36 days which would cover the entire life cycle of a planktonic crustacean or the entire embryo-larval stage of a fish, so significant chronic effects are clearly possible. However, if the 36 days of high concentrations is associated with a number of episodes, the exposure durations are reduced. The 7-day duration of the standard EPA subchronic aqueous toxicity tests could be taken as an approximate lower limit for chronic exposures, so the proportion of the year with high copper concentrations could be divided into five episodes and still induce significant chronic effects on a large proportion of species. More precise interpretations would require knowledge of the actual duration of episodes of high concentrations and of the rate of induction of copper effects on sensitive life stages.

### **2.1.2 Fish Body Burdens**

Although nearly all toxicity data for fishes are expressed in terms of aqueous concentrations, fish body burdens potentially provide an exposure metric that is more strongly correlated with effects (McCarty and Mackay 1993). This is particularly likely to be the case for chemicals that bioaccumulate in fish and other biota to concentrations greater than in water. For such chemicals, dietary exposure may be more important than direct aqueous exposures, and concentrations that are not detectable in water may result in high body burdens in fish. Three contaminants that have been determined to be COPECs in screening ecological assessments of Oak Ridge sites accumulate in that manner: mercury, polychlorinated biphenyls (PCBs), and selenium. Since the individual body burden measurements correspond to an exposure level for an individual fish, the maximum value is used for screening purposes and the risk estimate is based on the distribution of individual observations for each measured species. Measurements may be performed on muscle (fillet), carcass (residue after filleting), or whole fish. Since whole fish measurements are most commonly used in the literature, whole fish concentrations either measured directly or reconstructed from fillet and carcass data should be used to estimate exposure.

No standard benchmarks for effects on fish of internal exposures are available. The body burdens associated with effects in toxicity tests and field studies and body burdens found at other sites should be presented in the toxicity profiles. To be consistent with EPA practices in calculating CVs, thresholds for toxic effects can be expressed as geometric means of body burdens measured at the no observed effects concentration (NOEC) and lowest observed effects concentration (LOEC). However, other expressions that are more clearly related to effects may also be used.

Few data are available for fish toxicity based on body burdens; therefore, calculation of an HQ is usually sufficient integration of exposure and effects. However, the relevance of the data needs to be carefully reviewed and described.

## **2.2 BENTHIC INVERTEBRATES**

Two different expressions of sediment contamination may be used: whole sediment concentrations and pore water concentrations. The use of pore water is based on the assumption that chemicals associated with the solid phase are largely unavailable and therefore sediment toxicity can be estimated by measuring or modeling the pore water concentration. EPA uses this approach to calculate sediment quality criteria. Whole sediment concentrations do not account for effects of sediment properties on bioavailability. However, they are required by EPA Region IV and may provide a better estimate of risk for highly particle-associated chemicals.

For purposes of screening chemicals, the appropriate estimate of exposure is a concentration that protects the most exposed organisms. Because benthic invertebrates are relatively immobile and inhabit a medium that changes little over time, the maximum concentration is used. For risk estimation, the estimate of exposure of the community is the percentage of samples exceeding particular effects levels.

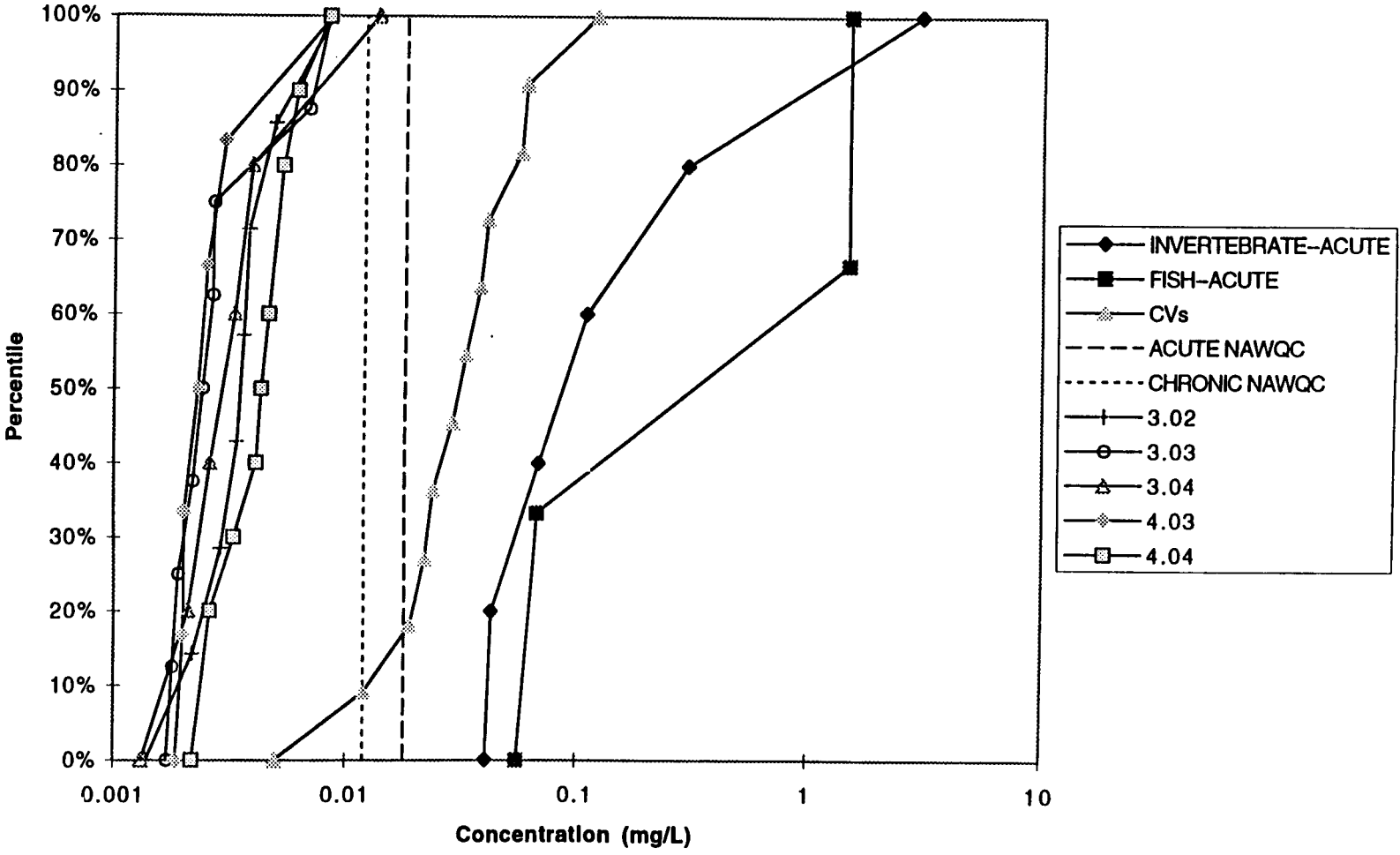
Sets of alternative sediment benchmarks were derived for each chemical (Jones et al. 1996). Whole sediment concentrations are compared to these alternative benchmarks. Pore water concentrations are compared to the ecotoxicological benchmarks for aquatic biota. The use of multiple benchmarks provides greater assurance of detecting all COPECs. Sediment quality criteria are corrected for site-specific conditions. Sediment benchmarks derived using the equilibrium partitioning method are calculated using location-specific percent organic carbon. As in the aqueous chemical screening, hardness-dependent criteria are corrected to site-specific hardness. For Clinch River and Poplar Creek, that value was 100 mg/L hardness based on the pore water mean across reaches of 134 mg/L and range of 82 to 160 mg/L.

For each COPEC, the distribution of observed concentrations in whole sediment and pore water is compared to the distributions of effective concentrations. In the case of exposure of benthic invertebrates to sediment pore water, the exposure distributions are interpreted as distributions over space since sediment composition varies little over the period in which samples were collected, but samples were distributed in space within reaches. The effects distributions are the same as for surface water—distributions of species sensitivities in acute and chronic aqueous toxicity tests. Therefore, overlap of the distributions indicates the proportion of locations in the reach where concentrations of the chemical in pore water are acutely or chronically toxic to a particular proportion of species. For example, copper concentrations in sediment pore water from more than 90% of locations in Reach 4.04 are below chronically toxic concentrations for more than 90% of aquatic animal species (Fig. 4). If the samples are collected by random or some other equal probability sample, these proportions can be interpreted as proportions of the area of the reach. Therefore, an alternate expression of the result is that less than 10% of the reach is estimated to be toxic to as much as 10% of benthic species.

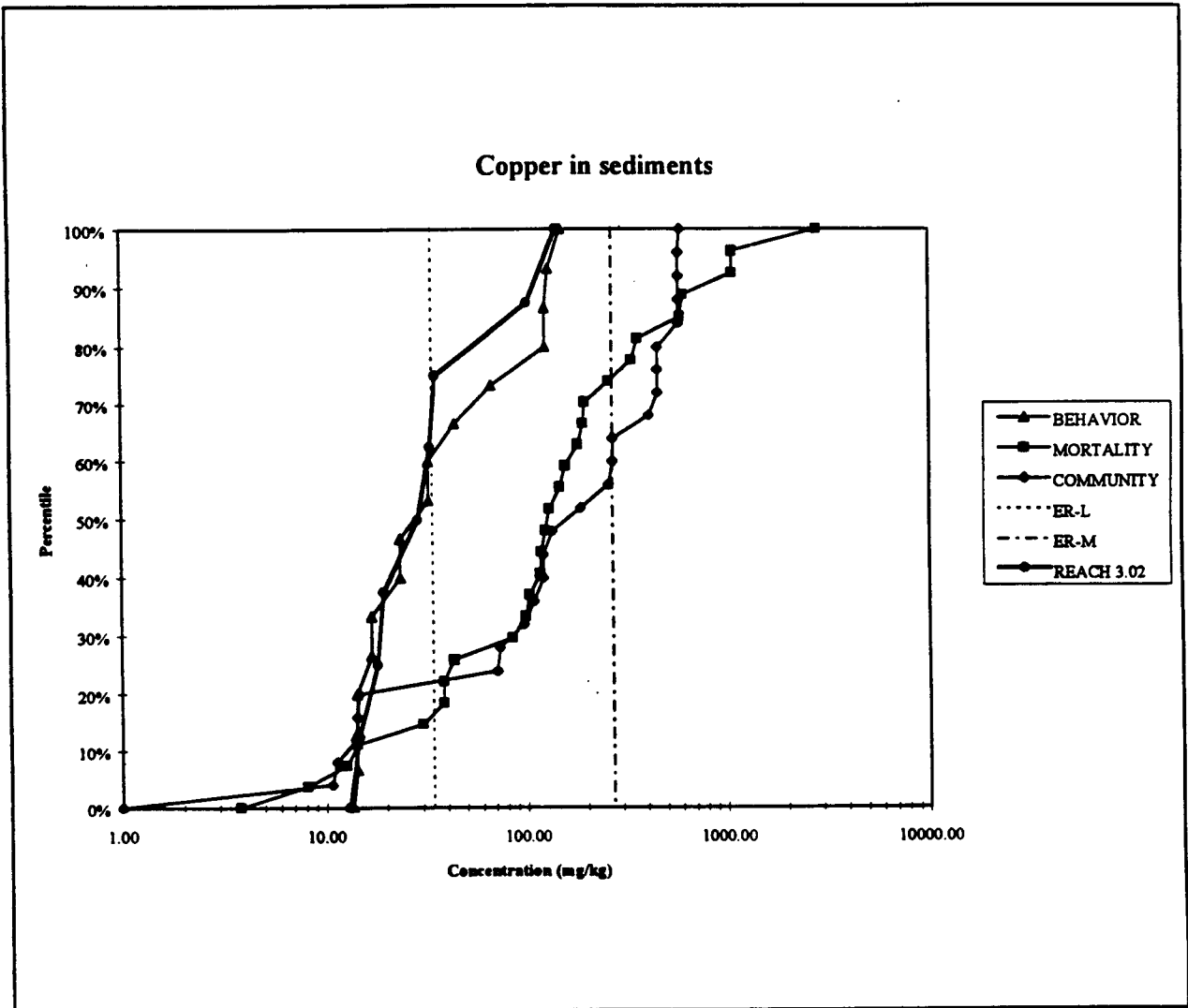
In the case of exposure of benthic invertebrates to chemicals in whole sediment, the exposure distributions are, as with pore water, distributions in space within reaches. If sufficient data are available, three effects distributions are presented for each sediment COPEC: a distribution of concentrations reported to be thresholds for reductions in benthic invertebrate community parameters in various locations, a distribution of concentrations reported to be thresholds for lethal effects in toxicity tests of various sediments, and a distribution of concentrations reported to be thresholds for behavioral effects in toxicity tests of various sediments.

If we assume that the effects data set are drawn from studies of a random sample of sediments so that the site sediments can be assumed to be a random draw from the same distribution, and if we assume that the reported community effects correspond to the community effects defined in the assessment endpoint, then the effects distributions can be treated as distributions of the probability that the chemical causes significant toxic effects on the endpoint at a given concentration. Overlap of the exposure and effects distributions represents the probability of significant alteration in the benthic communities at a given proportion of locations in a reach. For example, copper concentrations in whole sediment from half of locations in Reach 3.02 of Poplar Creek are above the concentration at which there is approximately a 20% likelihood of effects on community composition (Fig. 5).

### Copper in Pore Water



**Fig. 4. Empirical distribution functions for acute toxicity and chronic toxicity of copper to fish and aquatic invertebrates and for individual measurements of copper in sediment pore water from five reaches. Vertical lines are acute and chronic national ambient water quality criteria.**



**Fig. 5. Empirical distribution functions for toxicity of copper in sediment to sediment-associated organisms and for individual measurements of copper in sediment. Effects are reported thresholds for effects on behavior, survival, and community structure from MacDonald (1994). Vertical lines are National Oceanographic and Atmospheric Administration (NOAA) effects range low (ER-L) and effects range median (ER-M) values.**

The other two effects curves are not direct estimates of the endpoint, but they provide independent supporting evidence. Copper concentrations in whole sediment from half of locations in Reach 3.02 of Poplar Creek are above the concentration at which there is approximately a 50% likelihood of behavioral effects on benthic invertebrates and a 15% likelihood of lethal effects.

### **2.3 TERRESTRIAL PLANTS, SOIL INVERTEBRATES, AND MICROBIAL PROCESSES**

Exposures to organisms rooted in or inhabiting soil are expressed as whole soil concentrations. For screening purposes, the maximum observed surface soil concentration is appropriate because these organisms are essentially immobile, so some organism will occupy and complete its life cycle at the most contaminated point. These concentrations are compared to soil screening benchmarks for plants, invertebrates, and microbial processes (Will and Suter 1995a&b).

For estimation of risks, the distribution of observed concentrations should be used. These should be compared to the distributions of effective concentrations for plants, invertebrates, and microbial processes. The exposure distributions are interpreted as distributions over space since soil composition varies little over the period in which samples were collected, but samples and contaminants are distributed in space within areas. For plants, the effects distributions are distributions of species-soil combinations in soil toxicity tests. If we assume that site soils are drawn from the same distribution of soil properties as the tested soils and that test plants have the same sensitivity distribution as site plants, then the threshold concentrations for effects on site plants can be assumed to be drawn from the distributions of threshold concentrations for effects on plants in the toxicity tests. Therefore, overlap of the distributions indicates the proportion of locations in an area where concentrations of the chemical are expected to be toxic to a particular proportion of species in the site community.

Assumptions for invertebrates are similar except that earthworms are assumed to be representative of soil invertebrates and the distribution of test earthworm species sensitivities are assumed to be representative of site earthworms. The distributions of concentrations toxic to microbial processes are not quite equivalent because some processes are carried out by individual microbial strains and others are carried out by entire microbial communities. Therefore, the distributions of effects data are allocations of process-soil combinations rather than species-soil combinations. The overlap of the distributions indicates the proportion of locations in an area where concentrations of the chemical are expected to be toxic to a particular proportion of microbial processes.

### **2.4 SOURCES OF VARIANCE IN EFFECTS DISTRIBUTIONS**

The reader should note that an implicit assumption has not been adequately recognized in discussions of the use of sensitivity distributions in ERAs. The distributions include not only biological variance but also physical variance and methodological variance. The conventional interpretation that these distributions are distributions of species sensitivity or community sensitivity is based on the assumption that biological variance is dominant. For aqueous toxicity that is probably true. The test methods and endpoints for aquatic toxic effects are reasonably consistent, so methodological variance should be relatively low. In addition, variance in test water chemistry is relatively low, particularly when hardness is normalized for metals, so physical variance should be relatively low. However, for both sediments and soils, the testing and survey methods and the endpoints are highly variable, the media have highly variable texture and chemistry, and normalization methods are not available. Therefore, the physical and methodological variances may be significant contributors to the effects distributions in sediments and soils. The methodological

variance is extraneous. The physical variance is an actual property of soils and sediments; it could be considered extraneous as well.

Conversely, if one takes an ecosystem perspective, then the distributions resulting from the combination of biological and physical variance can be thought of as distributions of benthic ecosystem sensitivity, soil-plant system sensitivity, etc.; that concept is adopted in the interpretations presented previously in this report, and methodological variance is assumed to be small relative to the combined biological and physical variance. However, it would be highly desirable to disaggregate those sources of variance by standardizing methods and developing methods to normalize soils and sediments.

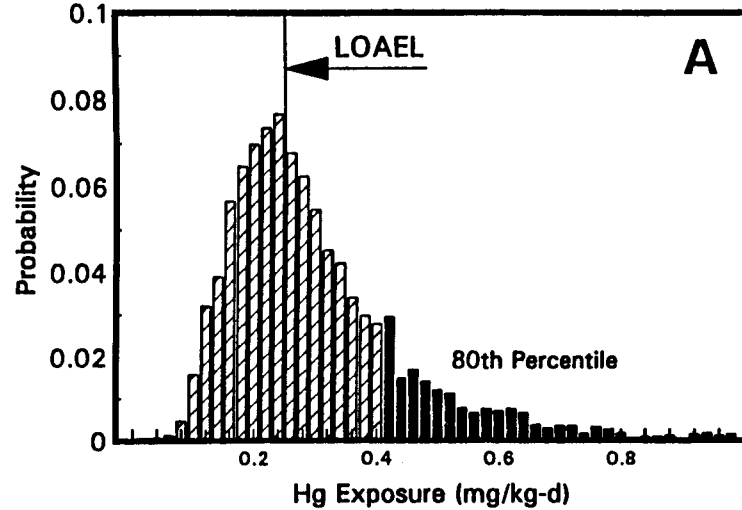
## **2.5 WILDLIFE**

Exposure of piscivorous wildlife to contaminants may be expressed as the rate of ingestion of contaminated food, soil, and water. First, conservative exposure estimates should be generated using point-estimates of exposure parameters and conservative assumptions. These conservative estimates are compared to no observed adverse effects levels (NOAELs) to identify COPECs (Sample et al. 1996). More realistic estimates of exposure at each reach or area should then be generated for the COPECs using Monte Carlo simulation in place of conservative assumptions. The Monte Carlo simulations incorporate the variability in the contaminant concentrations and use more realistic foraging and life history data. Exposure distributions generated by these distributions are compared to NOAELs and lowest observed adverse effects levels (LOAELs) to identify reaches or areas where exposure is sufficiently high to present a risk. Finally, Monte Carlo simulation of exposure from adjacent reaches or areas may be performed to address how foraging behavior and movements of wildlife influence contaminant exposure.

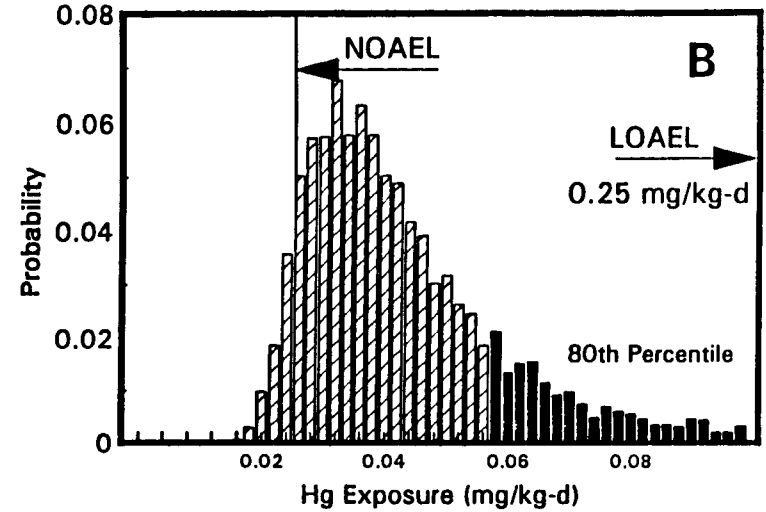
The interpretation of these distributions depends on whether risks to populations or individuals are being assessed. In the case of wildlife populations exposed to chemicals in food and water, the exposure distributions are allocations of total intake rate of the chemical across individuals in the populations based on the distributions of observed concentrations in water and various food items. If we assume that the members of a population occurring in each modeled area independently sample the water and food items over the entire area, then the proportions of the exposure distributions represent estimates of the proportion of a population receiving a particular intake rate. In keeping with practice in wildlife toxicology, the effects are treated as point estimates (i.e., NOAELs and LOAELs). Therefore, the intersection of these points with the exposure distributions represents estimates of the proportion of the exposed population with exposure levels less than the NOAEL and LOAEL. For example, <1% of the members of a rough-winged swallow colony located in Reach 3.02 would receive a dose rate greater than the LOAEL but approximately half of those in Reach 3.01 would (Fig. 6).

The case of threatened and endangered wildlife is a little different from that of other wildlife species because risks to these species are assessed at the organism level. Therefore, the exposure models represent exposure of individuals within their foraging areas, and the exposure distribution is a distribution of the likelihood of various exposure rates to individuals foraging in the specified areas. Effects are specified by the same point estimates as with other wildlife, so the intersection of the point estimates of effects with the exposure distributions represents the likelihood that an individual will have an intake rate less than the no observed effects level (NOEL) and lowest observed effects level (LOEL). For example, the likelihood that the osprey foraging exclusively in Reach 3.02 of Poplar Creek will receive a dose rate greater than the LOEL is greater than 50% (Fig. 7).

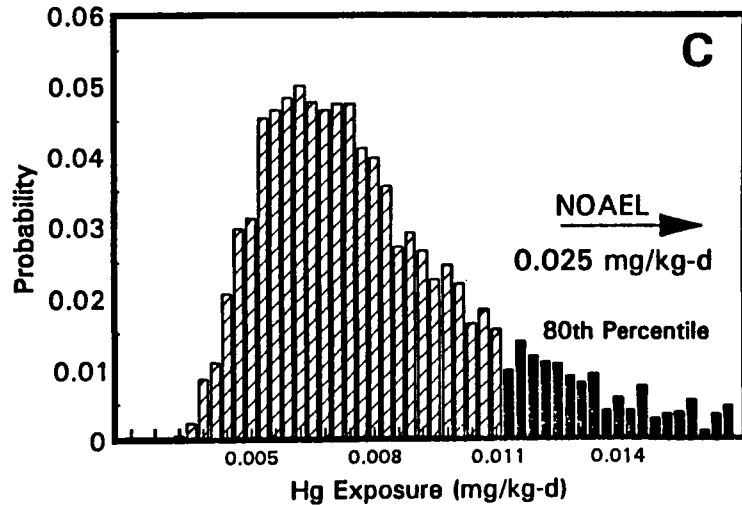
Contaminant = Mercury  
Location = Subreach 3.01, Poplar Creek, EFPC



Contaminant = Mercury  
Location = Subreach 3.02, Poplar Creek-Mitchell Br



Contaminant = Mercury  
Location = Subreach 3.03, Poplar Creek-CNF



Contaminant = Mercury  
Location = Subreach 3.04, Poplar Creek-Mouth

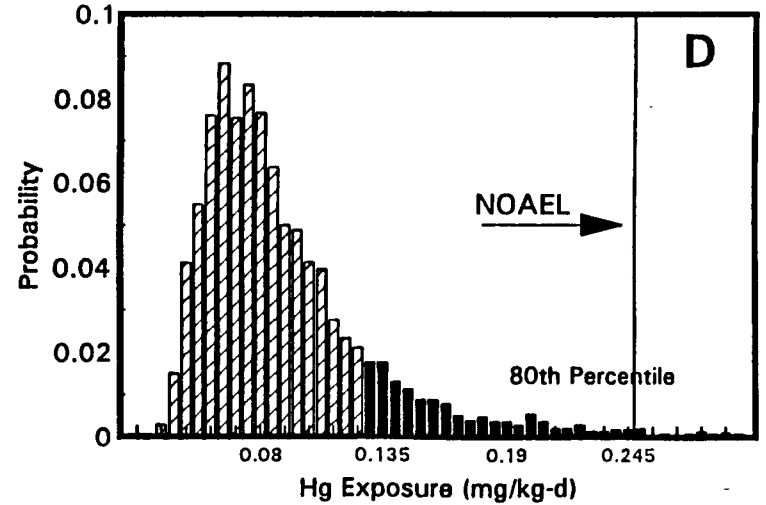


Fig. 6. Distributions of estimated mercury exposures for rough-winged swallows in Poplar Creek subreaches.

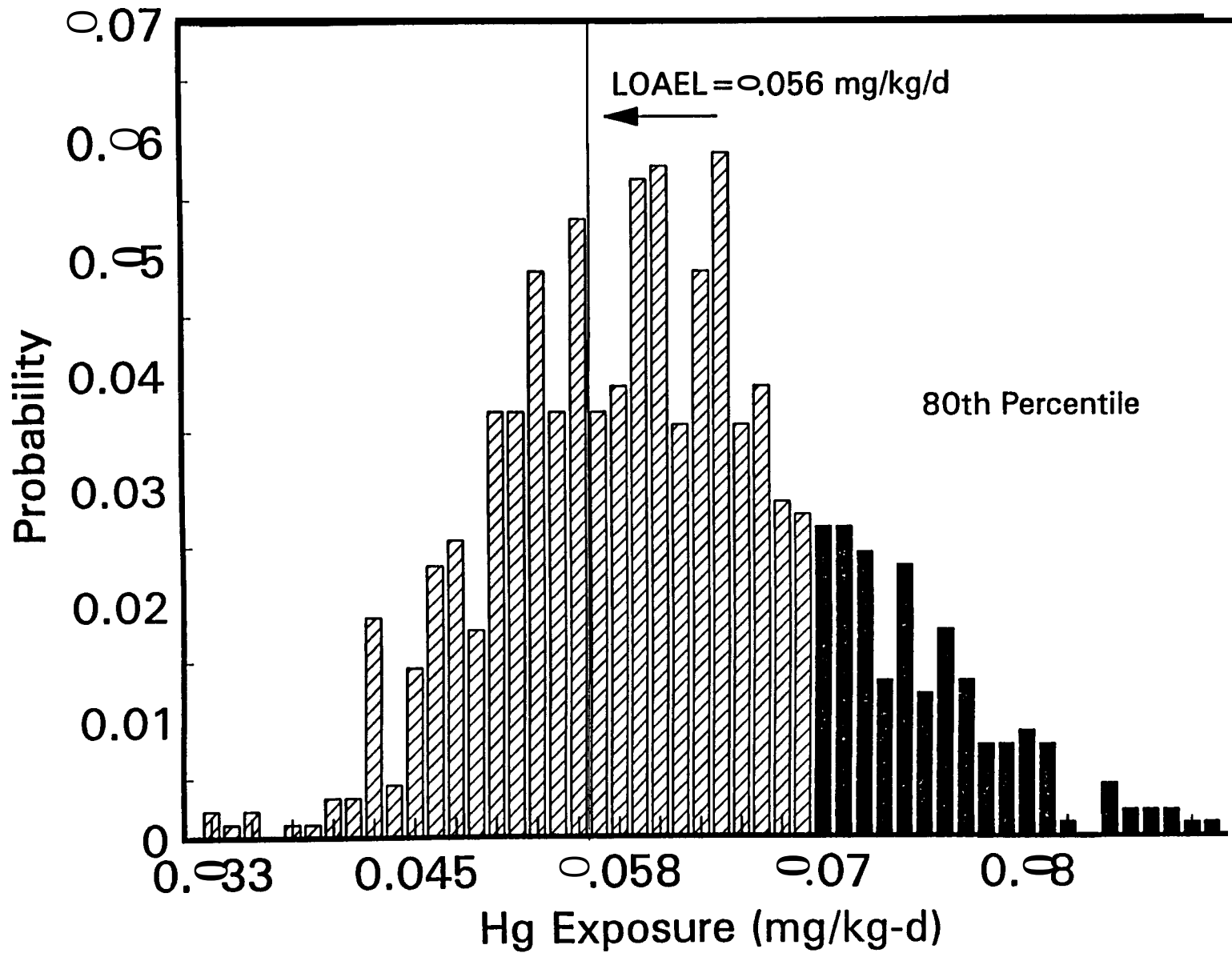


Fig. 7. Distributions of estimated mercury exposures for osprey in Poplar Creek subreach 3.02.



One must clearly understand the nature of the variances in the input parameters to ensure that the interpretations discussed in this section are correct. For example, the variance in body weight among individual mink or the proportion of fish in their diet would be appropriate to include if the intent is to estimate the proportion of individual mink receiving a particular dose. If variation due to ignorance or uncertainty about fundamental processes such as uptake efficiency were included, the distributions would no longer reflect the distribution of exposure among individuals so the distributions could not be interpreted as the proportions of individuals exposed to a particular dose. In that case, the probabilities drawn from the distributions would be best considered credibilities of a particular dose given both variation among individuals and scientific uncertainties.

### **3. AMBIENT MEDIA TOXICITY TESTS**

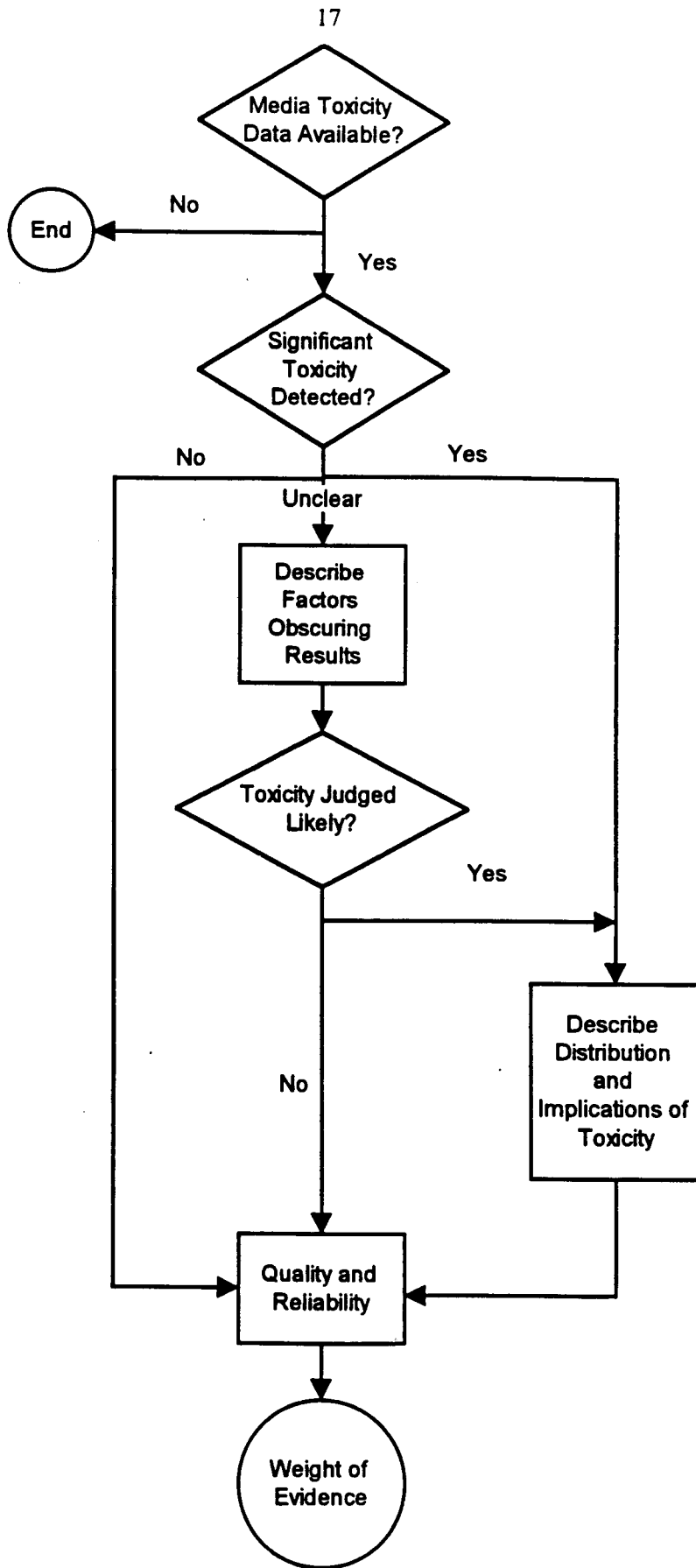
Risk characterization for this line of evidence begins by determining whether the tests show significant toxicity (Fig. 8). Toxicity is not significant if the effects relative to controls are less than 20% (e.g., less than 20% mortality) and the effects are not statistically significantly different from controls. That is, effects are considered significant if (1) the hypothesis of no difference between responses in contaminated media and in either reference media or control media is rejected with 95% confidence (i.e., statistical significance) or (2) an effect of 20% or greater in survival, growth, or reproduction relative to either reference media or control media is observed (i.e., biological significance).

If no significant toxicity was found, the risk characterization consists of determining the likelihood that the result constitutes a false negative. False negatives could result from not collecting samples from the most contaminated sites or at times with the highest contaminant levels, handling the samples in a way that reduced toxicity, or using tests that are not sufficiently sensitive to detect effects that would cause significant injuries to populations or communities in the field.

If significant toxicity occurs in the tests, the risk characterization should describe the nature and magnitude of the effects and the consistency of effects among tests conducted with different species in the same medium.

Toxicity tests may produce ambiguous results in some cases because of poor performance of organisms in control media (e.g., may be caused by diseases, background contamination, inappropriate reference or control media, or poor performance of the test protocol). In such cases, expert judgment by the assessor in consultation with the individuals who performed the test should be used to arrive at the best interpretation of the test results.

If significant toxicity is found at any site, then the relationship of toxicity to exposure must be characterized. The first way to do this is to examine the relationship of toxicity to concentrations of chemicals in the media. The manner in which this is done will depend on the amount of data available. If numerous toxicity tests are available, the level of effects or the frequency of tests showing toxic effects could be defined as a function of concentrations of one or more COPECs. For example, if fathead minnow larvae experienced more than 20% mortality in one or more tests, the percent mortality could be plotted against the concentration of each of the COPECs in water samples collected in conjunction with the test. If there is a positive relationship, an appropriate statistical model should be fit to the points. If multiple chemicals with the same mode of action may be jointly contributing to toxicity, the aggregate concentration [e.g., total polyaromatic hydrocarbons (PAHs)]



**Fig. 8. Risk characterization based on toxicity testing of ambient media.**

could be used as the independent variable. In general, if toxicity is occurring, then it should be possible to identify exposure-response relationships. However, there are a number of reasons why a true causal relationship between a chemical and a toxic response may not be apparent (Table 2). Therefore, the lack of an exposure-response relationship does not disprove that one or more of the COPECs caused an apparent toxic effect.

**Table 2. Why contaminant concentrations in ambient media may not be correlated with toxicity of those media**

---

Variation in bioavailability

Due to variance in medium characteristics

Due to variance in contaminant age among locations (contaminants added to soil and sediments may become less bioavailable over time due to sequestration)

Due to variance in transformation or sequestration rates among locations

Variation in the form of the chemical (e.g., ionization state)

Variation in concentration over time or space (i.e., samples for analysis may not be the same as those tested)

Spatial heterogeneity

Temporal variability (e.g., aqueous toxicity tests last for several days but typically water from only one day is analyzed)

Variation in composition of the waste (concentrations of components of the waste other than the individual COPEC that is believed to be the principle toxicant may vary over space and time thereby obscuring the relationship)

Variation in co-occurring contaminants (concentrations of contaminants from upstream sources may vary over time)

Inadequate detection limits (Even if the chemicals are detected when toxic, no correlation will be found if they are not detected when there is low or no toxicity)

Inherent variation in toxicity tests

Variation in toxicity test due to variance in medium characteristics (e.g., hardness, organic matter content, and pH)

---

An alternative and potentially complementary approach to relating toxicity to exposure is to determine the relationship between the occurrence of toxicity and sources of contaminants (e.g., springs, seeps, tributaries, spills) or of diluents (i.e., relatively clean water or sediments). This may be done by simply creating a table of potential sources of contamination or dilution and indicate for each test whether toxicity increases, decreases, or remains unchanged below that source. The same information may be conveyed graphically. For a stream or river, toxicity may be plotted as a function of reach (if reach boundaries are defined by source locations) or distance downstream (with locations of sources marked on the graph) (Fig. 9).

Finally, when sources of toxic water have been identified, and tests have been performed on dilution series of those waters, the transport and fate of toxicity can be modeled like that of individual chemicals (DiToro et al. 1991). Such models of toxicity can be used to explain ecological degradation observed in streams and apportion causation among sources.

To facilitate the weight-of-evidence analysis and make the bases clear to the reader, it may be useful to summarize the results of this integration for each reach or area where significant toxicity was found using information from Table 3.

# Medaka Survival

Proportion significantly reduced Mean proportional reduction

Proportion

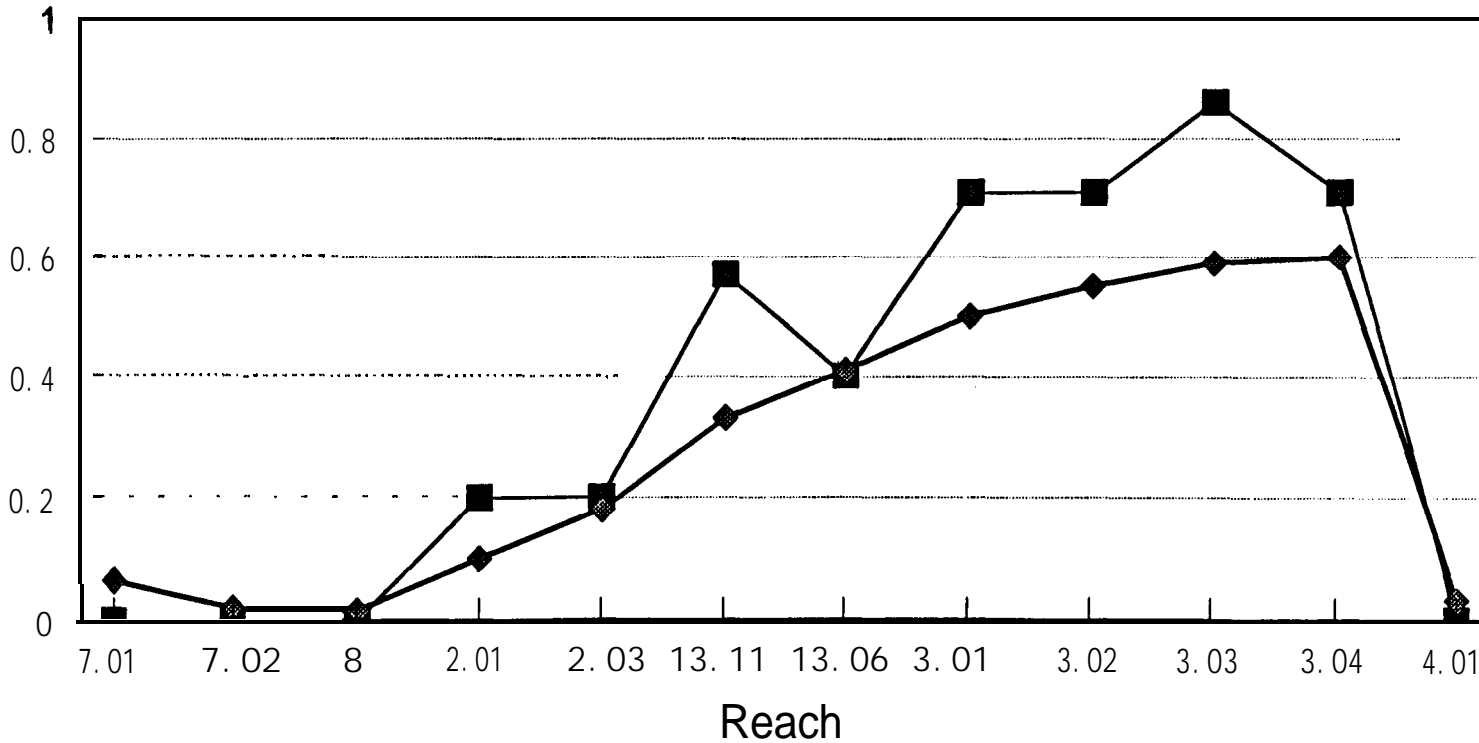


Fig. 9. Mean proportional reduction in survival of medaka eggs in ambient water relative to controls and proportion of tests in which survival was significantly reduced (i.e., either reduced at least 20% or statistically significantly reduced) in numbered reaches and subreaches.

**Table 3. Summary table for integration of ambient media toxicity test results**

<b>Issue</b>	<b>Result</b>
Species affected	<i>List species and life stages affected in the tests</i>
Severity of effects	<i>List types and magnitudes of effects</i>
Spatial extent of effects	<i>Meters of stream, square meters of land, etc. for which media samples were toxic</i>
Frequency of effects	<i>Proportion of time or number of distinct episodes</i>
Association with source or chemical	<i>Spatial and temporal relationship to hypothesized sources</i>
Estimated effect	<i>Summarize the nature and extent of estimated toxic effects and credible upper bounds</i>
Confidence in results	<i>Rating and supporting comments</i>

#### 4. BIOLOGICAL SURVEYS

If biological survey data are available for an endpoint species or community, then the first question to be answered is whether the data suggest that significant effects are occurring (Fig. 10). For some groups, notably fish and benthic invertebrates, there are abundant data from reference streams for comparison. For most other endpoint groups, references must be established *ad hoc* and the lack of temporal or spatial replication may make inference tenuous. For some taxa such as most birds, traditional survey data are not useful for estimating risks from wastes because mobility, territoriality, or other factors obscure demographic effects. However, survey results may be more reliable if efforts are made to control extraneous variance such as by setting out nest boxes on contaminated and reference sites to monitor reproductive success.

Care must be taken to consider the sensitivity of field survey data to toxic effects relative to other lines of evidence. Some biological surveys are very sensitive (e.g., surveys of nesting success of colonial nesting birds or electrofishing surveys of wadeable streams), others are moderately sensitive (e.g., benthic macroinvertebrates), and still others are insensitive (e.g., fish community surveys in large reservoirs and small mammal surveys). However, even relatively insensitive surveys may be quite useful in assessments. For example, if the concentrations of chemicals suggest that a medium should be highly toxic but toxicity tests of the medium find no toxicity, then even a relatively insensitive survey that found a community that was not highly modified would tend to confirm that the chemical analyses were misleading and the toxicity test data were correct (e.g., the chemical was not in a bioavailable form or consisted of a less toxic species).

Conversely, a highly modified community in the absence of high levels of analyzed chemicals would suggest that combined toxic effects, toxic levels of unanalyzed contaminants, or episodic contamination had occurred. However, field surveys interpreted in isolation without supporting data could be misleading, particularly if the absence of statistically significant differences were interpreted as an absence of effects.

If biological survey data are consistent with significant reductions in abundance, production, or diversity, associations of apparent effects with causal factors must be examined. First, the distribution of apparent effects in space and time must be compared to the distribution of sources

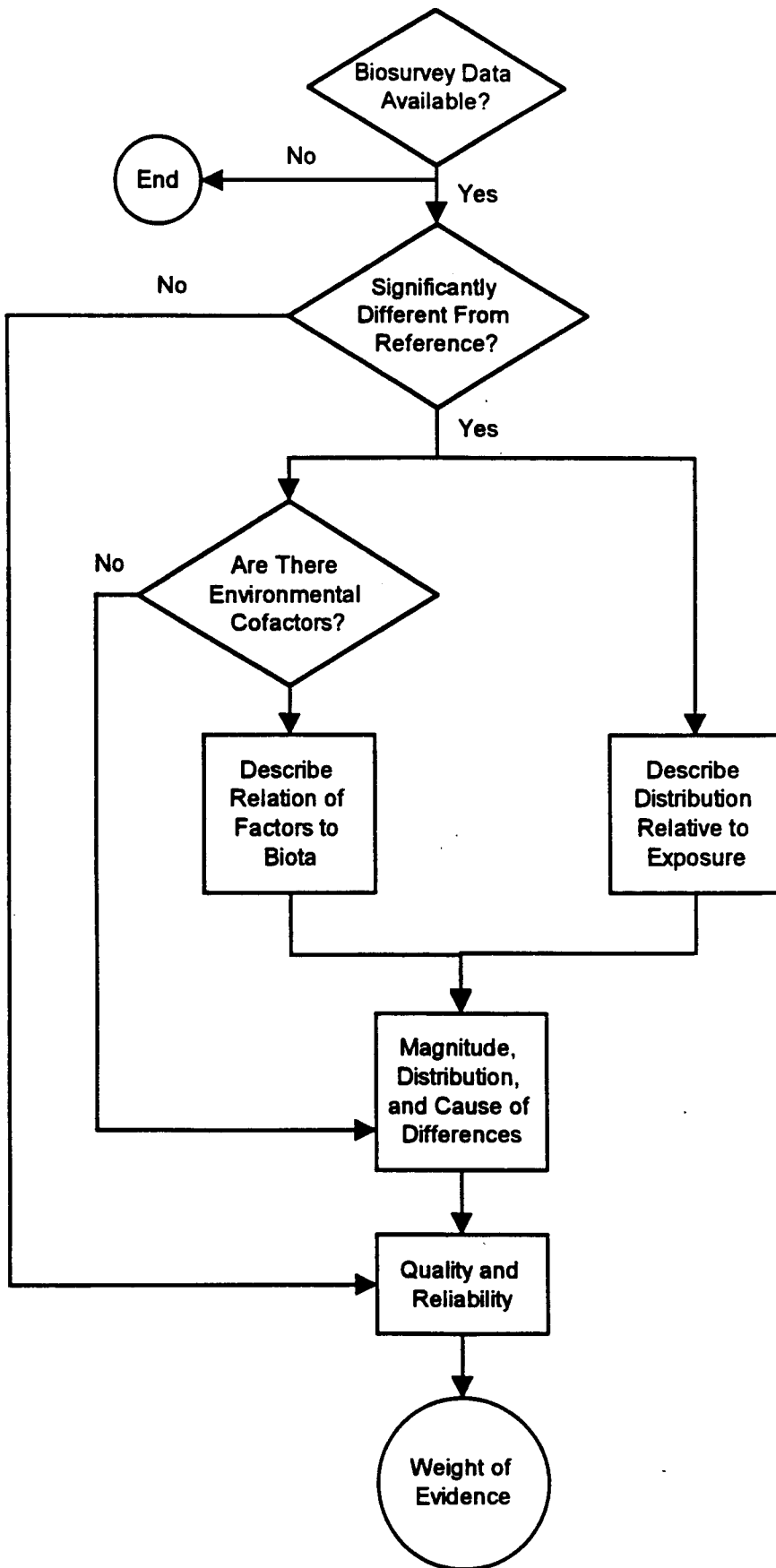


Fig. 10. Risk characterization based on biological survey data.

or of contaminants. Second, the distribution of apparent effects must be compared to the distribution of habitat factors that are likely to affect the organisms in question such as stream structure and flow. Finally, the natural variability of the endpoint populations and communities and the accuracy of the survey methods must be examined to estimate the likelihood that the apparent effects are due to chance.

To facilitate the weight-of-evidence analysis and make the bases clear to the reader, it may be useful to summarize the results of this integration for each reach or area using the following table.

**Table 4. Summary table for integration of biological survey results**

<b>Issue</b>	<b>Result</b>
Taxa and properties surveyed	<i>List species or communities and measurement endpoints</i>
Nature and severity of effects	<i>List types and magnitudes of apparent effects</i>
Minimum detectable effects	<i>For each measurement endpoint, define the smallest effect that could have been distinguished from reference</i>
Spatial extent of effects	<i>Meters of stream, square meters of land, etc.</i>
Number and nature of reference sites	<i>List and describe reference sites including habitat differences</i>
Association with habitat characteristics	<i>Describe any correlations or qualitative associations of apparent effects with habitat variables</i>
Association with source or chemical	<i>Describe any correlations or qualitative associations of apparent effects with sources or chemical concentrations</i>
Most likely cause of apparent effects	<i>Based on the associations described in previous items, present the most likely cause of the apparent effects</i>
Estimated effects	<i>Summarize the nature and extent of estimated toxic effects and credible upper bounds</i>
Confidence in results	<i>Rating and supporting comments</i>

## 5. BIOMARKERS

Biomarkers are physiological or biochemical measures, such as blood cholinesterase concentration, that may be indicative of exposure to contaminants. They are seldom useful for estimating risks by themselves, but they can be used to support other lines of inference. The inference begins by asking if the levels of the biomarkers significantly differ from those at reference sites (Fig. 11). If they do, then it is necessary to determine whether they are diagnostic or at least characteristic of any of the COPECs or of any of the habitat factors that are thought to affect the endpoint biota. If the biomarkers are characteristic of contaminant exposures, then the distribution and frequency of elevated levels must be compared to the distributions and concentrations of contaminants. Finally, to the extent that the biomarkers are known to be related to overt effects such as reductions in growth, fecundity, or mortality, the implications of the observed biomarker levels for populations or communities should be estimated.

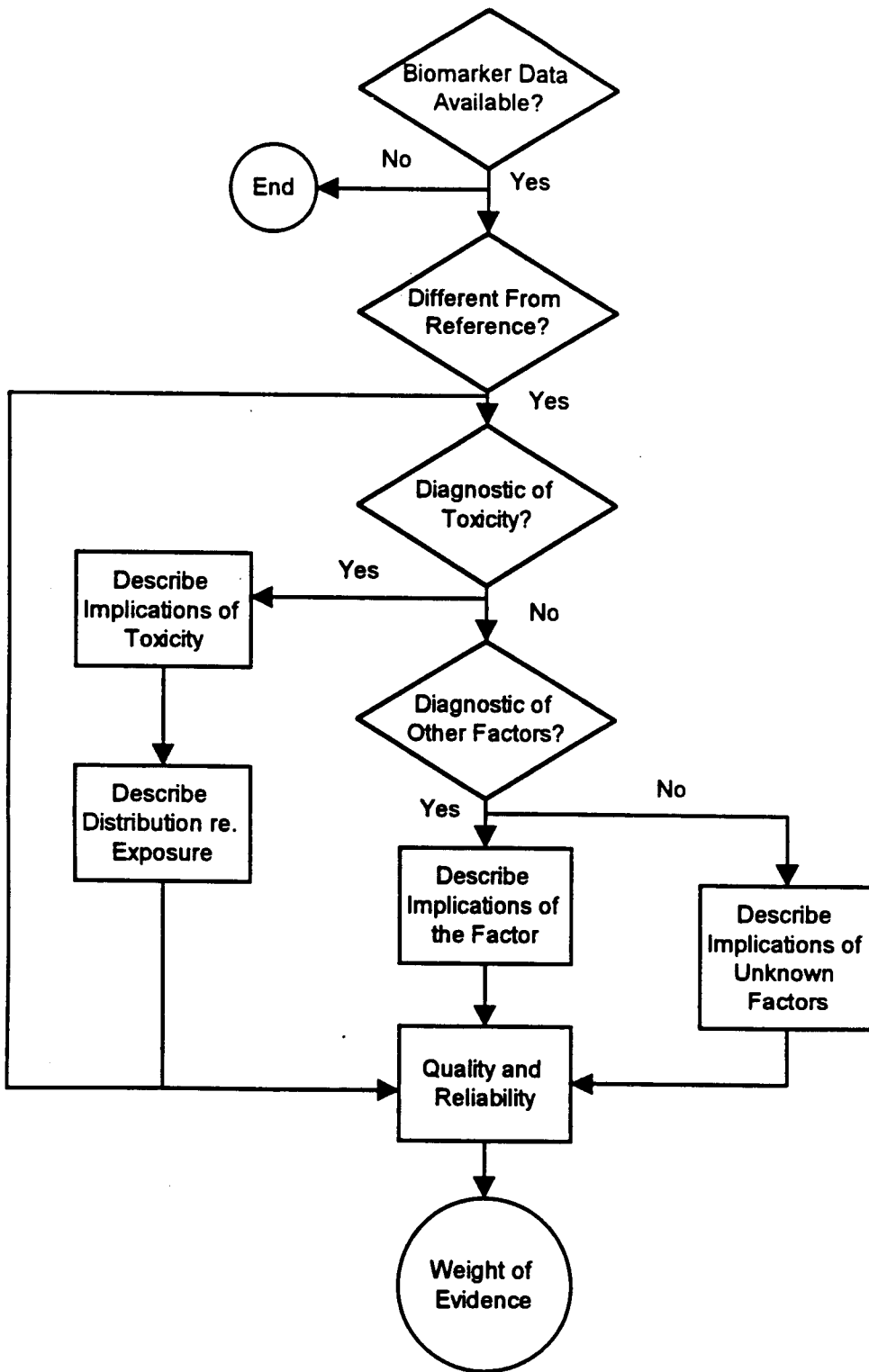


Fig. 11. Risk characterization based on biomarker data.



To facilitate the weight-of-evidence analysis and to make the bases clear to the reader, it may be useful to summarize the results of this integration for each reach or area using the following table.

**Table 5. Summary table for integration of biomarker results**

<b>Issue</b>	<b>Result</b>
Taxa and biomarkers significantly responding	<i>List the species and specific responses</i>
Implications of biomarker responses for organisms and populations	<i>Describe, as far as possible, the relationship between the biomarkers and population or community endpoints</i>
Chemicals that induce the observed response	<i>List chemicals or chemical classes that are known to induce the biomarker response</i>
Number and nature of reference sites	<i>List and describe reference sites including habitat differences</i>
Association with habitat or seasonal variables	<i>List habitat or life cycle variables that may affect the level of the biological response at the site</i>
Association with source or chemical	<i>Describe any correlations or qualitative associations of apparent effects with sources or chemical concentrations</i>
Most likely cause of response	<i>Based on the associations described in previous items, present the most likely cause of the apparent effects</i>
Estimated effects	<i>Summarize the estimated nature and extent of effects associated with the biomarker and credible upper bounds if they can be identified</i>
Confidence in results	<i>Rating and associated comments</i>

## 6. WEIGHT OF EVIDENCE

The weighing of evidence begins by summarizing the available lines of evidence for each endpoint (Fig. 12). Given that one has estimated risks based on each line of evidence, the process of weighing the evidence amounts to determining what estimate of risks is most likely given those results. If the assessment endpoint is defined in terms of some threshold for significance, then the process can be conducted in two steps. First, for each line of evidence, determine whether it is consistent with exceedence of the threshold, inconsistent with exceedence, or ambiguous. Second, determine whether the results as a whole indicate that it is likely or unlikely that the threshold is exceeded. If the results for all lines of evidence are consistent or inconsistent, the result of the weighing of evidence is clear. Assuming that there is no bias in the assessment that affects all lines of evidence, agreement among multiple lines of evidence is strong support for a conclusion. However, if there are inconsistencies, the true weighing of evidence must occur. The weights are determined based on the following considerations (Menzie et al. 1996; Suter 1993):

**Relevance**—Evidence is given more weight if the measurement endpoint is more directly related to (i.e., relevant to) the assessment endpoint.

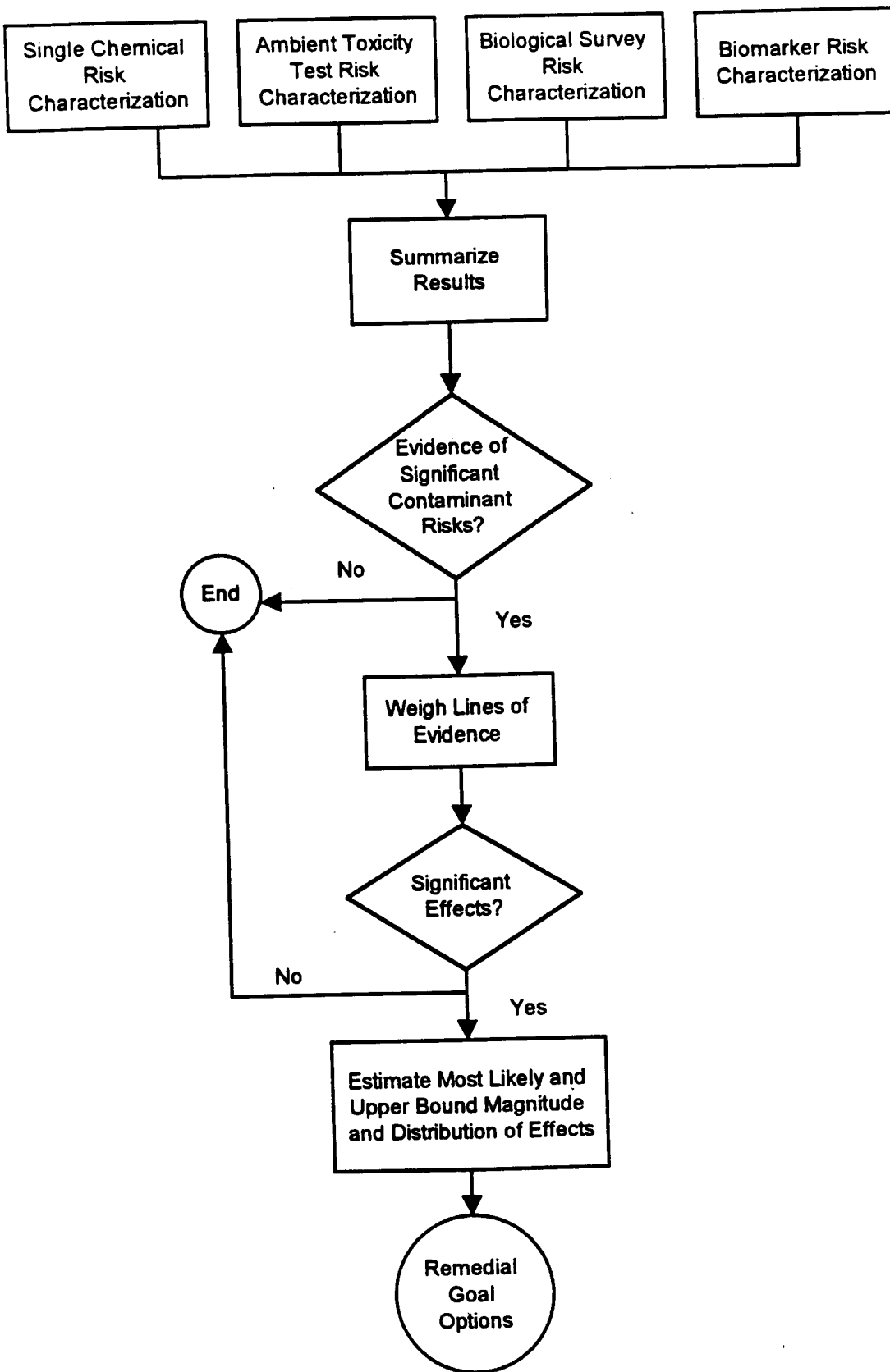


Fig. 12. Risk characterization based on weighing of multiple lines of evidence.

- Effects are relevant if the measurement endpoint is a direct estimate of the assessment endpoint or if validation studies have demonstrated that the measurement endpoint is predictive of the assessment endpoint. Note that a measurement endpoint based on statistical significance (e.g., a NOEC) is less likely to bear a consistent relationship to an assessment endpoint than one that is based on biological significance (e.g., and  $EC_x$ ).
- The mode of exposure may not be relevant if the media used in a test are not similar to the site media. Normalization of media concentrations may increase the relevance of a test if the normalization method has been validated. Similarly, the relevance of tests in solution to sediment or soil exposures is low unless the models or extraction techniques used to estimate aqueous phase exposures have been validated.
- Measurement endpoints derived from the literature rather than site-specific studies may have used a form of the chemicals that is not relevant to the chemical detected in the field. For example, is it the same ionization state and has the weathering or sequestration of the field contaminant changed its composition or form in ways that are not reflected in the test?

When the relationship is unclear, relevance may be evaluated by listing the ways in which the results could be wrong because they are fundamentally inappropriate or so inaccurate as to nullify the results and evaluate the likelihood that they are occurring in this case. For example, single chemical toxicity tests could be performed with the wrong form of the chemical; in media differing from the site media in ways that significantly affect toxicity; or the tests may be insensitive due to short duration, a resistant species, or the lack of measures of sublethal effects.

**Exposure/Response**—As in all toxicological studies, a line of evidence that demonstrates a relationship between the magnitude of exposure and the effects is more convincing than one that does not. For example, apparent effects in media toxicity tests may be attributed to the chemical with measured concentrations that exceed benchmarks by the greatest margin, but unless the tested medium is analyzed and an exposure/response relationship demonstrated, it may be suspected that effects are a result of other contaminants, nutrient levels, texture, or other properties. If an exposure-response relationship has not been demonstrated, then consideration should be given to the magnitude of the observed differences. For example, if medium test data include only comparisons of contaminated and uncontaminated soils, the observed differences are less likely to be due to extraneous factors if they are large (e.g., 100% mortality rather than 25% less growth).

**Temporal Scope**—Determine whether the data encompass the relevant range of conditions. For example, if contaminated and reference soils are surveyed during a period of drought, few earthworms will be found at any site so toxic effects will not be apparent. Temporal scope may also be inadequate if aqueous toxic effects occur when storm events flush contaminants into streams but water for toxicity testing is not collected during such events.

**Spatial Scope**—Determine whether the data adequately represent the area to be assessed including not only the directly contaminated area but also indirectly contaminated areas and indirectly affected areas. In some cases, the most contaminated or most susceptible areas were not sampled because of access problems or because of the sampling design (e.g., random sampling with few samples).

**Quality**—Evaluate the quality of the data in terms of the protocols for sampling, analysis, and testing; the expertise of the individuals involved in the data collection; the adequacy of the quality control during sampling, sample processing, analysis, and recording of results; and any other issues that are known to affect the quality of the data for purposes of risk assessment. Although use of standard methods tends to increase the likelihood of high quality results, they are no guarantee. Standard

methods may be poorly implemented or may be inappropriate to a site. In contrast a well-designed and -performed site-specific measurement or testing protocol can give very high quality results.

**Quantity**—Evaluate the adequacy of the data in terms of the number of observations taken. Results based on small sample sizes are given less weight. The adequacy of the number of observations must be evaluated relative to the variance as in any analysis of a sampling design, but it is also important in studies of this type to consider their adequacy relative to potential biases in the sampling (see previous discussion on spatial and temporal scope).

These and other considerations can be used as points to consider in forming an expert judgment or consensus about which way the weight of evidence tips the balance. Table 6 presents an example of a simple summary of the results of weighing evidence based on this process. The lines of evidence are listed, and a symbol is assigned for each: + if the evidence is consistent with significant effects on the endpoint, - if it is inconsistent with significant effects, and ± if it is too ambiguous to assign to either category. The last column presents a short summary of the results of the risk characterization for that line of evidence.

**Table 6. A hypothetical summary of a risk characterization by weight of evidence for a soil invertebrate community in contaminated soil**

<b>Evidence</b>	<b>Result<sup>a</sup></b>	<b>Explanation</b>
Biological Surveys	-	Soil microarthropod taxonomic richness is within the range of reference soils of the same type, and is not correlated with concentrations of petroleum components..
Ambient Toxicity Tests	-	Soil did not reduce survivorship of the earthworm <i>Eisenia foetida</i> . Sublethal effects were not determined.
Organism Analyses	±	Concentrations of PAHs in depurated earthworms was elevated relative to worms from reference sites but toxic body burdens are unknown.
Soil Analyses/Single Chemical Tests	+	If the total hydrocarbon content of the soil is assumed to be composed of benzene, then deaths of earthworms would be expected. Toxicity data for other detected contaminants are unavailable.
Weight-of-Evidence	-	Although earthworm tests may not be sensitive, they and the biological surveys are both negative and are both more reliable than the single chemical toxicity data used with the analytical results for soil.

Notes:

+ indicates that the evidence is consistent with the occurrence of a 20% reduction in species richness or abundance of the invertebrate community.

– indicates that the evidence is inconsistent with the occurrence of a 20% reduction in species richness of abundance of the invertebrate community.

± indicates that the evidence is too ambiguous to interpret.

If indirect effects are part of the conceptual model, they should be summarized in their own line of the table. For example, effects on piscivorous wildlife could be due entirely or in part to toxicity to fish. The last line of the table presents the weight of evidence–based conclusion concerning whether significant effects are occurring and a brief statement concerning the basis for the conclusion. This

conclusion is not based simply on the relative number of + or - signs. The “weight” component of weight of evidence is the relative credibility and reliability of the conclusions of the various lines of evidence as discussed previously. Additionally, those considerations can be used to grade the weight to be assigned to each line of evidence (e.g., high, moderate, or low weight) (Table 7). This still leaves the inference to a process of expert judgement or consensus but makes the bases clearer to readers and reviewers.

**Table 7. Example of a table summarizing the risk characterization for the fish community in a stream at a waste site**

Evidence	Result <sup>a</sup>	Weight	Explanation
Biological Surveys	-	H	Fish community productivity and species richness are both high in reach 2, relative to reference reaches. Data are abundant and of high quality
Ambient Toxicity Tests	±	M	High lethality to fathead minnow larvae in a single test at Site 3.3, but variability is too high for standard statistical significance. No other aqueous toxicity was observed.
Water Analyses/Single Chemical Tests	+	M	Only Zn is believed to be potentially toxic in water and only to highly sensitive species. Few water samples were analyzed.
Weight-of-Evidence	-		Reach 2 supports a clearly high quality fish community. Other evidence which suggests toxic risks is much weaker (single chemical toxicology) or inconsistent and weak (ambient toxicity tests).

Notes:

+ indicates that the evidence is consistent with the occurrence of the endpoint effect.

- indicates that the evidence is inconsistent with the occurrence of the endpoint effect.

± indicates that the evidence is too ambiguous to interpret.

Finally, a scoring system could be developed that would formalize the weighing of evidence. For example, a numerical weight could simply be assigned to each line of evidence based on quality, relevance, and other factors; a + or - assigned depending on whether the evidence is consistent or inconsistent with the hypothesized risk; and the weights summed across lines of evidence. A quantitative system has been developed by a group consisting of representatives of the state of Massachusetts, the private sector, and U.S. government agencies (Menzie et al. 1996). Such systems have the advantage of being open, consistent, and less subject to hidden biases, but they may not give as reasonable a result in every case as a careful *ad hoc* weighing of the evidence would. However the weighing of evidence is performed, it is incumbent on the assessment scientist to make the basis for the judgment as clear as possible to readers and reviewers. Where multiple subsites or reaches are assessed, it is helpful to provide a summary table for the weighing of evidence across the entire site as shown in Table 8 so the consistency of judgment can be reviewed.

**Table 8. Summary of weight-of-evidence analyses for reaches exposed to contaminants in the Clinch River/Poplar Creek Operable Unit**

Reach	Biological Surveys	Bioindicators	Ambient Toxicity Tests	Fish Analyses	Water Analyses/Single Chemical Toxicity	Weight of Evidence
2. Upper Clinch R. Arm	±	±		±	-	-
3. Poplar Creek Embayment	+	±	+	±	+	+
4. Lower Clinch R. Arm	-	±	-	±	+	-
7. McCoy Branch Embayment			±		-	-

**Notes:**

+ indicates that the evidence is consistent with the occurrence of the endpoint effect.

- indicates that the evidence is inconsistent with the occurrence of the endpoint effect.

± indicates that the evidence is too ambiguous to interpret.

Blank cells indicate that data were not available for that line of evidence.

The use of quantitative weighing of evidence or of an equivalent expert judgment about which lines of evidence are most reliable is based on an implicit assumption that the lines of evidence are logically independent. Another approach to weighing multiple lines of evidence is to determine whether there are logical relationships among the lines of evidence. Based on knowledge of site conditions and of environmental chemistry and toxicology, one may be able to explain why inconsistencies occur among the lines of evidence. For example, one may know that spiked soil tests tend to overestimate the availability and hence the toxicity of contaminants and may even be able to say whether the bias associated with this factor is sufficient to account for discrepancies with tests of site soils. This process of developing a logical explanation for differences among lines of evidence is potentially more convincing than simple weighing of the evidence because it is mechanistic. However, it is important to remember that such explanations can degenerate into just-so stories if the relevance of the proposed mechanisms is not well supported.

In general, a logical analysis of the data should proceed from most realistic (i.e., site-specific) to most precise and controlled (e.g., single chemical and species toxicity tests). Field surveys indicate the actual state of the receiving environment, so other lines of evidence that contradict the field surveys, after allowing for limitations of the field data, are clearly incorrect. For example, the presence of plants that are growing and not visibly injured indicates that lethal and gross pathological effects are not occurring but does not preclude reductions in reproduction or growth rates; these effects could be addressed by more detailed field studies of growth rates and seed production and viability.

The presence of individuals of highly mobile species such as birds indicates almost nothing about risks because dispersal replaces losses of individuals or reduced reproduction. Ambient media toxicity tests indicate whether toxicity could be responsible for differences in the state of the receiving environment, including differences that may not be detectable in the field. However, field effects are usually more credible than negative test results because field exposures are longer and otherwise more realistic and site species and life stages may be more sensitive than test species and life stages.

Single chemical toxicity tests indicate which components of the contaminated ambient media could be responsible. Because they are less realistic than other lines of evidence, single chemical toxicity tests are usually less credible than the other lines of evidence. They do not include combined toxic effects, the test medium may not represent the site media, the exposure may be unrealistic, and the chemicals may be in a different form than at the site. However, because these studies are more controlled than those from other lines of evidence, they are more likely to detect sublethal effects. In addition, single chemical toxicity tests may include longer exposures, more sensitive responses, and more sensitive species than tests of contaminated ambient media. These sorts of logical arguments must be generated *ad hoc* because they depend on the characteristics of the data and the site.

After the lines of evidence have been weighed to reach a conclusion about the significance of risks to an assessment endpoint, it is usually appropriate to proceed to estimate the nature, magnitude, and distribution of any effects that were judged to be significant. A significant risk is sufficient to prompt consideration of remedial actions, but the nature, magnitude, and distribution of effects determine whether remediation is justified given remedial costs and countervailing risks. In general, it will be clear that one line of evidence provides the best estimate of effects and that is likely to be the most site-specific line of evidence. However, other lines of evidence may contribute by setting bounds on the estimate.

## 7. FUTURE RISKS

Baseline ERAs for the Oak Ridge Reservation focus primarily on current risks. However, future baseline risks should be characterized when:

- contaminant exposures are expected to increase in the future (e.g., a contaminated groundwater plume will intersect a stream),
- biological succession is expected to increase risks (e.g., a forest will replace a lawn), or
- significant recovery is expected to occur in the near term without remedial actions (i.e., the expense and ecological damage associated with remedial actions may not be justified).

Although these future baseline risks cannot be characterized by measuring effects or by testing future media, all lines of evidence that are useful for estimating current risks may be extended to them. As in human health risk assessments, risk models derived by epidemiological methods can be applied to future conditions and even applied to different sites. For example, if concentrations are expected to change in the future, the exposure-response relationship derived from biosurvey data (e.g., a relationship between contaminant concentration and fish abundance) may supply a better estimate of future effects than a concentration-response relationship derived from laboratory test data. Results of toxicity tests of currently contaminated media may also be used to estimate future effects. The utility of the various risk models depends on their reliability (as suggested by the weight-of-evidence analysis) and their relevance to the future conditions.

## 8. REFERENCES

- Alabaster, J. S., and R. Lloyd. 1982. *Water Quality Criteria for Freshwater Fish*. Butterworth Scientific, London.
- DiToro, D. M., J. A. Hallden, and J. L. Plafkin. 1991. Modeling *Ceriodaphnia* toxicity in the Naugatuck River II. copper, hardness and effluent interactions. *Environ. Toxicol. Chem.* 10:261-274.
- Finney, D. J. 1971. *Probit Analysis*, 3rd Edition. Cambridge University Press, Cambridge.
- HECD (Health and Criteria Division). 1992. *Interim guidance on interpretation and implementation of aquatic life criteria for metals*. U.S. Environmental Protection Agency, Washington, D.C.
- 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, TN.
- MacDonald, D. D. 1994. *An approach to the assessment of sediment quality in Florida coastal waters* (Four Volumes). Florida Department of Environmental Protection, Tallahassee.
- McCarty, L. S., and D. Mackay. 1993. Enhancing ecotoxicological modeling and assessment: body residues and modes of toxic action. *ES&T* 27:1719-1728.
- Menzie, C, M. H. Henning, J. Cura, K. Finkelstein, J. Gentile, J. Maughan, D. Mitchell, S. Petron, B. Potocki, S. Svirsky, et al. 1996. A weight-of-evidence approach for evaluating ecological risks: report of the Massachusetts Weight-of-Evidence Work Group. *Human Ecol. Risk Assess.* 2(2):277-304.
- Prothro, M. G. 1993. Office of Water Policy and Technical Guidance on Interpretation and Implementation of Aquatic Life Metals Criteria. Memorandum to Water Management Division Directors and Environmental Services Division Directors, Regions I-X, October 1, 1993.
- Risk Assessment Forum. 1992. *Framework for Ecological Risk Assessment*. EPA/630/R-92/001. United States Environmental Protection Agency, Washington, D.C.
- Sample, B. E., D. M. Opresko, and G. W. Suter II. 1996. *Toxicological benchmarks for wildlife*. ES/ER/TM-86/R3. Oak Ridge National Laboratory, Oak Ridge, Tennessee.
- Suter, G. W., II. 1993. *Ecological Risk Assessment*. Lewis Pub., Boca Raton, Florida.
- Suter, G. W., II., B. E. Sample, D. S. Jones, T. L. Ashwood, and J. M. Loar. 1995. *Approach and strategy for performing ecological risk assessments for the Department of Energy's Oak Ridge Reservation: 1995 Revision*. ES/ER/TM-33/R2. Environmental Restoration Division, Oak Ridge, Tennessee.
- Suter, G.W., II. 1995. *Guide for performing screening ecological risk assessments*. ES/ER/TM-153. Environmental Restoration Division, Oak Ridge, Tennessee.
- 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, Tennessee.



Will, M. E., and G. W. Suter II. 1995a. *Toxicological benchmarks for screening potential contaminants of concern for effects on soil and litter invertebrates and heterotrophic processes*. ES/ER/TM-126. Oak Ridge National Laboratory, Oak Ridge, Tennessee.

Will, M. E., and G. W. Suter II. 1995b. *Toxicological benchmarks for screening potential contaminants of concern for effects on terrestrial plants*. ES/ER/TM-85/R1. Oak Ridge National Laboratory, Oak Ridge, Tennessee.