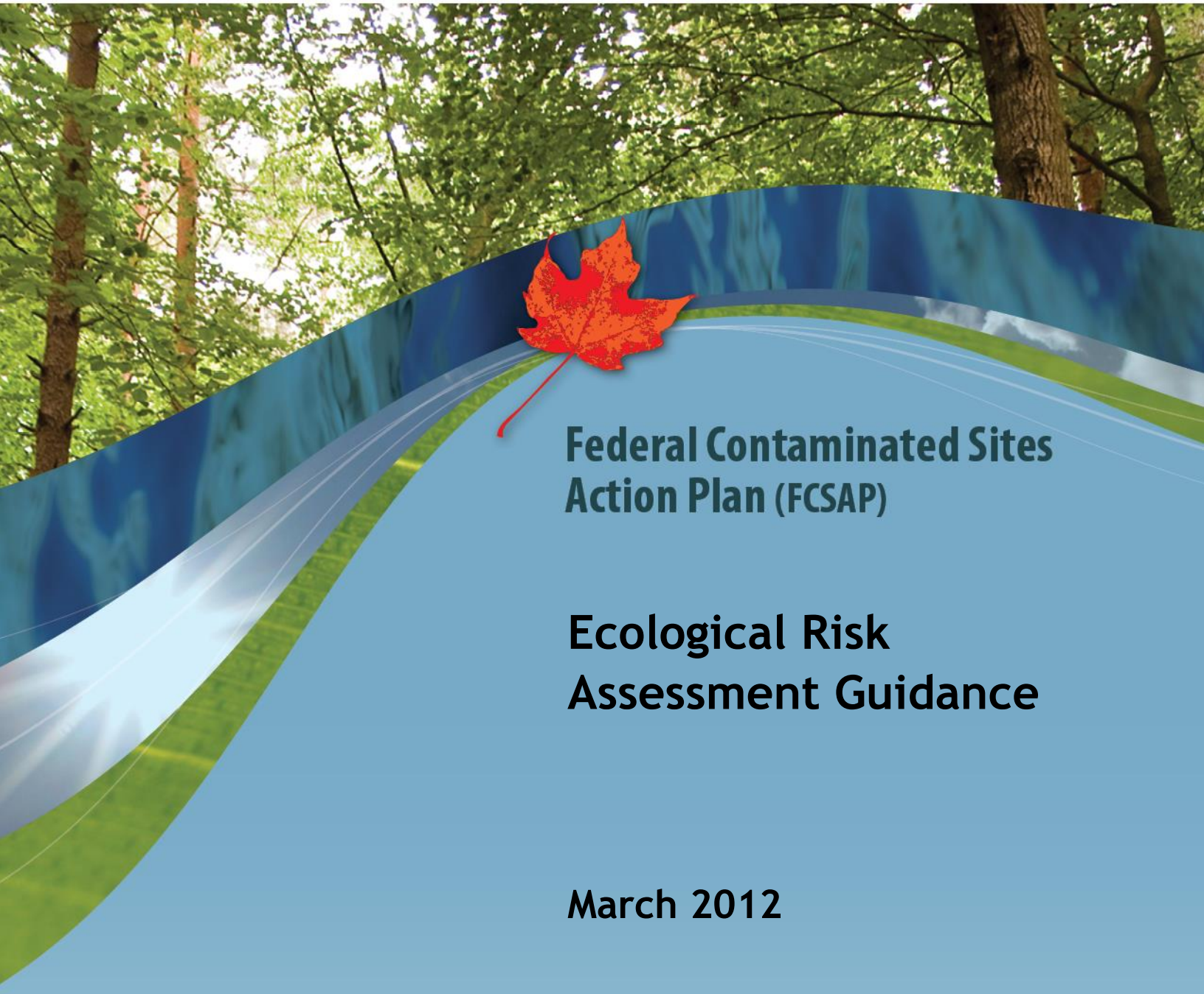
Government of Canada | Gouvernement du Canada

Canada



Federal Contaminated Sites Action Plan (FCSAP)

Ecological Risk Assessment Guidance

March 2012

LIBRARY AND ARCHIVES CANADA CATALOGUING IN PUBLICATION

Federal Contaminated Sites Action Plan (FCSAP) Ecological Risk Assessment Guidance

Issued also in French under title:
Plan d'action pour les sites contaminés fédéraux (PASCF) : Document d'orientation sur l'évaluation du risque écotoxicologique

ISBN no. 978-1-100-22282-0

Cat. no. En14-19/1-2013E-PDF

DISCLAIMER

Her Majesty is not responsible for the accuracy or completeness of the information contained in the reproduced material. Her Majesty shall at all times be indemnified and held harmless against any and all claims whatsoever arising out of negligence or other fault in the use of the information contained in this publication or product.

The information in this document does not constitute legal advice; following this guidance will not necessarily ensure compliance with federal, provincial, or any other regulatory requirements. In case of discrepancy between this information and any Acts of Parliament, most notably the *Canadian Environmental Protection Act, 1999* or the *Fisheries Act* or regulations made under these Acts, the Acts of Parliament and associated regulations take precedence. Notwithstanding any other regulatory or permitting requirements, any deposits, discharges and releases from your operations or activities must comply with all applicable federal Acts and regulations.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF TECHNICAL MODULES

Module 1: Toxicity Test Selection and Interpretation

Module 2: Selection or Development of Site-Specific Toxicity Reference Values

Module 3: Standardization of Wildlife Receptor Characteristics

Module 4: Causality Assessment[1]

[1] Remained under development in January 2012. Additional modules will be developed in the future. Contact FCSAP.PASCF@ec.gc.ca for a complete list of technical ERA modules.

ACKNOWLEDGEMENTS

Environment Canada would like to thank the many reviewers in government and private sector in Canada and the United States who provided valuable comments as part of the public peer review process.

ACRONYMS

ADME	Absorption, distribution, metabolism, and excretion
AEC	Area of environmental concern
AEL	Acceptable effect level
ANOVA	Analysis of variance
APEC	Area of potential environmental concern
ASTM	American Society for Testing and Materials
AVS	Acid volatile sulfide
BAF	Bioaccumulation factor
BCF	Bioconcentration factor
BPJ	Best professional judgement
BLM	Biotic ligand model
BSAF	Biota-sediment accumulation factor
CABIN	Canadian Aquatic Biomonitoring Network
CBR	Critical body residue
CCA	Canonical correspondence analysis
CCME	Canadian Council of Ministers of the Environment
CEC	Cation exchange capacity
CEQG	Canadian environmental quality guideline
CERCLA	Comprehensive Environmental Response, Compensation, and Liability Act
COC	Contaminant of concern
CSM	Conceptual site model
DAF	Dose adjustment factor
DDT	Dichlorodiphenyltrichloroethane
DFO	Department of Fisheries and Oceans (Canada)
DQO	Data quality objective
EC	Environment Canada
EC_X	Effect concentration, with percent *X* of organisms affected
ED	Effect dose
Eh	Redox potential
EPT	Ephemeroptera, Plecotera, Trichoptera, known collectively as the EPT taxa (mayflies, caddisflies, and stoneflies)

ERA	Ecological risk assessment
ERED	Environmental Residue-Effects Database
FCSAP	Federal Contaminated Sites Action Plan
FRF	Foraging range factor
GIS	Geographic Information Systems
HC	Health Canada
HC_p	Hazardous concentration (affecting percentage *p* of population)
HI	Hazard index
HQ	Hazard quotient
IC	Inhibitory concentration
ID	Inhibitory dose
K_{OW}	Octanol-water partition coefficient
LC	Lethal concentration
LD	Lethal dose
LOE	Line of evidence
LOAEC / LOAEL	Lowest observed adverse effect concentration / level
NOAEC / NOAEL	No observed adverse effect concentration / level
NRC	National Research Council
OMOE	Ontario Ministry of Environment
PAH	Polycyclic aromatic hydrocarbon
PBET	Physiologically based extraction test
PCA	Principal components analysis
PCB	Polychlorinated biphenyl
PF/SAP	Problem formulation & sampling and analysis plan
pH	Power of hydrogen; negative logarithm of the hydrogen ion concentration
PQRA	Preliminary quantitative risk assessment
PRP	Potentially responsible party
PSAMP	Puget Sound Assessment and Monitoring Program
PWGSC	Public Works and Government Services Canada
QA	Quality assurance
QC	Quality control

RCA	Reference condition approach
ROC	Receptor of concern
SAB-CS	Science Advisory Board (British Columbia Contaminated Sites)
SAP	Sampling and analysis plan
SARA	Species at Risk Act
SEM	Simultaneously extracted metal
SQG	Sediment quality guideline
SSD	Species sensitivity distribution
SSTL	Site-specific target level
SWAC	Spatially weighted average concentration
SWAMP	Surface Water Ambient Monitoring Program (California)
TCCR	Transparency, clarity, consistency, reasonableness
TEQ	Toxic equivalents
TIE	Toxicity identification evaluation
TRA	Tissue residue approach
TRG	Tissue residue guideline
TRV	Toxicity reference value
UCLM	Upper confidence limit of the mean
UF	Uptake factor
USEPA	United States Environmental Protection Agency
VEC	Valued ecosystem component
WOE	Weight of evidence

GLOSSARY

Abiotic medium – Any environmental medium not associated with biological tissue (e.g., soil, sediment, water, air).

Acute – Relating to a small increment of time required to elicit an adverse environmental response. With respect to toxicity testing, the term describes tests applied over a short duration, typically less than ten percent of an organism's lifespan. Note, however, that some short-term tests may be defined as chronic rather than acute if they are conducted using a sensitive life stage; definitions of acute versus chronic vary widely by jurisdiction.

A priori – Refers to prior knowledge about a condition, rather than that estimated by recent observation. In ERA, the term *a priori* is used to describe knowledge or models of biological systems considered prior to the conduct of the analysis phases of the risk assessment.

Acceptable effect level (AEL) – The magnitude (or rate) of effects that would be acceptable for a specific measurement endpoint or assessment endpoint. The AEL operationalizes a protection goal.

Analysis of variance (ANOVA) – A statistical method used for a single dependent variable that performs comparisons and tracks the effects of one or more discrete factors (independent variables), each of which may have a number of levels and may interact to affect the dependent variable.

Application factor – see *Safety factor*.

Area of potential environmental concern (APEC) – A portion of a site where contamination is suspected or confirmed.

Assessment endpoint – An assessment endpoint is an explicit expression of the environmental value to be protected. An assessment endpoint must include an entity (typically a receptor or receptor group – i.e., a 'thing' to be protected) and a specific property of that receptor (an attribute). For example, if the entity is a fish community, attributes could include the number of species, the trophic structure, etc. An assessment endpoint may also have an explicit spatial or temporal component.

Assessment factor – see *Safety factor*.

Aryl hydrocarbon receptor (*Ah* receptor) – A member of the family of basic-helix-loop-helix transcription factors. The *Ah* receptor binds to certain chemicals such as dioxins and PCB congeners, causing the receptor to translocate into the nucleus of organism cells, eventually leading to genetic damage. The mechanism of toxicity via the *Ah* receptor underpins the use of the toxic equivalents system for evaluating responses of chlorinated organic substances to vertebrates.

Attribute – A quality of an endpoint that reflects one aspect of its value for informing the risk assessment.

Best professional judgement (BPJ) – The thorough application of critical judgement in professional practice, in which an experiential, reflective, self-corrective, and purposeful thinking process is applied to consider knowledge, context, evidence, methods, conceptualizations, and criteria. BPJ is a means by which a practitioner can incorporate a diverse range of information without articulating a mechanical process for processing the information.

Bias – A systematic tendency that distorts the interpretation of results. In ERA, a bias occurs in two main forms. In the study design or interpretation, bias is a perjorative term that reflects partiality of a practitioner that prevents objective consideration of an issue or situation. In statistical measurement, bias reflects a systematic under- or over-prediction of a true parameter value. Both forms of bias introduce systematic error into risk estimates.

Bioaccumulation – The process by which substances accumulate in the tissues of living organisms. Bioaccumulation occurs when the concentration of a COC in an organism is higher than the concentration in the surrounding environment. Most substances bioaccumulate to some extent, whereas few biomagnify.

Bioaccumulation factor (BAF) – The quotient obtained by dividing the concentration of a substance in an organism (or specified tissue) by its concentration in a specified exposure medium, for example, air, food, sediment, soil, water (definition from ASTM 2011).

Bioconcentration factor (BCF*)* – Equivalent to an uptake factor, for the case where water (only) is the abiotic exposure medium.

Biomagnification – Refers to the process by which chemical concentrations in plants or animals increase relative to food from transfer through the food web (e.g., predators have greater concentrations of a particular chemical than their prey).

Biota-sediment accumulation factor (BSAF) – Equivalent to an uptake factor, where the abiotic medium is sediment, and where both the tissue and sediment concentrations are normalized to carbon pools (lipid and total organic carbon, respectively).

Biotic medium – Any biological medium (i.e., tissues) where COCs may be found.

Category of evidence – A type of related lines of evidence within a *weight-of-evidence* framework.

Causation – The act or fact of causing; the production of an effect by a cause. Causation differs from association (correlation) in that the latter does not imply a mechanistic linkage between observations. An assessment of causation in an ecological risk assessment attempts to distinguish between associations that are coincidental or caused

by external factors and associations that are driven by underlying predictable mechanisms.

Chronic – Relating to an extended time duration. In the context of toxicity testing, the term is used to describe tests that expose organisms over a substantial portion of their life cycle, for example more than 10% of the life cycle or throughout a sensitive life stage. Definitions of chronic vary widely.

Cluster analysis – A class of statistical techniques that can be applied to data that exhibit "natural" groupings based on an assessment of interdependence. Cluster analysis sorts through the raw data and groups them into clusters of relatively homogeneous cases or observations. Whereas factor analysis reduces the number of variables by grouping them into a smaller set of factors, cluster analysis reduces the number of observations or cases by grouping them into a smaller set of clusters.

Coherence – A concept that relates to the way in which multiple lines of evidence are congruent; approaches to the assessment of coherence include evaluations of causation, ecological relevance, logical interpretations, and best professional judgement. In a WOE approach, coherence analysis is applied following the "face-value" interpretation of results to determine whether the lines of evidence are consistent and/or provide a unified interpretation of findings.

Concentration-response – The relationship between an effects measure and exposure (measured as concentration) across a range of exposure concentrations.

Conceptual site model (CSM) – A narrative and graphical representation of the relationships between contaminant sources, fate, exposure pathways, and receptors.

Conservative – Adjective expressing the tendency to deliberately overstate the potential for environmental harm. The overestimate is intended to provide a margin of error to buffer against uncertainty in the analysis, and to provide increased confidence that estimates or predictions of risk are not understated. In ERA practice, it is common to apply conservatism in parameter estimation. However, when conservatism is too great, either through unrealistic assumptions or through compounding of multiple conservative assumptions, an analysis is deemed to be ultra-conservative, and therefore suspect.

Contaminants of Concern (COCs) – Contaminants that have been selected for evaluation in the ERA, usually based on a completed problem formulation.

Control – As a noun, an aspect of a controlled scientific experiment conducted for the purpose of determining the effect of a single variable of interest on a particular system, used to minimize the unintended influence of other variables on the same system. Negative controls confirm that the procedure is not causing an unrelated effect, and are intended to reduce incidence of *false positives*. The term control (as a verb) can also be used in experimental design to refer to manipulation of treatments intended to mitigate the confounding effect of external variables.

Correspondence analysis – A *multivariate* statistical technique that is conceptually similar to principal components analysis, in which data are scaled such that rows and columns are treated equivalently.

Critical body residue (CBR) – An internal body or tissue concentration that is associated with a toxicological response in a receptor.

Deterministic methods – Methods in which all biological, chemical, physical, and environmental parameters are assumed to be constant and can be accurately specified. Deterministic methods commonly apply to either a "most likely" value for a parameter, or a conservative value intended to guard against uncertainty.

Dichotomous – Adjective characterizing a parameter with only two possible states.

Dilution series – An experimental design and technique in which an abiotic medium is divided into multiple exposure magnitudes by diluting the full strength medium using clean material. A series of concentrations is specified using graded dilutions, with responses characterized for each treatment on a volume/volume, mass/volume, or mass/mass basis.

Diversity – An attribute referring to variation within an ecological community. In general, high diversity is associated with high richness (number of taxa) and evenness of abundance among taxonomic groups. Diversity is often used as a measure of ecosystem health. A number of numerical diversity indices have been developed, each of which has different theoretical underpinnings.

Dose-response – The relationship between an effects measure and exposure (measured as dose) across a range of dose values.

Ecological relevance – The degree to which a type of information used in an ERA (i.e., a measurement endpoint or line of evidence) can be meaningfully extrapolated to the biological scale of interest (i.e., the assessment endpoint).

Ecological Risk Assessment (ERA) – The process of evaluating the potential adverse effects on non-human organisms, populations or communities in response to human-induced stressors. ERA entails the application of a formal framework, analytical process, or model to estimate the effects of human actions on natural organisms, populations or communities and interprets the significance of those effects in light of the uncertainties identified in each study component.

Effect size – The absolute or relative magnitude of response to a stressor for a *measurement endpoint*.

Effects assessment – For any line of evidence, the component of a risk assessment that characterizes the nature of effects elicited by each contaminant under an exposure condition that is relevant to each receptor of concern.

Exposure assessment – For any line of evidence, the component of a risk assessment that quantifies the degree to which an organism encounters a stressor.

Exposure pathways – The routes through which a receptor of concern encounters COCs in environmental media (e.g., soil, water, air, sediment). Examples of exposure pathways include ingestion and inhalation.

Exposure point concentration – The value that represents a conservative estimate of the chemical concentration or dose available to an organism from a route of exposure.

Extrapolation – Inference or estimation by extending or projecting known information to a domain (spatial, temporal, biological, or chemical) that has not yet been studied. In statistics, extrapolation entails estimation (of a value of a variable outside a known range) from values within a known range, and requires an assumption that the estimated value follows logically from the known values.

Extrapolation factor – see *Safety factor*.

False negative – The error (often called a Type II error) in which a response occurs but is not detected.

False positive – The error (often called a Type I error) in which a response is deemed to occur when in fact there was no response. The term is often used to describe a situation in which an inappropriate conclusion was rendered based on available information.

Feeding guild – A group of organisms that use the same ecological resource in a similar way for feeding (e.g., insectivores, granivores, detritivores, carnivores); or, a group of species that overlap significantly in their niche requirements.

Gradient – A concept of experimental design in which treatments are planned to include a range of exposures from low to high, or a spatial range (e.g., near to far).

Guideline – A regulatory value that is recommended for the screening of environmental data, such as tissue residues or concentrations in abiotic media. A guideline usually differs from a standard in that a guideline does not convey a legal requirement or formal responsibility. Canadian Environmental Quality Guidelines are intended as nationally endorsed science based goals for environmental quality. The term is also used to describe a technical practice that is recommended to facilitate consistency among practitioners, but that is not strictly required.

Hazard quotient (HQ) – A numerical ratio that divides an estimated environmental concentration or other exposure measure by a response benchmark. Typically the response benchmark is a value assumed to be protective of the receptor of concern. HQ values below one (1.0) indicate negligible potential for harm, whereas HQ values above one indicate that an adverse response is possible and that more precise or accurate evaluation of risks may be warranted to address uncertainty.

Hazardous concentration (HC_p) – a threshold concentration from a species sensitivity distribution. The concentration is derived considering a proportion of the species affected (p) and an effect size of interest (i.e., acceptable level of response).

Hazard index (HI) – The arithmetic sum of individual *hazard quotients*, used to aggregate the individual responses of multiple stressors, that implicitly assumes linear additivity of response. A hazard index is applied where the mode of toxic action is considered to be similar among COCs.

Home range – The geographic area to which an organism normally confines its activity; for *exposure assessment* the activity of interest is usually the foraging area over a defined period of time, such as feeding range during the reproductive period.

Hypothesis – A proposed explanation for an observable phenomenon; in experimental design, a hypothesis is set forth and subsequently tested (either singly or along with multiple alternate hypotheses) to determine if the new data support or contradict the hypothesis.

Interpolation – To estimate a value of (a function or series) between two known values. The term can also be applied more generically to the assignment of qualities to members of a group based on observations of other members of the same group. Interpolation requires the underlying assumption that members of a group are similarly influenced by the processes under investigation.

Likelihood – In common usage, synonymous with the probability or frequency of an event. In statistical usage, likelihood is distinguished from probability, and refers to the estimation of unknown parameters based on known outcomes.

Line of evidence (LOE) – Any pairing of exposure and effects measures that provides evidence for the evaluation of a specific assessment endpoint. Typically a line of evidence requires use of one or more measurement endpoints. If the focus of the LOE is an effects measure (e.g., a toxicity test), the paired exposure measure may be quantitative (e.g., contaminant concentrations) or categorical (e.g., on-site versus a reference condition).

Linear model – A category of statistical methods that underlies many of the statistical analyses that are used in applied sciences. It is the foundation for the Student t-test, analysis of variance (ANOVA), regression analysis, and many multivariate methods. Linear models assume that the relationship between a response variable and explanatory variables (or factors) is linear, or can be approximated as linear following appropriate data transformation.

LOE group – a cluster of closely related LOEs that have a particular measurement endpoint (or multiple endpoints) in common and therefore incorporate some redundancy in a weight-of-evidence evaluation. Individual LOEs in a group should individually contribute sufficient incremental information (i.e., informing the evaluation of the

assessment endpoint) to warrant inclusion as separate LOE. An LOE grouping provides organization of related LOEs and flags potential for redundancy.

Lowest-observed-adverse-effect level (LOAEL) – Lowest amount, dose, or concentration of an agent, found by experiment or observation, that causes an adverse alteration of morphology, functional capacity, growth, development or life span in an organism, system, or (sub)population. Methods vary for identifying a LOAEL, but often apply statistical significance as a criterion.

Measurement endpoint – A measurement endpoint is a parameter that measures or describes exposure of, or an effect on, a receptor of concern. Alternatively, the term describes a change in an attribute of an assessment endpoint (or its surrogate) in response to a stressor to which it is exposed.

Model – A simplified description of a system, theory, or phenomenon that accounts for its known or inferred properties and that may be used for further study of its characteristics. In all cases, a model is a simplification of a more complex system, and the details not represented by the model structure are considered to be errors/variations not central to the problem at hand. Models include statistical models (numerical processes used to simulate or approximate complex processes) and conceptual models (graphical or schematic representation of key processes and pathways).

Monte Carlo Analysis – A probabilistic analysis technique where parameter values are drawn at random from defined input probability distributions, combined according to a model equation, and the process repeated iteratively until a relatively smooth distribution of solutions results.

Multivariate – A form of statistics encompassing the simultaneous observation and analysis of more than one statistical variable. In ecological risk assessment, the most common multivariate methods are clustering, correspondence analysis, factor analysis, principal components analysis, and multidimensional scaling.

Narcosis – A condition of deep stupor or unconsciousness produced by a drug or other chemical substance.

No-observed-adverse effect level (NOAEL) - An exposure level at which there are no statistically or biologically significant increases in the frequency or severity of adverse effects between the exposed organisms or population and the appropriate control; some effects may be produced at this level, but they are not considered to be adverse. Methods for identifying a NOAEL vary, but often apply statistical significance as a criterion.

Ordination – In *multivariate analysis*, ordination is a method complementary to data *cluster analysis*, and orders objects on multiple variables such that similar objects are near each other and dissimilar objects are farther from each other. These relationships between the objects are plotted on multiple axes and can be characterized numerically and/or graphically.

Point estimate – A single numerical value used to represent the state of a random variable. A point estimate collapses (or ignores) all of the variability and incertitude regarding a parameter or variable.

Post hoc – An adjective referring to a retrospective assessment (i.e., after the fact). Post hoc analysis, in the context of design and analysis of experiments, refers to statistical examination of data after the experiment has concluded, and may include searching for patterns that were not known in advance (*a priori*).

Potentially responsible party (PRP) – Refers to all industries, site owners, point sources, and legally responsible entities associated with contamination at a site.. The term is commonly used as part of contaminated sites legislation in the United States (*Comprehensive Environmental Response, Compensation, and Liability Act* [CERCLA] and Superfund).

Practitioner – The primary investigator in an ecological risk assessment responsible for the design, implementation, and interpretation of results. The practitioner, who may be a consultant, interacts with the responsible party for the site (client), the regulators, and other interested parties.

Precision – The quality of being repeatable in degree or value; the ability of a measurement to be consistently reproduced. Note that precise results are not necessarily accurate, as a precise measurement can be consistently biased.

Prescriptive – Pertaining to giving directives or rules, without flexibility or subjective analysis. Prescriptive approaches have a high degree of repeatability and consistency among investigators, but low degree of adaptability to site-specific conditions.

Probabilistic – Adjective describing a procedure in which the state of a random variable is described not as a point estimate (fixed value), but rather as a distribution of possible values. Using probabilistic methods, important biological, chemical, physical, and environmental parameters are assumed to vary or are uncertain and therefore are specified using distributions.

Probability – A mathematical way of expressing knowledge or belief that an event or outcome will occur or has occurred. In statistical usage, probability is distinguished from likelihood, and refers to the prediction of unknown outcomes based on known parameters.

Problem formulation – The first step in ERA that clarifies the nature of issues associated with contamination at a site and how those issues will be addressed.

Protection goal – A narrative statement that defines the desirable level of protection for a receptor or receptor group (see also *acceptable effect level*).

Qualitative – Adjective describing an approach that is narrative, referring to the characteristics of something being described, rather than numerical measurement.

Quantitative – Adjective describing an approach that is numerical (applies mathematical scores, probabilities, or parameters) in the derivation or analysis of risk estimates.

Receptor of Concern (ROC) – In ERA, any non-human individual organism, species, population, community, habitat or ecosystem that is potentially exposed to contaminants of concern and that is considered in the ERA. Identification of an organism as an ROC does not mean that it is being harmed, only that a pathway exists such that there is potential for harm.

Reference (condition) – A location, group of locations, or experimental treatment designed to reflect the ambient physical and chemical conditions of a contaminated medium or location in the absence of the stressors of concern in the risk assessment. For example, in a study of soil contamination, the reference condition should reflect the climate, substrate, and habitat factors relevant to the site but with no incremental contamination relative to background conditions. In some cases, the term reference may is used in the context of an altered local background condition (i.e., where the local conditions surrounding a site are not pristine due to non-point sources of contaminants). In other cases, the term reference is used to refer to pristine conditions in the absence of both site-specific contamination and non-point sources of contaminants.

Regression – A form of statistical modeling that attempts to evaluate the numerical relationship between one variable (termed the dependent variable) and one or more other variables (termed the independent variables).

Response profile – The relationship between COC concentrations and ecological effects.

Richness – In analysis of biological communities, refers to the variety of organisms present in a sample (e.g., the variety of plants or invertebrates); the value of richness can be determined by summing the number of unique taxa present in the sample.

Risk characterization – The process of estimating the magnitude (and where relevant, the probability) of adverse ecological impacts based on the information obtained from the exposure and effects assessments. Risk characterization also translates complex scientific information into a format that is useful for risk managers, by conveying the ecological consequences of the risk estimates along with the associated uncertainties.

Safety factor – Also called an application factor, uncertainty factor, or extrapolation factor. A numerical factor sometimes used in effects assessment and applied to observed endpoints in order to derive an exposure concentration below which adverse effects are unlikely to occur. The factor is applied in the face of uncertainty, and applied in order to not underestimate risk. As the quantity and quality of test data increases and their relevance to the organisms of interest improves, the size of the extrapolation factor diminishes. This guidance advises against indiscriminate use of safety factors and recommends other techniques for assessing uncertainty.

Sensitivity – The quality of being able to reliably detect perturbations in a parameter.

Spatial – Relating to space, particularly in terms of the lateral (horizontal) dimension. In ERA, the term spatial is often used to refer to level of resolution (grain) and extent (area)..

Species sensitivity distribution (SSD) – A cumulative probability distribution of toxicity values for multiple species.

Standard – An environmental benchmark subject to regulatory enforcement. Most standards are associated with specific environmental legislation that conveys the responsibilities of site owners.

Statistical power – The probability that the test will properly reject a false null hypothesis (i.e., that it will not make a Type II error). The probability of a Type II error is referred to as the false negative rate (β). Therefore power is equal to 1 − β. Although there are no formal standards for power, many researchers assess the power of their tests using 0.80 as a standard for adequacy. Factors influencing the power of a given test (or study design) include: (1) the statistical significance criterion for probability of a Type I error (α); (2) the magnitude of the effect of interest in the population; (3) the sample size (*n*); (4) the variation of the underlying data, as determined by measurement error and stochasticity.

Stochasticity – Random natural variations; stochastic processes can be simulated, but the variations cannot be reduced through additional analysis, only better described.

Stressor – any substance or process that may cause an undesirable response to the health or biological status of an organism.

Surrogate ROC – a surrogate ROC that is representative of a receptor type (e.g., a shrew may be used as a surrogate ROC for insectivorous mammals). More than one surrogate ROC may be used to represent a particular receptor type.

Taxon (plural is taxa) – A grouping of organisms given a formal taxonomic name (biological classification) such as species, genus, family, and identified as genetically distinct from other organisms.

Temporal – Relating to time, particularly in terms of changes or variations observed over a time period of interest.

Threshold – Dividing line (in units of exposure concentration or dose) between a zone of potential response and a zone of negligible response. Thresholds may be estimated using theory, data, or a combination of both. In nature, thresholds generally do not occur as precise or static entities, due to the variations among individuals and environmental factors that influence responses. Therefore, a threshold is usually expressed as a best estimate considered protective of most of the population, and often includes a margin of safety in the derivation.

Tissue residue guidelines (TRG) – Regulatory criteria or guidelines that refer to an internal body or tissue concentration in a receptor.

Toxicity – The observation of a chemically-induced physiological or biological response that impairs the health of an organism.

Toxicity identification evaluation (TIE) – A tool in which physical/chemical manipulation of a sample is conducted to isolate and to identify toxic substances in a test medium. A biological test, in this case a toxicity test, is used as the "indicator" to determine whether the manipulation changed toxicity.

Toxicology – The field of science that explores the relationship between substances of environmental concern and the responses elicited to organisms.

Toxicity reference value (TRV) – An exposure concentration or dose that is not expected to cause an unacceptable level of effect in receptor(s) exposed to the contaminant of potential concern. A TRV is a specific type of *threshold*, as defined above.

Type I error – Synonymous with false positive – The error of rejecting a null hypothesis when it is actually true. A Type I error occurs when we are observing an apparent difference when in truth there is no difference, thus indicating a test of poor specificity.

Type II error –Synonymous with false negative – The probability that the test will not reject an invalid null hypothesis. The probability of a Type II error is referred to as the false negative rate (β). This is the error of failing to observe a difference when in truth there is one, thus indicating a test of poor sensitivity.

Uncertainty – Uncertainty is a term used in subtly different ways in a number of scientific fields. Generally, it refers to imperfect knowledge regarding a given parameter, process, or condition. In risk assessment, uncertainty is the state of having limited knowledge where it is impossible to exactly describe an existing state or future outcome. Uncertainties come in many forms, including measurement uncertainty, random variations, conceptual uncertainty, and ignorance.

Univariate – Statistical tests that address one variable at a time. The term also applies to statistical tests for comparing two or more groups with respect to a single property, including the Student t-test, analysis of variance, sign test, Wilcoxon rank test, and the Mann-Whitney U-test.

Upper confidence limit of the mean (UCLM) – A statistical measure of the upper bound of a confidence interval for the mean value of an environmental parameter, such as the expected environmental concentration of a substance.

Uptake factor – A factor used to extrapolate contaminant concentrations from a single abiotic exposure medium to a tissue concentration in an organism. Several types of uptake factors exist, including the BCF, BAF, and BSAF.

Valued ecosystem component (VEC) – for purposes of ERA, this term should be considered synonymous with *receptor of concern* (ROC). The term VEC originates in

Environmental Impact Assessment literature. Either term can be used by practitioners, but ROC is used exclusively in this guidance document.

Weight – The degree of emphasis placed on a finding or *line of evidence* relative to others. The weight is a function of the overall value (information, reduction of uncertainty) in terms of addressing an *assessment endpoint*, and is determined by assessing the *attributes* relevant to the study.

Weight-of-evidence (WOE) – A systematic procedure used to aggregate or synthesize a number of different types of evidence, with the objective of developing a single unified conclusion or explanation to an environmental characterization. WOE is one of the tools applied during the risk characterization stage of ERA.

Wildlife – In the context of ERA, the term is generally applied to birds and mammals, and sometimes defined to include reptiles and amphibians. Generally it excludes fish and invertebrates.

1. INTRODUCTION

This section presents the context for use of ecological risk assessment (ERA) at federal contaminated sites in Canada. The general framework for ERA is introduced, and the scope and intent of the guidance is articulated.

1.1. Background

The Federal Contaminated Sites Action Plan (FCSAP) was developed to support federal departments, agencies and consolidated crown corporations in their efforts to reduce the risks to human health and the environment, as well as the financial liabilities associated with federal contaminated sites. Under FCSAP, ecological risk assessments (ERAs) are commonly used as a site management tool at federal contaminated sites. The basic ERA framework is described in CCME (1996, 1997a). This technical guidance document is intended to support federal custodians and risk assessment practitioners when conducting ecological risk assessments.

1.2. Why Conduct an ERA?

> ***Key Concept:***
>
> ERA helps determine whether and to what extent remediation or other risk management measures are needed.

Once a site is classified as contaminated, ERA is used to determine whether and to what extent remediation or other risk management efforts are warranted to mitigate current or future ecological risks. In some cases of limited contamination, the cost of remediation is low relative to the cost of further assessment, and the ecological impacts of remediation are negligible – these cases are not destined for ERA, or at least not for detailed ERA. Rather, ERA is appropriate for sites where the costs and/or ecological impacts of remediation are likely to be large relative to the cost of assessment. ERA provides a basis for determining whether remediation or other risk management measures are warranted (i.e., are there ecological risks?) and to what extent (e.g., which parts of a site should be remediated?). The costs of remediation or other risk management measures may ultimately be much lower using a risk-based approach compared to an approach based on comparison of contaminant concentrations to environmental quality guidelines.

1.3. ERA at Federal Contaminated Sites

There are numerous potential drivers for the use of ERA at federal contaminated sites, such as regulatory triggers (e.g., contamination of an off-site property), due diligence, or divestiture. In cases where sites will remain under federal jurisdiction, the regulatory

context will be mainly federal. On the other hand, in cases of divestiture where sites will be transferred from federal to another jurisdiction, or where there is off-site migration of contaminants beyond federal site boundaries, the ERA may be driven in part or entirely by provincial or territorial regulations and policy. This guidance is not constrained by federal or other contexts; rather the guidance attempts to focus on technical aspects of ERA that are likely to be applicable in many contexts, depending on the complexity of the ERA.

> ***Key Concept:***
>
> Many federal sites, particularly those intended for possible divestiture, may be subject to both federal and provincial ERA regulations and policy.

1.4. Roles and Responsibilities

The implementation of ERA involves various parties. Site custodians are responsible for the management of federal contaminated sites. The four FCSAP expert support departments – Health Canada (HC), the Department of Fisheries and Oceans (DFO), Environment Canada (EC), and Public Works and Government Services Canada (PWGSC) – provide technical and scientific advice on various aspects of contaminated sites management including ERAs. DFO and EC provide advice and guidance for ERAs, HC focuses on human health risk assessment, and PWGSC assists with project management issues. All four expert support departments are also responsible for providing advice and guidance on the application and interpretation of federal and provincial policies, guidelines and programs that may relate to federal contaminated sites; for promoting compliance with regulatory requirements and guidance; and for serving as a liaison with provincial and territorial governments. Risk assessors are responsible, under the direction of the site custodians, for implementing ERA at contaminated sites. In addition to site custodians, FCSAP expert support, risk assessors, and provincial regulators (in cases of divestiture), there may be specific stakeholders or other interested parties associated with particular sites. Consequently, it is appropriate for site custodians, expert support departments and risk assessors to proactively encourage communication and early involvement of the various parties in the ERA process. The site-specific consultation needs may include up-front dialogue prior to commencement of work, as well as dialogue at milestones during the ERA process (e.g., review of a problem formulation).

1.5. Introduction to the ERA Framework

> ***Key Concept:***
>
> Most ERAs warrant a weight of evidence (WOE) approach, whereby multiple lines of evidence are used to support the assessment. The WOE approach is entirely compatible with the standard conceptual framework for ERA.

The standard conceptual framework for ERA (**Figure 1-1**[2]**)** applies to federal contaminated sites. This framework is consistent with existing CCME risk assessment guidance; however, the science of ERA is constantly evolving, and the last two decades have included a significant increase in the complexity of risk assessments and in the tools and methods used to characterize risks. Whereas the conceptual framework appears simple, its application to multiple receptor groups via multiple exposure pathways using various lines of evidence can be quite complex. Consequently, in practice, the ERA framework is often applied using a weight of evidence (WOE) approach (**Figure 1-2**). In application, the WOE approach may be simple (e.g., a couple of lines of evidence for a single assessment endpoint) or complex (i.e., for complex sites).

The weight of evidence approach (**Figure 1-2**) integrates with the standard ERA framework as follows:

- Problem formulation defines the problem to be addressed and develops the scope for the ERA.
- For each receptor group or assessment endpoint, one or more lines of evidence (LOEs)[3] are used in the risk assessment. Each LOE must combine information on exposure and effects. The exposure information typically characterizes the extent to which receptors are exposed to contaminants via various exposure pathways.

> ***Definitions:***
>
> A receptor of concern in ERA is any non-human individual organism, species, population, community, habitat or ecosystem that is potentially exposed to contaminants of potential concern and that is considered in the ERA. *Examples* – meadow vole population; benthic invertebrate community.
>
> An assessment endpoint is an explicit expression of the environmental value to be protected, and must include a receptor or receptor group (i.e., an 'entity to be protected) and a specific attribute of that entity. *Example* – abundance and viability of small mammal populations. Spatial and temporal elements may also be included.

[2] The precise terminology and delineation of the components of ERA vary across different jurisdictions and applications, but the vocabulary outlined here is used relatively consistently.

[3] The relationship between assessment endpoints, measurement endpoints and lines of evidence is considered in detail in Section 2.

Figure 1-1. Generic Framework for Ecological Risk Assessment (simplified)

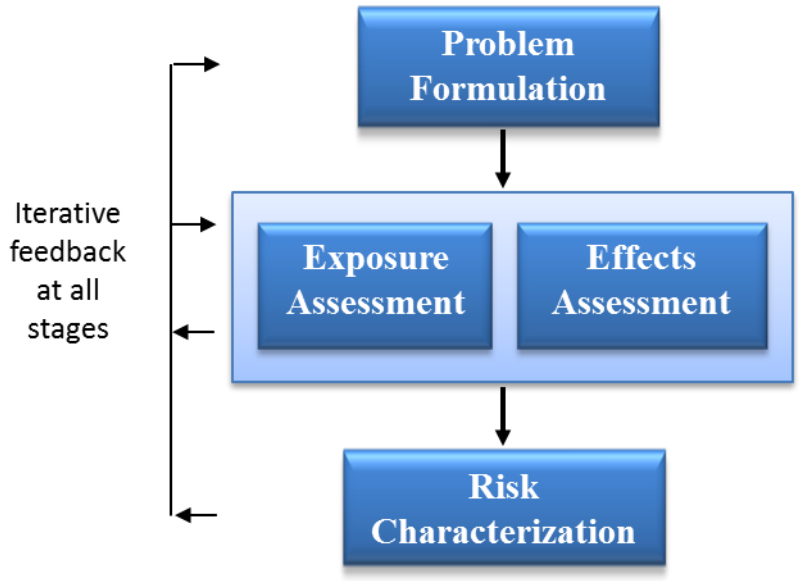
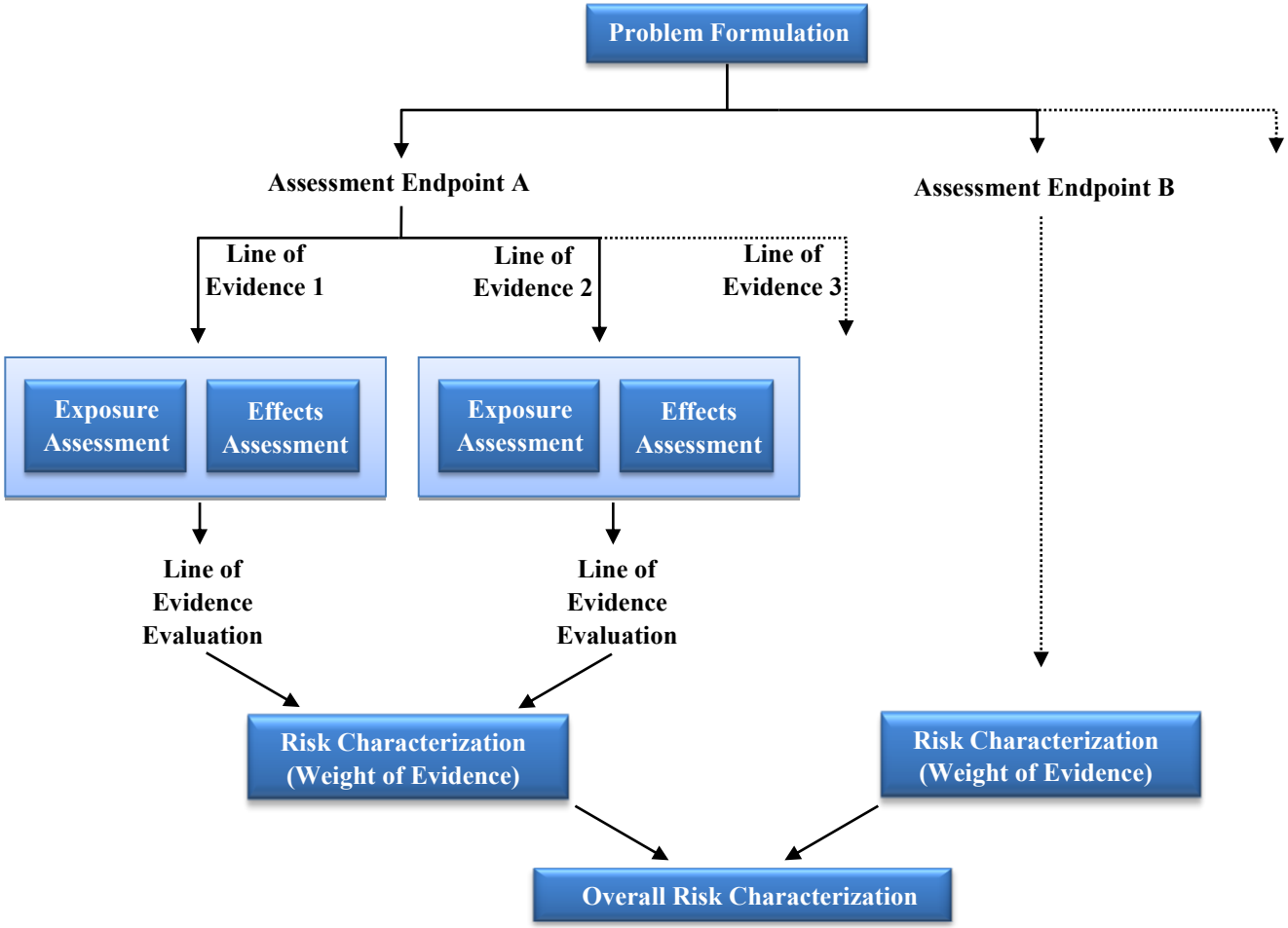


Figure 1-2. Weight of Evidence Approach to Ecological Risk Assessment



The effects information characterizes the nature of effects observed or expected at the site.

Once individual lines of evidence are evaluated, the findings across all lines of evidence are evaluated in an integrated fashion to characterize risks for a particular assessment endpoint or receptor group.

> ***Key Concept:***
>
> Each Line of Evidence (LOE) in an ERA combines information on exposure and effects in order to evaluate evidence for risk for that LOE.

1.6. Iterative Approach to Risk Assessment

In many cases, risk assessment follows a tiered approach, in which screening tools are applied at an early stage to determine if further work is needed or to prioritize future investigations. For example, a screening-level risk assessment might use conservative, simplifying assumptions to assess risks. If risks are acceptable using conservative simplifications, there is probably no need for further work. On the other hand, if potential effects are identified, a more detailed and accurate risk assessment may be warranted.

> ***Key Concept:***
>
> If an ERA based on conservative, simplistic assumptions shows no potential for risks, then a more detailed ERA is probably not warranted.

This document does not categorize types of risk assessments according to scope or level of detail (e.g., screening-level versus detailed risk assessment). Some regulatory or policy frameworks may have specific requirements in this regard. In practice, the process of tiering an ERA and the appropriate level of detail for each iteration is driven by many factors and is case-specific (Hill *et al.* 2000). The various parties involved in any particular ERA should agree on the expectations for each iteration of an ERA, particularly with respect to the type and degree of uncertainty that is expected to be resolved at each stage of investigation. Generally, it is important for each iteration of an ERA to address issues and uncertainties that are important from a risk management or decision-making perspective. In other words, each iteration of an ERA should significantly advance the usefulness of an ERA to support sound environmental management of contaminated sites.

1.7. Scope of Guidance

1.7.1. Level of Detail

This guidance document and supporting technical modules contain a high level of detail regarding many aspects of ERA. Consequently, some of the methods and approaches presented may be applicable only to complex sites where detailed ERA is warranted. Importantly, and in accordance with the iterative approach to ERA articulated in the previous section, the level of complexity in an ERA should be commensurate with the level of complexity of the site and its associated risks, taking into account the role of the ERA in supporting risk management decision-making. Practitioners must judge the appropriate level of detail for each ERA on a site-specific basis.

As mentioned in the previous section, this guidance document does not distinguish screening-level and detailed-level ERAs. Earlier CCME guidance on ERA (CCME 1996, 1997) differentiated screening-level and detailed-level ERA based on particular methods (e.g., hazard quotients were envisioned to be used only as part of screening-level ERAs). Although current guidance for human health risk assessment (Health Canada 2004) differentiates "preliminary" and "detailed" quantitative risk assessments, such differentiation would be much less practical for ERA given the range of receptors considered and the range of methods used to evaluate risks.

1.7.2. Organization of Document

This guidance document is organized around the conventional ERA framework, with major sections addressing problem formulation, exposure assessment, effects assessment, and risk characterization. This is intentional because the types of tools used for exposure assessment and effects assessment fall into major categories that can conveniently be discussed simultaneously. Because exposure, effects, and risk characterization apply to each line of evidence in an ERA, the concepts in all of the sections of this guidance must be understood prior to undertaking an ERA.

The introductions in each of **Sections 2 to 5** provide an overview of problem formulation, exposure assessment, effects assessment, and risk characterization respectively. Those introductions, together with text boxes on "key concepts" scattered throughout the text provide a sense of the scope and content of the guidance without complex technical details.

This guidance document aims to provide relatively detailed technical guidance, but it cannot be comprehensive in all aspects. Selected aspects of ERA are addressed in detail in the Technical Modules that are appended to the guidance document, and other modules may be added in future.

1.7.3. Other Sources of Guidance

The basic elements of ERA are described in numerous publications. Many documents have been developed in the context of particular regulatory regimes, and so the policy aspects of such documents may be inapplicable to some or all sites in Canada. At the federal level in Canada, the general framework has been described by the CCME (1996, 1997a, 1997b) – although these documents are still relevant and useful, they are not comprehensive and do not necessarily reflect all aspects of current ERA practice in Canada.

Guidance on ERA from Canadian provinces or territories tends to focus on specific aspects of ERA and not on the overall ERA framework. Such guidance is referred to as appropriate in **Sections 2 to 5** of this guidance document in the context of particular technical issues. There are a few cases where provincial guidance is more comprehensive – one is the guidance document on Detailed Ecological Risk Assessment developed in British Columbia (SAB-CS 2008). Although the policy elements of that guidance document are not always relevant outside of BC, the technical content is relatively detailed and reflects best available practice in ERA. Another relevant provincial guidance document is guidance in Ontario for implementation of risk assessment under the *Environmental Protection Act* (OMOE 2005); the policy elements of that document would also not be relevant for federal sites except in certain cases in Ontario. The province of Quebec has also developed guidance on an overall ERA framework (CEAEQ 1998). Finally, although specific to sediments, the Canada-Ontario decision-making framework for assessment of Great Lakes contaminated sediment also covers the key aspects of an overall framework for ERA (Environment Canada and OMOE 2008).

Numerous guidance documents on ERA have been developed in the United States, in particular by the Environmental Protection Agency (USEPA) and the National Research Council (NRC). The basic ERA framework is described in USEPA (1992) and USEPA (1998). Many of the other USEPA documents are particular to certain cases (e.g., Superfund sites), and most of the NRC documents address particular issues in the practice of ERA (e.g., NRC 2009). Practitioners are encouraged to consult the websites of the USEPA and the NRC to evaluate the potential usefulness of these documents and others that will be developed over time. The USEPA has recently compiled a table listing documents relevant to ecological risk assessment (appendix C in USEPA 2011).

In addition to guidance provided by government agencies, there are numerous books that address the process and technical elements of ERA. Two of the most common reference books are Suter (2007) addressing the generic framework for ERA, and Suter *et al.* (2000) addressing ERA for contaminated sites. Advanced ERA practitioners interested in detailed technical guidance on particular aspects of ERA are referred to the technical modules appended to this guidance, the technical appendices of SAB-CS (2008), and Suter *et al.* (2000).

2. PROBLEM FORMULATION

2.1. Introduction

2.1.1. Purpose

Problem formulation is the important first step in ERA that clarifies the nature of issues associated with contamination at a site and how those issues will be addressed. The specific objectives of the problem formulation are to:

> ***Key Concept:***
>
> It is important for the practitioner to document all of the assumptions and decisions made during the problem formulation so that site custodians and reviewers can understand the rationale and judge whether the scope of the ERA is adequate.

- Frame the issues, including the goals, context, and nature of potential effects.
- Design and plan an approach to assess risks, specifying the tools that will be used and how the results will be evaluated.

2.1.2. Overview of Problem Formulation

Problem formulation generally entails the following elements:

- Description of the ***site management goal(s)*** and the specific assessment goal of the ERA. For example, if a site management goal is to reclaim a site as parkland, the assessment goal of the (initial) ERA may be to assess whether there are potential effects under current conditions relative to *protection goals* for parkland.
- Review of the ***regulatory context*** for the site and the ERA, including applicable legal instruments and policy.
- Review of ***existing site information***, including at a minimum a list of relevant documentation; a site description; and a summary of key findings from previous investigations. For some complex ERAs, such a review may warrant a stand-alone chapter or document attached to the problem formulation.
- Selection of ***contaminants of concern*** (COCs) and description of their characteristics which are relevant to the ERA (i.e., transport and fate, as well as effects).

> ***Definition:***
>
> A contaminant of concern is a contaminant that has been selected for evaluation in the ERA.

Definitions:

A surrogate ROC is a receptor of concern that is representative of a receptor type (e.g., a shrew may be used as a surrogate ROC for insectivorous mammals).

Exposure pathways are the routes of exposure from environmental media (e.g., soil, water, air, sediment) to the receptors of concern. Examples of exposure pathways include ingestion and inhalation.

A conceptual site model (CSM) is a narrative and/or graphical representation of the relationships between contaminant sources, exposure pathways, and receptors.

- Selection of ***receptors of concern*** (ROCs) that could be affected by contamination and that will be evaluated in the ERA. Receptors can be identified at the level of individual organisms, species, populations, communities or habitats. Importantly, it is usually not feasible (or necessary) to include every possible species in an ERA; therefore a subset of candidates are selected as ***surrogate ROCs*** for particular types of receptors.
- Identification of the ***exposure pathways*** by which COCs may come into contact with the receptors of concern. Examples of exposure pathways include water and food consumption (for wildlife) and direct contact (for invertebrates).
- Development of a ***conceptual site model*** (CSM) that shows the potential links between source of contaminants, exposure pathways, and receptors of concern.

Definitions:

A protection goal is usually a narrative statement that defines the desirable level of protection for a receptor or receptor group. An acceptable effect level operationalizes the protection goal by specifying the magnitude (or rate) of effects that would be acceptable for a specific measurement endpoint or assessment endpoint.

A measurement endpoint is a parameter that measures or describes exposure for, or an effect on, a receptor of concern, or that measures or describes a change in an attribute of an assessment endpoint or its surrogate in response to a stressor to which it is exposed.

A line of evidence (LOE) is any pairing of exposure and effects measures that provides evidence for the evaluation of a specific assessment endpoint. Typically a line of evidence requires use(s) of one or more measurement endpoints. If the focal point of the LOE is an effects measure (e.g., a toxicity test), the paired exposure measure may be quantitative (e.g., contaminant concentrations) or categorical (e.g., on-site versus reference).

- Clarification of ***protection goals*** and associated ***acceptable effect levels*** (AELs). Typically, protection goals and AELs may vary by land use or by receptor (e.g., species at risk may be afforded a higher level of protection than common species).
- Identification of ***assessment endpoints***, which are attributes of receptors (the entities that are to be protected), often with specific spatial and temporal components. An ERA may have one assessment endpoint for a receptor group (e.g., ecological function of the soil invertebrate community) or there may be more than one assessment endpoint for a receptor or group of receptors.

> ***Key Concept:***
>
> An assessment endpoint describes an attribute of a receptor or receptor group, but does not articulate a desired state for that attribute.

- Identification of ***measurement endpoints***, which are the tools used to measure exposure for, or effects on, a receptor, or to measure changes in attributes of assessment endpoints.
- Development of ***lines of evidence*** for each assessment endpoint, which specify how measurement endpoints will be used to evaluate potential risks.

> ***Key Concept:***
>
> It is helpful to view measurement endpoints as tools, and lines of evidence as the use of those tools in one or more ways.

- Articulation of the ***general strategy*** for the ERA including how risk characterization will be conducted, and a ***sampling and analysis plan*** (SAP). In some cases, for example for complex ERAs with many components, the SAP may be prepared as a stand-alone document separate from the problem formulation.

The rest of **Section 2** explores each of these problem formulation elements in more detail. Although the elements of a problem formulation are presented in a linear fashion, in fact most elements need to be developed together using an iterative process. Furthermore, because almost all of the planning for an ERA occurs during the problem formulation, the content of **Sections 3 to 5** must be fully considered during development of the problem formulation.

> ***Key Concept:***
>
> Begin with the end in mind – a proper problem formulation does not simply result in a list of tools to be used for the ERA, but also specifies how the results will be evaluated.

2.2. Framing the Issues

2.2.1. Site Management Goals

> ***Key Concept:***
>
> A specific and clearly defined site management goal provides direction to risk assessors and site custodians.

At a broad level, ERA is guided by the overall site management goals. In the context of ERA, a management goal for a contaminated site is the overall planning objective for the site, usually worded as a statement about the desired condition of the ecosystem or its components in the context of future site use. Site management goals may be relatively generic and stated at a high level (e.g., "maintain a sustainable aquatic community adjacent to a ferry terminal"). In other cases, more specific management goals may be identified, such as:

- To determine whether contaminants (COCs) present in the surface soil layer require remediation within the existing provincial and federal regulatory frameworks for a particular land-use category;
- To determine whether intrusive remediation is warranted at a contaminated wetland adjacent to a federal airport.
- To develop a management plan for a Department of National Defense facility, that is protective of a specific federally-listed species at risk.

These more specific site management goals are generally preferable because they provide direction to risk assessors and site custodians. Specific site management goals are often developed through discussion with regulators, site custodians and stakeholders. Such dialogue can clarify how the ERA will be used to support risk management and decision-making. For example, if there are only two management options for a site, the ERA could be tailored to inform a decision about which option is preferable.

Site management goals provide the overall framework under which the components of the problem formulation are developed. A site management goal should not be confused with a protection goal (which is related to the desired level of protection for ecological receptors – see Section **2.3.1**), although protection goals are derived in part based on understanding of the site management goals.

2.2.1.1. Determining the Broad Assessment Goals

One of the potential pitfalls of ERA in practice is that individual practitioners are prone to applying identical approaches at different sites, even where site-specific considerations (including management issues) require different techniques. Therefore, it is imperative that risk assessors do the following:

- Consider the overall purpose of the risk assessment before selecting or interpreting measurement endpoints, tools, or techniques;
- Plan the study design to consider the fundamental underlying questions of interest; and
- Provide output in a format useful to the risk manager for making decisions.

In the case of FCSAP sites, a common goal for ERA is to determine what, if any, management or remedial action is needed to reduce liability. However, this may not be the only goal, and different risk assessments could require different approaches (and tools) depending on the context of the investigation. The goal of a risk assessment is not as a purely scientific endeavor, but rather as a management tool, in which the analysis should proceed only to the point at which information needs for risk managers can be met. Barnthouse (2008) notes that "ERA is best viewed as a bridge between science and management," rather than as a conventional scientific discipline like chemistry, toxicology, or ecology. Management decisions accommodate multiple goals and constraints, with a need to reconcile information collected across disciplines, scales, and types of evidence.

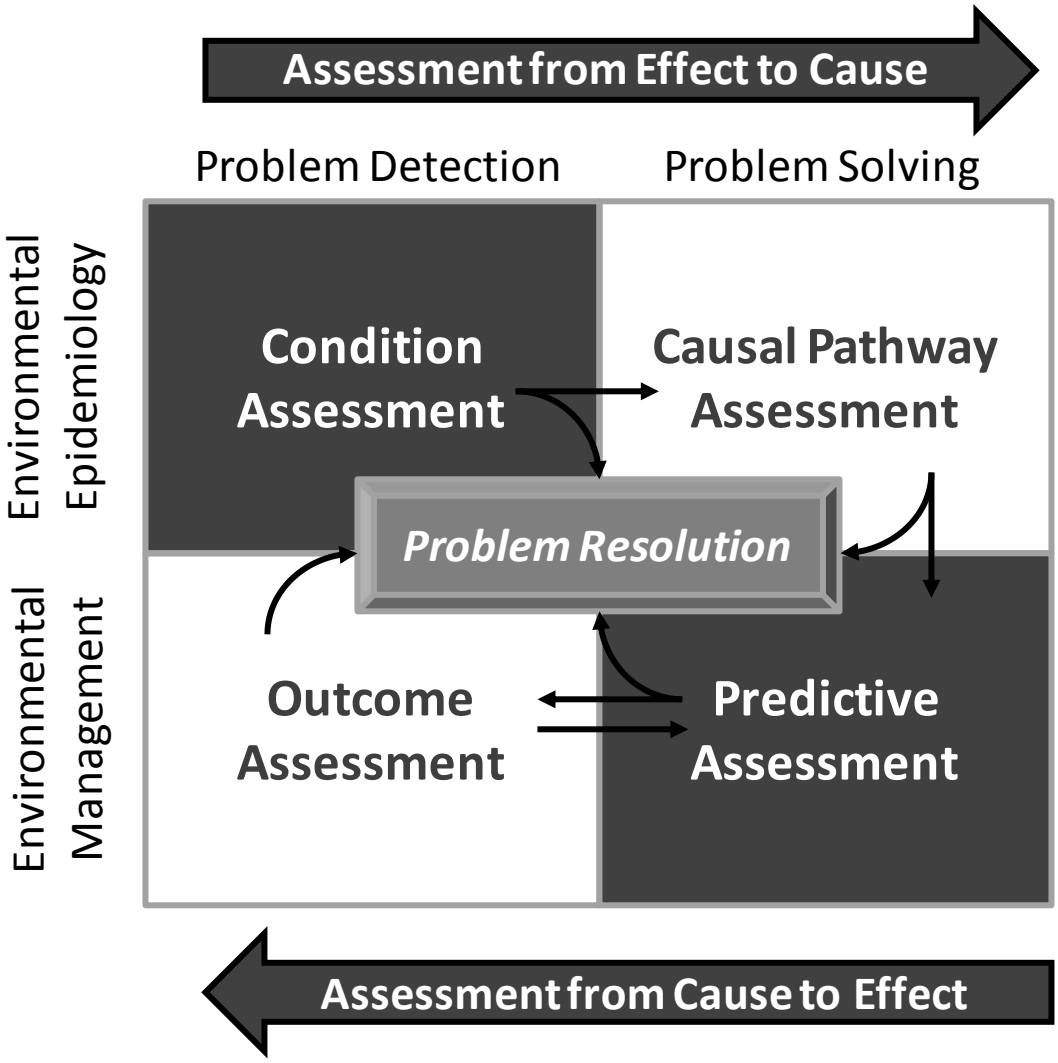
> ***Key Concept:***
>
> Every ERA should be tailored to site-specific considerations and context. Evaluation of 'current conditions' should not be assumed to be the default assessment goal. There may be need to evaluate risks under various scenarios.

To assist in framing management needs, a framework derived by Cormier and Suter (2008) conceptualizes four different themes of environmental risk assessments[4], as depicted in **Figure 2-1**:

- **Condition assessment** – purpose is to detect chemical, physical, and biological impairment, through analysis of environmental monitoring data;

[4] The term environmental risk assessment is used to distinguish human and ecological risk assessments from broader environmental assessments that incorporate social, cultural, and economic analyses. However, environmental assessments, as defined by Cormier and Suter (2008), include any science-based assessments used to inform environmental management decisions, where such decisions accommodate multiple goals and constraints. The framework depicted in Figure 2-1 applies to all types of environmental assessment, although we have emphasized the application to contaminated sites herein.

Figure 2-1. Environmental Risk Assessment Framework as Developed by Cormier and Suter (2008).



- **Causal pathway assessment** – purpose is to determine proximate causes and identify their sources, and where possible, characterize the causal pathways that connect them;
- **Predictive assessment** – purpose is to estimate environmental, economic, and societal risks, and benefits associated with different management alternatives. Acceptability of actions may be determined through evaluating the risks in light of social, economic, and legal considerations; and
- **Outcome assessment** – purpose is to evaluate the results of a past management action, through estimation and/or direct measurement.

The framework recognizes that all assessment types can potentially resolve an environmental issue, or prompt a refined assessment in a subsequent tier of investigation.

These four broad themes of assessments are organized based on two key questions:

- Are we interested in explaining what has already occurred in terms of environmental effects (environmental forensics) or in extrapolating our knowledge to the prediction or optimization of future conditions (environmental management)?
- Are we interested in simply detecting an environmental response (condition assessment), or in assigning or allocating responsibility among sources (causal pathway assessment)?

This framework is useful for focusing the risk assessment objectives, particularly by asking the two key questions above and then organizing the study design and evaluation to answer them. It is also possible to tier an investigation, such that resolving issues of causation and prediction is deferred pending the results of a preliminary risk assessment.

This framework is applicable to all environment risk assessment scenarios, but is particularly well suited to the assessment of contaminated sites, as discussed below:

- **Condition assessment** – this may determine whether a site is sufficiently contaminated to warrant further assessment, or it may include a biological condition assessment to determine whether there is evidence of impairment from the site. In general, the initial stages of condition assessment rely on chemical characterization, but progress to toxicological and biological tools at more detailed stages of investigation;
- **Causal pathway assessment** – some level of causal assessment is incorporated into all contaminated site investigations because the preliminary and/or detailed site investigations identify sources (i.e., areas of potential environmental concern) and contamination pathways, at least at a broad level. However, the importance of causal assessment increases for some contaminated site scenarios, such as where multiple responsible parties contribute to contamination, or where non-

contaminant stressors may influence the pattern of observed responses. In these cases, the causal assessment explores in more detail (and with increased emphasis on quantitative methods and/or mechanistic understanding) the linkage between exposure and effect.

- **Predictive assessment** – the degree of prediction required in a contaminated sites assessment is a function of the range of potential site uses contemplated over time. Where a site is proposed for divestiture, but without a foreseeable change in site use, risk assessments may rely on empirical information from existing site characterization. Conversely, scenarios of significant redevelopment or remediation often trigger the need to model or predict future conditions of contamination and their influence on risk estimates. Changes in site use may result in revised assumptions regarding exposure (e.g., revised calculations of exposure concentrations or doses), effects (e.g., revised toxicity estimates based on changing contaminant fingerprint over time), and risk management alternatives (e.g., administrative controls on site use).
- **Outcome assessment** – in a contaminated sites application, an outcome assessment entails evaluation of multiple "what if" scenarios, with the objective of determining whether risk estimates can be meaningfully influenced by actions taken by the risk manager. For example, where baseline risks are considered to be unacceptable, a remedial options analysis can be undertaken to evaluate the impact of different management alternatives, including monitored natural recovery.

Although the Cormier and Suter (2008) framework is simplified and conceptual, it is possible to frame the core risk assessment needs through consideration of site-specific issues. The four themes are not mutually exclusive, so it is possible to draw elements from multiple themes (quadrants) to develop an assessment framework that is appropriately customized to the site context. To assist in refining the broad assessment goals beyond the simple four-theme framework, and to operationalize the framework to potential FCSAP site-specific issues, it is helpful to pose questions that inform the selection of tools. For example:

- Are there multiple potentially responsible parties associated with the contamination (industries, site owners, point sources, legally responsible entities)? [if yes, consider role of *causation*]
- Is there a need to extrapolate results to other parcels or conditions [if yes, consider importance of *predictive* tools]
- Is the study intended to detect environmental changes in response to source control or other management actions, such as remediation of contaminated soils up-gradient of a harbour facility, or monitoring of tailings treatment at an abandoned mine site? [if yes, consider *outcome assessment* tools]

- Is the site within a context of significant regional background contamination? [if yes, consider role of *causation*]
- Is the site large and/or complex in terms of physical, chemical, and biological conditions? [if yes, consider tools for extrapolating across space, time, and/or habitat / substrate type]
- Are the processes affecting contaminant transport, accumulation, and toxicity already well understood? [if yes, the need for *causation assessment* and/or refining *predictive tools* may be lower]
- Is the study designed to screen or rank priorities for future tiers of study, as opposed to detailed remediation design? [if yes, consider initial *condition assessment*]
- Does the site contamination affect off-site parties or sensitive habitats, as opposed to being a site-specific management issue? [if yes, the need for *causation assessment* and/or refining *predictive tools* may be greater]
- Are there known confounding factors to direct assessment of risks, such as physical habitat modifications, mechanical disturbance, etc.? [if yes, consider the need for *causation assessment* and/or refining *predictive tools*]
- Is there potential cost-savings through use of an adaptive management approach, and is there adequate time available for such an approach? [if yes, consider initial *condition assessment* to optimize resources, followed by other approaches]
- Does the ERA require an evaluation of stressors that are either not contaminants and/or that may confound the assessment of a primary contaminant (e.g., biological influence of cattle grazing, regional organic enrichment, or invasive species)? [if yes, consider the need for *causation assessment* and/or refining *predictive tools*]
- Is the ERA intended to evaluate the relative risks (or benefits) of alternative management approaches at the site? [if yes, consider the need for *predictive tools*, and *outcome assessment*, with use of a formal decision analysis framework to guide management decisions]

In practice, the type of assessment is also related to the costs of investigation relative to the potential cost of remedial action or environmental liability. If the site is small and potential remedial costs are low, there would be a tendency to apply relatively simple approaches in a condition assessment. Conversely, if the site is large, complex and has multiple potential stressors, the value of causative and predictive (extrapolative) assessments may increase. In the latter case, the return on investment tends to be greater for the site manager because significant liabilities can be eliminated through application of science to screen pathways, contaminants, or management units.

Predictive assessments that focus on future risk scenarios are common in ERA. Future risks may differ from current risks for many reasons including:

- Implementation of risk management measures such as remediation or fencing;
- Natural attenuation may occur due to physical or chemical processes (dechlorination, burial by settlement of relatively clean material);
- Changing human use of the site, including addition or removal of infrastructure;
- Natural ecological succession – for example, if a site is no longer subject to human use, natural ecological processes may result in changes to ecosystem type along a gradient of disturbance level (e.g., from plowed fields associated with agricultural land to a forest type that may become established);

In cases where both current and future risks are estimated, it is possible to estimate the expected change in risks that may occur. This comparative approach can be useful for evaluating the likely effectiveness of risk management measures.

Note that the assessment type is not necessarily static, but rather may progress from one type to another based on feedback from early stages of investigation, as implied by the arrows in Figure 2-1. FCSAP investigations often begin with condition assessments, in which the primary objective is to determine whether the current site conditions are acceptable. Depending on the results, subsequent tiers of analysis may require increased emphasis on causality and/or prediction of changes over time.

2.2.2. Regulatory Context

The regulatory context for an ERA is important for determining the scope of the ERA and technical constraints.

> ***Key Concept:***
>
> An ERA for a federal site that is intended for divestiture may be subject to provincial legislation and policy.

Legal Instruments and Policies – The problem formulation should acknowledge the various federal and other (e.g., provincial) legal instruments and policies that are applicable for a particular site, and should promote consistency of the ERA with those legal requirements and policies. Examples of relevant federal legislation include, but are not limited to, the *Fisheries Act*, the 1999 *Canadian Environmental Protection Act*, the *Species at Risk Act*, the *Migratory Birds Convention Act*, the *Canadian Environmental Assessment Act*, and the *Canada National Parks Act*. There are numerous other potentially relevant legal instruments at federal and provincial levels (see SAB-CS 2008 for some further discussion). The regulatory context can have direct implications on the technical details of the ERA. For example, the protection goal defined for a species at risk (e.g., a rare or endangered

species) may be much different than for a common species. The *Species at Risk Act* requires protection of individual organisms of a species at risk, whereas for some common species an ERA may aim for protection at population level. In addition, certain aspects of contamination may not be the primary focus of an ERA if they are addressed by other regulations (although cumulative impacts should not be ignored); for example, some discharges of contaminants are permitted under other regulations. Depending on the goals of the ERA, it may or may not be relevant to explicitly consider the effects of such discharges (e.g., a risk assessment of a federal water lot conducted in the vicinity of a municipal effluent discharge may need to consider the effects of the discharge in order to discriminate between potential sources of impairment).

Land Use – Land use designations are usually important in determining whether or not a terrestrial site is contaminated, because the screening guidelines for a given contaminant often vary by land use. In addition, land use (either the designation, or the actual land use) affects an ERA in other important ways. First, policy developed for technical aspects of ERA may be specific to land use. For example, protection goals may be different in a park compared to an industrial property. Second, actual land use at a site may limit the scope of risks (particularly exposure pathways) that need to be considered. For example, if a site does not have (in either current or potential future uses scenarios) any exposed surface soil, many receptor groups would not be present. Third, land use in the areas surrounding a site may also limit the scope of an ERA. For example, if a site exists in the middle of an urban centre, consideration of large mammal receptors may be unnecessary. Conversely, in a relatively remote setting, a similarly-sized parcel of developed land may require consideration of large mammals that inhabit adjacent areas but use a site for food. The problem formulation should, as appropriate, highlight relevant implications for the ERA of the current and/or potential future land use of any given site.

> ***Key Concept:***
>
> Assumptions regarding future land use can influence the selection of relevant exposure pathways for an ERA.

2.2.3. Review of Existing Site Information

Every problem formulation uses existing information about a site as its starting point. Although problem formulation is the first formal stage of risk assessment, from a practical perspective, different sites have varying degrees of baseline site investigation information from which to begin the problem formulation stage. Therefore, the purpose of the review is to summarize pertinent

> ***Key Concept:***
>
> The challenge for the practitioner is to make a succinct, risk-related summary of the most relevant results from site investigations.

information on contaminant sources and distribution, transport pathways, and biological attributes of the site.

The basic information includes:

- Documentation – a list of relevant available documents about the site;
- Site description – location, setting, etc.;
- Review of previous environmental site assessments and findings (e.g., site chemistry, historical and ongoing contaminant sources, screening guidelines used, etc.); and
- If applicable, review of risk-related data for the site (in cases where previous risk-related investigations have been conducted, or if the problem formulation process has been iterative).

For some complex ERAs, the review of existing information may warrant a stand-alone chapter separate from the problem formulation.

Because of the range in abundance, type, and quality of environmental site investigation data that may be available, an important decision-point is the determination of whether supplemental site investigation may be warranted prior to advancing with formal risk assessment activities such as a problem formulation deliverable. The potential need will be site-specific, depending on how the ERA will be used to support site management goals. If there are major data gaps identified, risk assessors and site custodians should consider the value of delaying finalization of the problem formulation pending supplemental data collection. In other cases, some aspects of a project or spatial components may progress along different timelines.

2.2.4. Contaminants of Potential Concern

As previously defined, contaminants of concern (COCs) are those contaminants that have been selected for evaluation in the ERA[5]. This section reviews the broad categories of sources of COCs that should be considered in an ERA, and then focuses on the COC selection process. Finally, this section reviews the characteristics of COCs that must be understood in order to proceed to subsequent components of the problem formulation.

[5] In some jurisdictions, terms such as COPC (Contaminants of Potential Concern) or PCOC (Potential Contaminant of Concern) refer to the initial list of substances considered, whereas the term Contaminants of Concern (COC) is used to refer to the final list after the selection process conducted as part of the problem formulation. In other jurisdictions the term COC is not used at all and the final list is referred to as the list of COPCs. In this guidance document, the term COC refers to the final list of substances retained for the risk assessment at the end of problem formulation.

2.2.4.1. Sources of COCs

Understanding of the sources of COCs is important for determining likely exposure pathways. Categories of sources of COCs at a site include:

- On-site point sources (e.g., historical spills, ongoing point source effluent discharges).
- On-site non-point sources (e.g., contaminated groundwater, sediments or water; surface water runoff).
- Underground artificial conduits such as sewers, pipelines and buried structures that may contribute to contamination.
- Preferential natural pathways such as fractures in limestone geology that facilitate transport of COCs.
- Significant off-site sources (including via long range air transport) that need to be considered for their potential to confound site-related contamination and/or as contributors to cumulative risks.

Identification of sources of contamination requires a thorough understanding not only of the site itself, but also of the surrounding land use and the location of the site – this is critical for identification of off-site sources in particular. Typically, the relevant information can be summarized from site investigation documents. In fact, it is common for site assessment documents to distinguish areas of a site based on historical use and other factors, and then to identify the specific COCs associated with each area of potential environmental concern (APEC). This level of resolution for COC sources is often relevant to the ERA as well.

2.2.4.2. Selection of COCs

Selection of contaminants of concern (COCs) is an important early step in the problem formulation process. The starting point is usually the list generated from site investigation reports, although it is important to confirm whether additional COCs may be relevant prior to proceeding with the risk assessment. Often, the initial list of contaminants generated by site investigation reports is referred to as a preliminary list of COCs or a list of contaminants of *potential* concern (COPCs), which is reduced to a final list of COCs during problem formulation. Although there may be regulatory requirements to consider all of the COPCs identified during site investigation, the final list of COCs for ERA may be different for several reasons. The process used to select COCs should be agreed upon with site custodians and regulators as early as possible in the ERA process. COC selection is important for ensuring that the ERA does not miss any important

contaminants while also preventing needless analysis of contaminants that do not warrant evaluation.

This section focuses on identifying the key considerations that should guide COC selection. Further discussion of many of these issues can be found in CCME (2012) and SAB-CS (2008).

> ***Key Concept:***
>
> Canadian Environmental Quality Guidelines (CEQGs) should be used as the default for screening COCs at federal sites

1. *Applicable guidelines*[6] – At typical FCSAP sites with no other stakeholders than the federal government, Canadian Environmental Quality Guidelines (CEQGs) should be used for screening. However, for some sites, regulators and stakeholders may agree to use both national[7] and provincial guidelines, or only provincial guidelines. The latter may be relevant, for example, if there is a plan to divest a site to a province, or to an owner that would be subject to provincial regulations.

   Within a set of guidelines, there may be multiple options according to land use, water use, soil texture, transport / exposure pathways, or other factors. In such cases, rationale is needed for determining exactly which guidelines are applicable and which are not. For cases where guidelines are based on consideration of both ecological and human health components, it may be reasonable to exclude the human health component if rationale is provided.

   Also, if there are site data from multiple media, it may be appropriate for one medium to take precedence (e.g., if there is a soil guideline for the soil-to-groundwater pathway, it may be appropriate to not screen data using that soil guideline if there is also a rigorous data set for groundwater directly).

2. *Substances for which there are no guidelines* – In some cases, site-related substances may be present at elevated concentrations but may not be addressed by CEQGs, in which case provincial, territorial or other guidelines may be used (CSMWG 1999). In other cases there may be no relevant guidelines from other jurisdictions. For example, at a marine site situated in an industrial harbour, it may be appropriate to consider tributyltin (TBT), which was historically used as an anti-fouling agent in marine paints and is known to be persistent, bioaccumulative, and toxic. In such cases, decisions must be taken regarding

[6] The term "guidelines' is used loosely in this section to include numeric environmental quality guidelines, criteria, standards or any other regulatory or policy benchmark that may be used for COC screening.

[7] The CEQG are national. Separate federal environmental quality guidelines are being developed for substances that are a federal priority under Canada's Chemical Management Plan and under Section 54 of the Canadian Environmental Protection Act.

whether to include the substance in the ERA or use alternative methods of screening. Alternative screening methods may include adopting guidelines from other jurisdictions – in such cases, the level of protection inherent in such guidelines should be understood, and if it is not similar to the level of protection inherent in the CEQG then rationale is needed for their use. Another approach that has been used is screening COCs in one media using guidelines for another media (e.g., application of a safety factor to surface water data to facilitate comparison to groundwater guidelines) – this approach is *not* recommended unless defensible rationale is provided. If no screening guidelines exist for a substance, the risk assessor should question why that is the case. Often, environmental quality guidelines may not have been implemented due to high uncertainty in the available scientific data. Consideration of substances should be extended beyond conventional chemical stressors to macronutrients (e.g., phosphorus), dissolved oxygen or other important indicators of habitat quality and quantity that could be potentially important contributors to total risk at a site. Risk assessors should provide rationale for including substances for which guidelines are lacking, but should also provide rationale for excluding those substances if they are site-related. Often, substances are considered in ERA that were not considered during site investigation.

> ***Key Concept:***
>
> Substances without guidelines often warrant consideration in ERA, regardless of whether they were considered during site investigation or not.

3. *Background concentrations* – In some cases, background concentrations of a substance may exceed generic guidelines. In such cases, it may be appropriate to compare to background concentrations rather than to guidelines. This may include comparison to 'regional' background data, or to more localized data. As an example, metal mines are typically located in areas of naturally elevated metals concentrations. Where background concentrations are potentially elevated, reference conditions should be defined carefully, or gradient-based sampling methods should extend far enough off-site to establish suitable local background concentrations of COCs.

> ***Key Concept:***
>
> Off-site data can provide important context in cases where substances are present at naturally elevated concentrations exceeding guidelines.

4. *Food chain uptake* – If environmental quality guidelines are based on protection of lower level receptors (e.g., invertebrates), it is important to determine whether and how to screen COCs for evaluation of higher level receptors via food chain uptake. For example, Canadian interim sediment quality guidelines (ISQGs; CCME 1999a) are not designed to protect higher trophic levels – for potentially

bioaccumulative substances, Canadian tissue residue guidelines for protection of wildlife consumers of aquatic biota are more appropriate (CCME 1999b), but do not cover all relevant substances. Some substances are known to be bioaccumulative or biomagnifying (or are named as such in policy or regulatory documents), but for other substances the need to consider food chain uptake may depend on site-specific characteristics. Screening for food chain pathways usually focuses on receptors with a definable home range size, so one consideration in screening for those receptors is whether to screen based on individual samples or summary statistics for an area (this issue is discussed in more detail below).

5. *Using statistics* – Most risk assessors conduct a preliminary screen of data using maximum concentrations for a COC in a particular medium. However, if the maximum concentration exceeds a guideline, and the receptor is mobile, consideration should be given to the use of summary statistics (e.g., compare the 95% upper confidence limit of the mean [UCLM] or the 90th percentile of the concentrations to the guideline). Risk assessors may consider using summary statistics on a case-by-case basis, in light of factors such as the number of samples, spacing of samples, seasonality or timing of samples (particularly for water), and the characteristics of receptors. Rationale should be provided for any decisions made. As a default, for immobile receptors (e.g., plants, small invertebrates) the maximum concentration for a COC should be used as a conservative starting point. For mobile receptors exposed to an area characterized by multiple samples, the maximum could be used as a conservative starting point if there are less than 10 samples, but for sample sizes of 10 or more the 95th upper confidence limit of the mean and the 90th percentile should be considered, with accompanying rationale. In the case of soil, practitioners should consult current CCME guidance for characterizing a volume of contaminated material using statistics (CCME 2012), as the concepts can be applied to characterizing an area of contamination to which a given receptor may be exposed. In the case of water, consideration should be given to the temporal nature of the data – if data were collected during two separate events, it may be appropriate to use summary statistics from each event separately. Finally, summary statistics that are based on all data may require consideration of data points where contaminant concentrations are below detection limits. In such

> ***Key Concept:***
>
> For immobile receptors, use of maximum concentrations in soil or other media may be appropriate.
>
> For mobile receptors, assuming exposure only to the worst-case soil or sediment concentration is highly conservative. When sample sizes are reasonable (e.g., 10 or more in the area used by a receptor), use of percentiles or upper confidence limits on mean concentrations is more realistic.

cases, risk assessors should provide clear rationale, including statistical rationale as appropriate, for methods used to deal with those data points (see CCME 2012 for additional guidance).

6. *Sampling depths* – For soil and sediment (and less commonly, water[8]), determination of the applicable depth is not always straight-forward. The depth of 'surface' soils or sediments may be standardized for many sites (e.g., default value for rooting depth of plants in soil based on policy determination), but exceptions can be expected. For example, the relevant surface soil horizon may be deeper where tap roots are present. Alternatively, if a site lacks deep rooting plants (or has a planned future use that excludes them), 'surface' soil depth ranges could be shallower. As another example, some COCs or receptors may be associated only with the humic soil layer and not with the underlying inorganic soil layer. In that case, the depth used for screening may not be a fixed depth, but may vary depending on the thickness of the humic layer.

   As a default, if there is no site-specific information available to define the depth of the surface soil layer on a site-specific basis, all data in the top 1.5 m for soil should be used to screen for surface soil exposure, consistent with the Canada-wide standard for petroleum hydrocarbons in soil (CCME 2008b). For sediment, as a default all available data in the top 1 m should be used to screen for surface sediment exposure. In all cases rationale should be provided and consultation with FCSAP expert support is recommended.

   > ***Key Concept:***
   >
   > Appropriate soil and sediment sample depths for screening COCs may vary by site and by receptor group. As a default all data in the top 1.5 m for soil should be considered relevant for screening surface soil exposure, and all data in the top 1 m for sediment should be considered relevant for screening surface sediment exposure. However, when possible the appropriate depth for screening can be defined on a site-specific basis, taking into account factors such as depth of bioturbation and patterns of deposition or erosion.

   The depth of soils and sediments considered for screening COCs may not be the same depth that is considered during exposure assessment for each receptor group. Consideration of soil and sediment depth during exposure assessment is elaborated in **Section 3**.

   For some ERAs, soil or sediment at depth will be explicitly considered in the ERA if:

[8] Depth of surface water sampling can be important for lakes, for example if there is stratification.

(a) there is a plan or a possibility for that soil or sediment to become exposed (e.g., through removal or erosion); or

(b) there is a possibility that contaminants at depth will be mobilized and reach receptors (e.g., via groundwater transport).

7. *Sampling Density and Coverage* – COC screening for ERA usually begins after site investigation has characterized the horizontal and vertical extent of contamination at a reasonable resolution. However, the density and coverage (horizontal, vertical) of sampling will vary by site, and in some cases ERA is initiated before site assessment activities are complete. Ideally, risk assessors get involved in site investigation planning early so that sampling density is sufficient for screening and characterizing exposure later during ERA. Specifically, sampling density must be sufficient, and the samples must be representative of site conditions relevant to the ERA. If data are limited, risk assessors should consider whether sufficient data are available to warrant exclusion of certain COCs based on lack of exceedances of guidelines for existing samples. Practitioners should consult current CCME sampling guidance when considering sampling density and coverage (CCME 2012).

8. *Data Quality* – Another important consideration when evaluating existing data is the quality of those data, particularly with respect to analytical detection limits. If many or all existing data do not have detection limits that are lower than relevant guidelines, then the utility of the data for ruling out COCs is diminished. Practitioners should ensure that sampling and analysis plans specify data quality objectives that meet the needs of the ERA and are consistent with current CCME sampling guidance regarding data quality (CCME 2012; also see EPA 2006 for further discussion). In some instances, higher resolution methods or specific extraction procedures may be required to achieve the accuracy and precision requirements of the ERA.

9. *Form of contaminants* – Risk assessors should be diligent in identifying the relevant form of contaminants and how contaminants are identified. The exposure pathways identified in the ERA will in part determine the relevant form(s) of the contaminant (e.g., total versus dissolved, oxidation state, or adjustment for environmental conditions such as pH). The type of contamination and the availability of toxicological data may determine the form of a contaminant considered in the ERA. For example, in the case of PCBs, it may be possible to conduct the ERA based on total PCBs, on one or more individual congeners, selected homologs, or on Aroclor mixtures. Alternatively, or in addition, the dioxin-like PCB congeners may be evaluated using the toxic equivalency (TEQ) model whereby the combined effects of all dioxin-like compounds are evaluated together. The appropriateness of each approach depends on the receptor type, the

chemical signature present at the site, and the availability of matched effects data for each quantitation method.

2.2.4.3. Characteristics of COCs

The characteristics of COCs are an important input to identifying receptors, exposure pathways and endpoints in the problem formulation. Characteristics can be separated into two types – (1) transport and fate (including bioavailability), and (2) effects.

> ***Key Concept:***
>
> An understanding of the characteristics of a COC is important for determining how a COC may pose risks, and which receptors are most likely to be affected.

Transport and Fate – The transport and fate characteristics of a COC determine how the contaminant will move from source(s) and partition into various environmental media such as soil, water, sediment and biota. The transport and fate characteristics help determine which receptors and exposure pathways are relevant in the ERA. For example, sediment benthic organisms may be relevant for contaminants that are transported from an upland site to the aquatic environment via groundwater. The transport and fate description is usually qualitative, but when possible quantitative metrics should also be used. For example, for organic compounds the octanol-water partition coefficient (K_{OW}) provides insight into potential for bioaccumulation and biomagnification (i.e., substances with a high K_{OW} tend to partition into organic matter). The CCME national classification system for contaminated sites (CCME 2008a) uses a threshold $\log(K_{OW})$ of 4, above which exposure via food chain transfer are considered more likely. The Persistence and Bioaccumulation Regulations under CEPA (SOR/2000-107) use a $\log(K_{OW})$ of 5 or higher as one method of classifying a contaminant as bioaccumulative (other methods rely on magnitude of bioaccumulation factors or bioconcentration factors). Whenever pathways are excluded from consideration on the basis of the transport and fate properties of a COC, rationale is essential.

Bioavailability is an important factor influencing the degree to which COCs will partition from abiotic media into tissues. A COC that is bound with soil particles may pass through the gut of a receptor whereas a COC in dissolved form in water may be much more bioavailable in the gut.

Consideration of the transport and fate characteristics should include potential degradation processes that are relevant to the substance. Some contaminants degrade into breakdown products that may be as or more toxic than the parent compounds. In some cases (e.g., PAH contamination in aquatic life), the metabolism of the parent product is highly-receptor specific.

The transport and fate of COCs will depend on their physical and chemical properties, and on the specific characteristics of the environmental media at the site. For this reason, conventional parameters collected during chemistry programs are important – examples include soil pH, water hardness, organic content of soil or sediment, or sediment grain size. These parameters affect the fate of COCs not only among abiotic media but also between abiotic and biotic media (e.g., by influencing bioavailability).

Effects – The review of effects characteristics of a COC emphasizes the types of organisms that may be affected by the COC and the relevant mechanisms of action. It is seldom necessary for the information to be compiled for every specific receptor. Rather, broad characteristics relevant to key receptor groups are usually adequate for the purposes of the problem formulation. In some cases, the concentrations or doses associated with particular adverse effects may be specified, helping to identify the effects endpoints that are expected to be most sensitive and therefore the potential candidates for the formal effects assessment. The effects characteristics provide important information for selection of receptors (e.g., which receptors groups are known to be sensitive to the COC) and for the selection of endpoints (i.e., those known to be caused by the COC and that are relevant to the ERA). While it is important to understand the basic environmental fate and toxicity of COCs at the problem formulation stage, more comprehensive reviews of effects literature are typically conducted as part of the effects assessment during the ERA (e.g., if needed for derivation of a dose-response relationship).

Sites can vary greatly in their degree and nature of contamination. As the nature of contamination changes from single COCs to several COCs to complex COC mixtures, so do the challenges of understanding potential effects. When multiple contaminants are present in an exposure medium, they may interact to produce antagonistic, additive or synergistic effects. Ultimately, not accounting for these interactions, or applying invalid models to account for such interactions, could both lead to erroneous risk conclusions. Some tools used in effects assessment are better suited than others to dealing with contaminant mixtures (see **Section 4**). Some contaminant interactions have been well characterized. Examples include the biotic ligand model (BLM; Di Toro *et al.* 2001, Paquin *et al.* 2003) for metals, the ΣPAH model (e.g., Swartz *et al.* 1995, Ozretich *et al.* 2000) for polycyclic aromatic hydrocarbons, and toxic equivalent (TEQ) approaches for dioxin-like effects (e.g., CCME 2002).

Understanding contaminant interactions in detail at the problem formulation stage is not usually warranted, and may not be warranted at all (i.e., even during effects assessment) depending on the scope of the ERA and the tools used. At a minimum, risk assessors should attempt to identify potentially important interactions when documenting the modes of action of COCs during problem formulation (e.g., Menzie *et al.* 2009). The risk assessor can then determine the best approach to integrating that information into the ERA.

2.2.5. Receptors of Concern

> ***Key Concept:***
>
> For wildlife receptors, surrogate ROCs can be used in the ERA to represent risks to a group of receptors with common characteristics such as feeding habits (e.g., small omnivorous mammals, piscivorous birds, etc.).

This section contains the technical guidance for receptor selection for use in ERA. Specifically, guidance is provided for determining which types of ROCs should be considered at a site, and for identifying appropriate surrogates as representatives of those receptor types.

As previously defined, for ERA a receptor of concern (ROC)[9] is any non-human individual, species, population, community, habitat or ecosystem that is potentially exposed to COCs. The level of biological organization at which an ROC is defined varies. In the case of lower trophic levels, the community is often identified as the ROC (e.g., zooplankton community, benthic invertebrate community). In the case of higher trophic levels, the ROC is usually defined at the species level (e.g., mink, eagle). In the latter case, a particular species may be selected for direct assessment of that species and/or for use as a representative (or surrogate) for similar organisms. As described in this section, where a surrogate organism is applied, the risk assessor should articulate the groups of organisms that the ROC is intended to represent. In most cases, the groups are selected on the basis of functional feeding groups (e.g., small omnivorous mammals, piscivorous birds, forage fish) rather than on taxonomic linkages. In selecting a specific surrogate ROC, the risk assessor considers the degree to which the surrogate may be assumed to be protective of related species on the basis of contaminant sensitivity and life history considerations (diet, foraging range, etc.). This section provides guidance on these issues.

The following subsections outline the following:

- Information compilation;
- Identification of receptor types;
- Criteria for selection of surrogate (representative) ROCs; and
- Linking ROCs to problem formulation.

2.2.5.1. Information Compilation

Consideration of potential receptors is site-specific and begins with an understanding of the ecological attributes of the site. The risk assessor should start by compiling information such as:

[9] The term Valued Ecosystem Component (VEC) has the same or similar meaning.

- General site characteristics (e.g., forest cover, roads, watershed, wetland areas);
- Regional and local habitat surveys and land use classifications;
- Records of environmental conditions / parameters measured on site that may be relevant to any level of biological organization;
- Species inventories (flora and fauna) and species range maps;
- Species that are at risk (e.g., listed as rare or endangered), or have some similar status (e.g., consult the *Species at Risk Act* (*SARA*) and provincial lists). Identifying the possibility of species at risk at this early stage provides an opportunity for specific consideration of these species in the ERA;
- Other jurisdictional lists of suggested or required ROCs (e.g., "Paramètres d'exposition chez les mammifères" and "Paramètres d'exposition chez les oiseaux" in the province of Quebec[10]; or "Rationale for the development of generic soil and groundwater standards for use at contaminated sites in Ontario" in the province of Ontario[11]);
- Information from local experts and residents of the area or surrounding properties; and,
- Potential presence of domestic animals (livestock, cats, dogs) that may warrant a particular level of protection (e.g., protection of individual organisms) or consideration of particular endpoints not usually considered for wildlife (e.g., cancer).

> ***Key Concept:***
>
> Habitat surveys by wildlife biologists can help risk assessors to identify relevant ROCs.

If site information is limited or simply not available at this point, practitioners should consider conducting a site visit with a qualified professional to obtain site information (e.g., basic site characteristics, habitat types represented, and receptors common to the site). Even if information is available, a site visit is effective in confirming existing information and providing a better basis for identifying receptors. Evaluation of the use of a site should take into account seasonality as some potential receptors may only use the site for a

[10] Document available from: http://www.ceaeq.gouv.qc.ca/index_en.htm

[11] Document available from: http://www.ene.gov.on.ca/stdprodconsume/groups/lr/@ene/@resources/documents/resource/stdprod_081546.pdf

portion of their life cycle. Methods for site-specific surveys for purposes of receptor identification are usually qualitative, but may also be quantitative[12].

2.2.5.2. Identification of Receptor Types

There are numerous types of receptors relevant to ERA and numerous surrogate ROCs that can be used to represent those receptor types (see **Tables 2-1 and 2-2** for aquatic and terrestrial ecosystems respectively). Surrogate ROCs are often particular species for higher trophic levels (e.g., birds and mammals) and are often communities for lower trophic levels.

> ***Key Concept:***
>
> Rationale should be provided to support inclusion and exclusion of receptor types in an ERA. Tables 2-3 and 2-4 are recommended for this purpose.

During problem formulation, the practitioner must initially consider all receptor types that could be included in the ERA and should then provide rationale for why particular receptor types are included in or excluded from the ERA (**Tables 2-3 and 2-4** are recommended templates for this purpose, for aquatic and terrestrial ecosystems respectively). Once receptor types are selected, surrogate ROCs should be selected for each receptor type.

ROC selection should be based on all of the information compiled about the site, and should consider:

> ***Definition:***
>
> A feeding guild refers to a group of organisms that use the same ecological resource in a similar way for feeding (e.g., insectivores, granivores, detritivores, carnivores, etc.).

- Representation from the various trophic levels, habitats, feeding guilds and environments that are appropriate for the site.
- Receptors that could be found off-site in adjacent properties, but that use the subject site or could be affected by on-site contamination.
- Receptors that are expected to be present during particular times or seasons.
- Receptors that are expected to be present under future scenarios / land use, if relevant for the ERA.

[12] For example, the US Fish and Wildlife Service has developed Habitat Suitability Index Models for fish and wildlife (http://www.nwrc.usgs.gov/wdb/pub/hsi/hsiintro.htm)

Table 2-1. Types of receptors and example surrogates for aquatic ecosystems

Aquatic Receptor Group	Aquatic Receptor Type[1]	Example Surrogate[2,3] ROCs for Aquatic Ecosystems	
		MARINE	**FRESHWATER**
		e.g., ocean, rocky shore, tidal flat, estuary	e.g., lake, pond, wetland, river, stream
Primary Producer	Phytoplankton	phytoplankton community	phytoplankton community
	Periphyton		periphyton community
	Plants and Algae	seaweed species; plant or algae community	algal species; aquatic plant community
Pelagic Invertebrate	Zooplankton	zooplankton community	zooplankton community
	Others	shrimp; jellyfish	shrimp
Benthic Invertebrate	Epifauna	mussel; crab	crayfish; benthos community
	Infauna	benthos community	bivalve; benthos community
Fish	Benthivorous	stickleback; sculpin; herring; flatfish	stickleback; sculpin; sucker
	Planktivorous	minnow	salmonid (e.g., kokanee)
	Piscivorous	salmonid	salmonid
Mammal	Herbivorous		muskrat*; beaver; moose
	Piscivorous	seal; otter	mink*; otter*
	Omnivorous	racoon*; bear*	racoon*; bear*
Bird	Herbivorous	goose*; brant	goose*
	Insectivorous	shorebird; diving duck	shorebird; swallow
	Piscivorous	grebe; cormorant; heron*; eagle; kingfisher*	grebe; loon; merganser; heron*; osprey; eagle; kingfisher*
	Omnivorous	dabbling duck	dabbling duck*; diving duck*
Amphibian	Carnivorous		frog; toad; salamander
Reptile	Omnivorous		turtle

Notes:

[1] Receptor types in lower trophic levels are classified by habitat, whereas those in higher levels are classified by feeding guild.

[2] Examples of surrogate ROCs that are commonly used to represent the receptor types; note that more than one surrogate ROC can be selected for a given receptor type.

[3] Surrogates are not always needed, particularly for lower trophic levels where the ROC is often defined at the community level. In this table, lower trophic level communities are listed to clarify what is typically evaluated, but the communities are not surrogates; rather, they are the receptors of interest.

* Receptors that are recommended in the province of Quebec. Refer to the documents below for comprehensive species-specific lists.

For mammal species refer to **Tableau 2** in "Paramètres d'exposition chez les mammifères", available from: http://www.ceaeq.gouv.qc.ca/index_en.htm

For bird species refer to **Tableau 2** in "Paramètres d'exposition chez les oiseaux", available from: http://www.ceaeq.gouv.qc.ca/index_en.htm

Table 2-2. Types of receptors and example surrogates for *terrestrial* ecosystems

Terrestrial Receptor Group	Terrestrial Receptor Type[1]	Example Surrogate[2,3] ROCs for Terrestrial Ecosystems The examples below may apply to urban landuses (e.g., industrial, commercial, agricultural, park, residential) or to wildlands (e.g., prairie, forest, tundra, alpine).
Primary Producer	Moss/Grass/Shrub/Tree/Forb	plant species; plant community
Invertebrate	Ground-dwelling	invertebrate community; particular species (earthworm; springtail; beetle)
	Aerial	dragonfly
Mammal	Herbivorous	vole*; mouse*; squirrel*; hare; cattle; sheep; deer*; caribou
	Insectivorous	shrew*; mole*; bat
	Carnivorous	marten; weasel*; domestic cat; domestic dog; coyote*; bobcat
	Omnivorous	fox*; skunk*; raccoon*; bear*
Bird	Herbivorous	canada geese
	Insectivorous	warbler; flycatcher; swallow
	Carnivorous	owl; hawk*; falcon
	Omnivorous	blackbird; sparrow*; crow*; grouse*; chickadee*; robin*
Amphibian	Carnivorous	frog; toad; salamander
Reptile	Carnivorous	snake; lizard

Notes:

[1] Receptor types in lower trophic levels are classified by habitat, whereas those in higher levels are classified by feeding guild.

[2] Examples of surrogate ROCs that are commonly used to represent the receptor types; note that more than one surrogate ROC can be selected for a given receptor type.

[3] Surrogates are not always needed, particularly for lower trophic levels where the ROC is often defined at the community level. In this table, lower trophic level communities are listed to clarify what is typically evaluated, but the communities are not surrogates; rather, they are the receptors of interest.

* Receptors that are recommended in the province of Quebec. Refer to the documents below for comprehensive species-specific lists.

For mammal species refer to **Tableau 2** in "Paramètres d'exposition chez les mammifères", available from: http://www.ceaeq.gouv.qc.ca/index_en.htm

For bird species refer to **Tableau 2** in "Paramètres d'exposition chez les oiseaux", available from: http://www.ceaeq.gouv.qc.ca/index_en.htm

See OMOE (2011) for a list of birds and mammals that were used to develop generic site condition standards in Ontario.

Table 2-3. Template for ROC selection and rationale in *aquatic* ecosystems

Aquatic Receptor Group	Aquatic Receptor Type[1]	Included in ERA? (Yes/No)	Rationale[2]	Surrogate ROC[3] (if applicable)
Primary Producer	Phytoplankton			
	Periphyton			
	Macrophyte			
Pelagic Invertebrate	Zooplankton			
	Others			
Benthic Invertebrate	Epifauna			
	Infauna			
Fish	Benthivorous			
	Planktivorous			
	Piscivorous			
Mammal	Herbivorous			
	Piscivorous			
	Omnivorous			
Bird	Herbivorous			
	Insectivorous			
	Piscivorous			
	Omnivorous			
Amphibian	Carnivorous			
Reptile	Omnivorous			

Notes:

[1] Each receptor type should be represented in an ERA if relevant to the site.

[2] A rationale must be provided whether the receptor type is being represented or not.

[3] A surrogate ROC is used to represent a receptor type. Surrogates are usually identified for fish and wildlife, but less often for lower trophic levels where the ROC is often defined at the community level; note that more than one surrogate ROC can be selected for a given receptor type.

Table 2-4. Template for ROC selection and rationale in *terrestrial* ecosystems

Terrestrial Receptor Group	Terrestrial Receptor Type[1]	Included in ERA? (Yes/No)	Rationale[2]	Surrogate ROC[3] (if applicable)
Primary Producer	Moss/Grass/Shrub/Tree/Forb			
Invertebrate	Ground-dwelling			
	Aerial			
Mammal	Herbivorous			
	Insectivorous			
	Carnivorous			
	Omnivorous			
Bird	Herbivorous			
	Insectivorous			
	Carnivorous			
	Omnivorous			
Amphibian	Carnivorous			
Reptile	Carnivorous			

Notes:

[1] Each receptor type should be represented in an ERA if relevant to the site.

[2] A rationale must be provided whether the receptor type is being represented or not.

[3] A surrogate ROC is used to represent a receptor type. Surrogates are usually identified for wildlife, but less often for lower trophic levels where the ROC is often defined at the community level; note that more than one surrogate ROC can be selected for a given receptor type.

2.2.5.3. Criteria for Selection of Surrogate ROCs

Based on the information review, there may be numerous possible surrogate ROCs that could be selected for each receptor type. It is often appropriate to include multiple surrogates for each receptor type due to variability among species. However, because assessing the ecological risks to an exhaustive list of potential ROCs is generally neither practical nor necessary, the following criteria should be used to select the appropriate types of ROCs and representative surrogates:

1. *Ecological relevance* – An ecologically 'relevant' organism is one that is an appropriate indicator of actual or potential exposures given the environmental conditions germane to the assessment. An ecologically relevant organism should be expected to be found at a site under reasonably foreseeable conditions (e.g., an arctic fox at a site in the arctic), whereas an ecologically irrelevant organism is one that would not be expected to be found at a site under normal circumstances (e.g., a wolf at a small urban site). An important distinction to be made is that an organism need not be actually observed to be considered ecologically relevant. If contamination is sufficiently great that the organism is extirpated, or if an organism is sufficiently secretive, there may be little or no evidence of its presence. A common error made in problem formulation is to assume that local absence of an organism equates with lack of ecological relevance. It is usual practice to select ROCs that represent key functional groups that are expected to be exposed to the COCs on site, or that would be expected to be present at the site in the absence of contamination. In addition, keystone species that are important to ecosystem stability may be preferentially selected as ROCs.

2. *Degree / mechanism of exposure to the COCs on site* – A number of factors have the potential to affect the degree to which ROCs are exposed to the COCs on the site, including:

- The status of the ROC (life stage, migratory versus resident);
- How the ROC uses the site (feeding guild, feeding behaviour, metabolism);
- How much / often the ROC uses the site (home range size, habitat suitability, off-site habitat characteristics); and
- Number and type of exposure pathways (environmental media, indirect / direct contact / consumption, bioaccumulation and biomagnification processes).

It is therefore important to consult ROC life history and background information, consider the intended use of a site in terms of its influence on habitat quality and availability, and understand which exposure pathways are relevant.

Individual receptors may be exposed to COCs through a number of pathways, all of which should be identified during problem formulation. For receptor selection, information on the relative importance of these exposure pathways is critical. For

example, if groundwater flow to aquatic life is an important fate pathway, this may indicate that intertidal receptors (e.g., benthos, mussels, kelp) would be more appropriate as ROCs than finfish. Exposure pathways are considered "open," "operable," or "complete" if a COC is present and there is a route of exposure by which a receptor of concern comes into contact with the COC. A common error in risk assessment is to confuse the distinction between (a) a pathway that is operable but with low exposure concentrations, and (b) a pathway that is inoperable due to lack of plausible transport pathway.

3. *Relative sensitivity to the COCs* – It is customary to include species or other receptor types that are relatively sensitive to the COCs if such information is known. For example, some birds are known to be sensitive to certain pesticides due to effects on egg shell thinning; some fish are known to be sensitive to selenium based on reproductive toxicity endpoints; mink are known to be sensitive to PCBs and mercury. The principle for selection of a sensitive species is that demonstration of lack of harm for a sensitive organism is considered an indication of protection for the less sensitive taxa in the same functional group. However, selection of ROCs based solely on sensitivity considerations is questionable. Sensitivity must be considered in terms of both the magnitude at which responses are observed and the type of response elicited. In addition, sensitivity may occur for particular life stages, therefore the mechanisms by which site-related COCs could affect that life stage must be considered.

4. *Relative importance from a conservation perspective* – If rare, endangered or threatened species (i.e., species at risk) or habitats are confirmed to be present, these species must be considered as potential ROCs. They should also be included if they are likely to be present in the future (based on information regarding geographic distribution, habitat preferences and site-specific habitat availability).

5. *Relative social, economic and/or cultural importance* – Any particular species or group that is of special importance (e.g., domestic pets, livestock, species of significance to Aboriginal Communities, species of commercial or recreational importance) would typically be included as an ROC and may be subjected to a different level of protection than other ROCs.

6. *Availability of ecotoxicological and life history data* – Where effects data will be literature-based, ROCs for which ecotoxicological data are readily available are preferentially selected (see **Section 4** regarding sources); otherwise the ability of ERA to assess effects on the ROC may be reduced. The benefit of selecting highly-specific ROCs is offset where data related to toxicity thresholds or exposure information is limited.

7. *Availability of appropriate measurement endpoints* – It is important to assess ROCs at an ecological scale that is relevant to management goals for the site and to select measurement endpoints that are aligned with those goals. An example of an ecosystem-level receptor would be "the wetland ecosystem," for instances where the measure of

effect reflects an ecosystem-level process such as nutrient cycling or primary productivity. An example of a community-level receptor would be "the benthic invertebrate community," for instances where the measure of effect is community-level attributes such as species diversity. An additional consideration related to measurement endpoints is the ability to distinguish effects from natural variation. For example, abundance of benthic organisms is often highly variable, particularly where habitat and substrate conditions vary; in these circumstances, the investigator must consider the statistical and practical (financial) constraints to detection of site-related impacts.

In addition to the criteria described above, rationale for selection of surrogate ROCs may be based on logistical considerations or other tools such as a site visit and habitat assessment by a qualified biologist. Local expertise and traditional knowledge may also be useful in identifying appropriate ROCs.

2.2.5.4. Carrying ROCs Forward in the Problem Formulation

A list of ROCs must be carried forward and linked to other elements of problem formulation. This occurs in at least two ways, as explained in the subsequent sections of this guidance. First, ROCs are a component of a conceptual site model, and are linked to sources of COCs via exposure pathways. Second, specific attributes of ROCs are identified to formulate the assessment endpoints for the ERA.

A tabular format can be useful in summarizing the ROCs that are carried forward. As recommended earlier, templates are useful (see **Tables 2-3 and 2-4** for aquatic and terrestrial ecosystems respectively) to guide the risk assessor during selection of surrogates that are used to represent some types of ROCs. Rationale should be provided even for receptor types that are not carried forward in the ERA.

> ***Key Concept:***
>
> Rationale should be provided to support inclusion and exclusion of receptor types in an ERA. Templates provided in Tables 2-3 and 2-4 are recommended for this purpose.

In some cases the ROC selection process may not be finalized during the problem formulation, but rather presented in conceptual form pending the results of more detailed study. For example, for remote sites it may be more efficient for the problem formulation to focus on selection of broad receptor groups and types only, deferring selection of specific surrogates until a wildlife biologist visits the site during an ERA field program. This approach may be particularly appropriate for higher trophic level receptors.

2.2.6. Exposure Pathways

This section provides guidance on identifying exposure pathways linking contaminant sources to ROCs. The identification of exposure pathways is highly inter-related to other elements of the problem formulation. The identification of pathways integrates information on:

- Sources of COCs;
- Contaminant fate and transport; and
- ROCs and their general characteristics.

These elements have been discussed in previous sections.

Rationale should be provided for the inclusion or exclusion of any potential pathways for each receptor group. The rationale may be based on quantitative considerations (e.g., magnitude of concentrations in groundwater and expected dilution before contact with the receiving environment), qualitative considerations (putative limitation of inhalation exposures to surface-dwelling wildlife), or a combination of these approaches. Rationale should indicate where pathways are considered to be:

- Complete (or operative / open), with a documented link between source and receptor, or
- Incomplete (or inoperative / closed), with no documented or anticipated link between source and receptor.

The following exposure pathways should be considered (adapted from SAB-CS 2008), although specific requirements may vary by jurisdiction[13]:

- Soil invertebrates and terrestrial plants are in direct contact with elevated COC concentrations in soil.
- Mammals, birds, amphibians, reptiles, fish and invertebrate macrofauna ingest elevated COC concentrations via soil / sediment ingestion (e.g., via consumption of soil covered plant roots).
- Mammals, birds, amphibians and reptiles ingest elevated COC concentrations via water ingestion.
- Mammals, birds, amphibians

> ***Key Concept:***
>
> For wildlife, ingestion pathways (water, food and soil/sediment) may often be the only relevant pathways. However, dermal exposure and inhalation are relevant in some cases.

[13] For example, policy in B.C. allows the ERA practitioner to exclude dermal exposure and inhalation for birds and mammals except for rare cases (SAB-CS 2008).

and reptiles ingest elevated COC concentrations via consumption of prey items (particularly for those chemicals known to bioaccumulate).

- Aquatic species (macrophytes, plankton, invertebrates, amphibians and fish) are in direct contact with elevated COC concentrations in surface water, sediment or sediment porewater.
- Some aquatic species (e.g., planktivores, piscivores) ingest elevated COC concentrations via consumption of prey items.
- Dermal exposure (direct contact with soil and sediment) of wildlife should be considered when relevant, for COCs that can be absorbed readily through this pathway. Dermal exposure can also be a relevant exposure pathway for amphibians and reptiles. Detailed guidance on how to assess dermal exposure is limited (SAB-CS 2008, Suter 1996). Approaches for this pathway should be taken on a site-specific basis with appropriate rationale and consultation.
- Inhalation exposure through wind-blown dust or inhalation of vapours can be a relevant pathway for some mammals, birds, reptiles and amphibians. In practice, this pathway has not been commonly assessed, but may be required in some jurisdictions in future, and should be considered where the conceptual model indicates potential widespread exposure. For example, a site with high concentrations of volatile compounds and good small mammal habitat may warrant consideration of vapour inhalation. Inhalation toxicity data are currently lacking for most contaminants, but some jurisdictions are developing guidance and screening values for soil and vapour. In addition, because small mammals generally construct their burrows to allow for air flow, characterizing exposure may be challenging.
- Indirect pathways such as food source depletion by toxicity of COCs to invertebrates should also be considered.

Tables are recommended for summarizing the pathway selection process – example templates are provided for aquatic ecosystems (**Table 2-5)** and for terrestrial ecosystems (**Table 2-6**).

Key Concept:

Rationale should be provided to support inclusion and exclusion of exposure pathways for each receptor group in an ERA. Tables 2-5 and 2-6 are recommended for this purpose.

Table 2-5. Example of tabular format for justifying exposure pathway selection in *aquatic* ecosystems

Receptor Group	Exposure Pathway	Included (yes/no)	Rationale
Primary Producer	Direct Contact (Water)		
	Direct Contact (Sediment)		
Pelagic Invertebrate	Direct Contact (Water)		
Benthic Invertebrate	Direct Contact (Water)		
	Direct Contact (Sediment)		
	Food Consumption (for macrofauna)		
Fish	Direct Contact (Water)		
	Direct Contact (Sediment)		
	Food Consumption		
	Incidental Sediment Ingestion		
Mammal	Water Consumption		
	Food Consumption		
	Incidental Sediment Ingestion		
Bird	Water Consumption		
	Food Consumption		
	Incidental Sediment Ingestion		
Amphibians & Reptiles	Direct Contact (Water)		
	Water Consumption		
	Food Consumption		
	Incidental Sediment Ingestion		

[1] This table should be adapted on a site-specific basis and in many cases should have additional detail for receptor types (e.g., benthic infauna, benthic epifauna, etc.) or additional pathways that may be relevant for particular contaminants (e.g., maternal transfer via eggs or lactation).

Table 2-6. Example of tabular format for justifying exposure pathway selection in *terrestrial* ecosystems

Receptor Group	Exposure Pathway	Included (yes/no)	Rationale
Primary Producer	Direct Contact (Soil, Soil Porewater or Groundwater)[2]		
Invertebrate	Direct Contact (Soil, Soil Porewater or Groundwater)[2]		
Mammal	Water Consumption		
	Food Consumption		
	Incidental Soil Ingestion		
	Dermal Exposure		
	Inhalation		
Bird	Water Consumption		
	Food Consumption		
	Incidental Soil Ingestion		
	Dermal Exposure		
	Inhalation		
Reptiles & Amphibians	Water Consumption		
	Food Consumption		
	Incidental Soil Ingestion		
	Dermal Exposure		
	Inhalation		

[1] This table should be adapted on a site-specific basis and in many cases should have additional detail for receptor types (e.g., benthic infauna, benthic epifauna, etc.) or additional pathways that may be relevant for particular contaminants (e.g., maternal transfer via eggs or lactation).

[2] For ERA purposes, this guidance defines any water in soil interstitial spaces in the biologically active zone as soil porewater. In other words, groundwater may be a source of contaminants, but in the biologically active zone of soil that water is considered to be porewater.

2.2.7. Conceptual Site Model

A conceptual site model (CSM)[14] guides implementation of the ERA, by clarifying the relationships between:

> ***Key Concept:***
>
> A conceptual site model brings together in one place the key information regarding contaminant sources, fate and transport pathways, exposure pathways and receptors of concern.

- Contaminant sources;
- Relevant fate and transport pathways;
- Receptors of concern; and
- Relevant exposure pathways linking sources to ROCs.

The CSM is a core component of most ERA frameworks (e.g., ASTM 2008, CCME 1996, SAB-CS 2008, Suter 1996, USEPA 1998). The CSM can be expressed in a table, matrix, diagram, or pictorial format. Importantly, the CSM should be supported with text that cross-references the rationale used to select ROCs and exposure pathways (e.g., the rationale detailed in **Tables 2-3 to 2-6**).

Because risk assessment is an iterative process, a CSM should be updated as more information becomes available to refine the problem formulation.

The overall complexity of a CSM should be proportional to the complexity of the site. A simplified food web diagram (showing significant interactions between the different trophic levels and feeding guilds) is often a useful component of a CSM for identifying links between COCs and ROCs at all trophic levels. For example, a CSM with a food web diagram may indicate that elevated COC concentrations in soil may impact both soil invertebrates as well as insectivorous small mammals.

Two main types of CSMs are pictorial and box-diagram, each with certain advantages and disadvantages as follows:

> ***Key Concept:***
>
> The preferred type of conceptual site model depends on the details of the site. For complex ERAs, it may be appropriate to use more than one type in order to convey all relevant information.

Pictorial: A graphical CSM that incorporates visual representations of the pathways and receptors. Pictorial CSMs should typically contain arrows and descriptive text to summarize linkages between

[14] Guidance developed for environmental site characterization (CCME 2011) distinguishes between a CSM and a Conceptual Exposure Model (CEM). ERA practitioners typically do not use the term CEM. The definition of a CSM used here is consistent with ERA practice and existing ERA guidance (CCME 1996; USEPA 1998; SAB-CS 2008).

sources, pathways and receptors. This style of CSM is well suited to communicating contaminant sources, exposure pathways, major fate processes, and food web dynamics to a non-technical audience. A disadvantage is that some fate processes and indirect effects, as well as information on complete / incomplete pathways, cannot be represented easily in a pictorial fashion. This disadvantage can be mitigated by augmenting the pictorial CSM with a tabular summary of exposure pathways, indicating where pathways are complete and significant for each receptor group considered in the ERA (an example of such a tabular summary is part of **Figure 2-4** discussed below). Examples of pictorial-style CSMs are provided in **Figures 2-2 and 2-3.**

Box Diagram: This type of CSM uses a "flowchart" style. An advantage of this approach is that it facilitates a more rigorous examination of the pathways and connections among and between contaminant sources, fate and exposure pathways, and receptors. This type of model may or may not incorporate a tabular summary indicating where pathways are complete and significant. The main disadvantage of this CSM form is that information is more difficult to interpret, especially for a lay audience. An example of this type of CSM is provided in **Figure 2-4**.

For particularly complex sites, the use of both types of CSMs should be considered (rather than just one) as each type of CSM has unique advantages.

Importantly, while **Figures 2-2 to 2-4** are typical examples of basic CSMs, additional information can be added to the CSMs, particularly for complex sites– for example, CSMs can be annotated with information about the COCs associated with each pathway, their chemical form in various media, or the types of effects that are considered for each receptor of concern. **Figure 2-5** provides a simple example of a CSM that is specific about the COCs, the receptor, and food chain linkages, and also gives an indication of how the exposure and effects assessments are conducted. Finally, a CSM may also be used to show indirect or secondary effects, for example effects on food supply for piscivorous birds associated with a contaminant-related decline in fish population density.

Software choices for creation of conceptual models vary, but generally include spreadsheet packages (e.g., such as Microsoft Excel®), presentation packages (e.g., Microsoft PowerPoint®), or graphic packages such as Corel Draw® and Microsoft Visio®[15]. Ultimately, the software which is used will be determined by the presentation format and ease of use. Typically, box diagrams are easily constructed using spreadsheets or presentation packages, whereas pictorial diagrams usually require graphics packages.

[15] This list is not comprehensive, but includes some of the more commonly used software packages.

Figure 2-2. Example of Pictorial-Style Conceptual Site Model

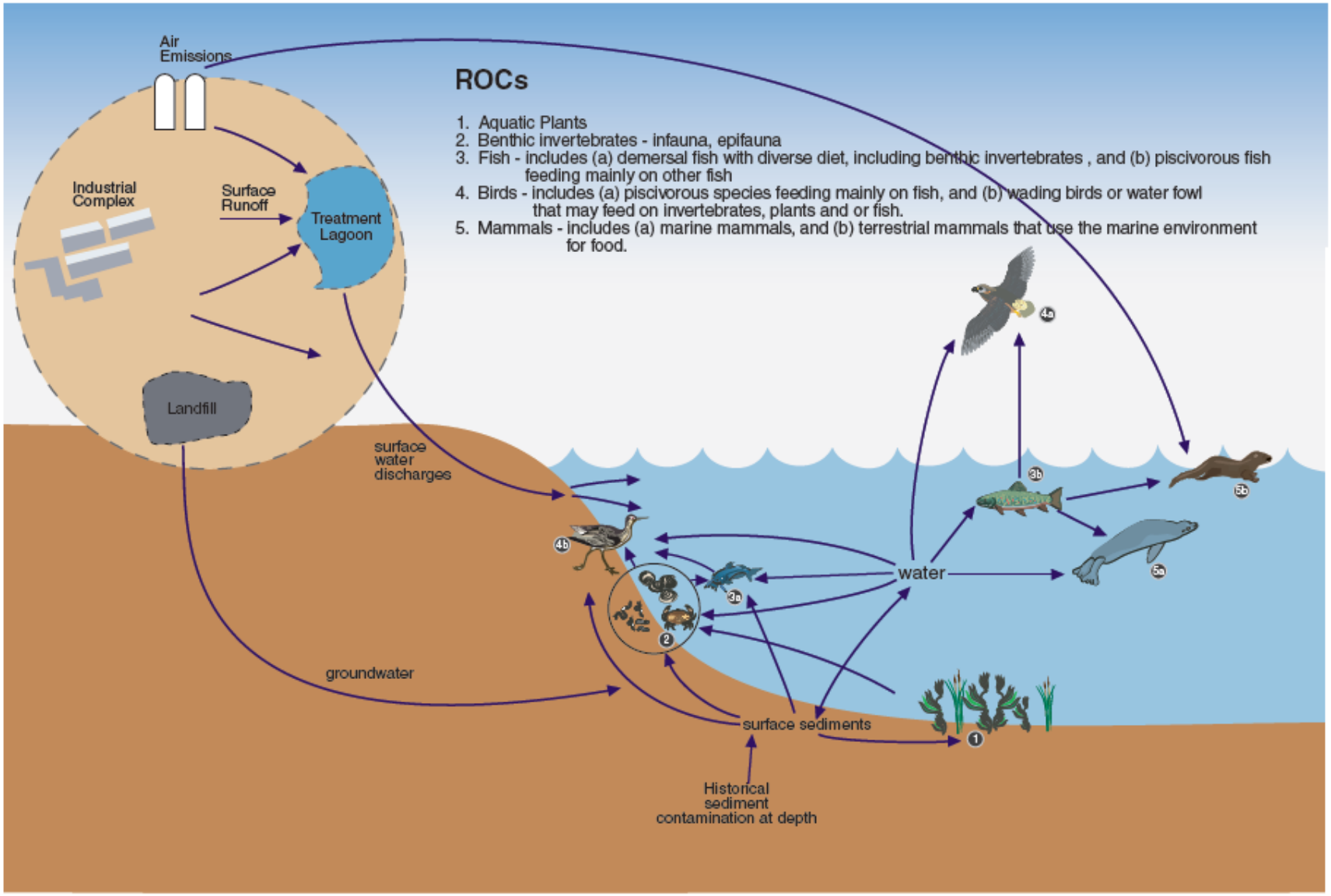


Figure 2-3. Example of Pictorial-Style Conceptual Site Model

1. Accumulation of COCs by soil invertebrates (ingestion, direct contact) and plants (root uptake).
2. Consumption of plants and soil invertebrates by small mammals and birds.
3. Consumption of small mammals and birds by carnivores.
4. Movement and accumulation of COCs from soil to hard-bottom benthic organisms via groundwater and surface water runoff.
5. Movement and accumulation of COCs from soil to soft-bottom benthic organisms via groundwater and surface water runoff.
6. Uptake from COCs by fish from water and food.

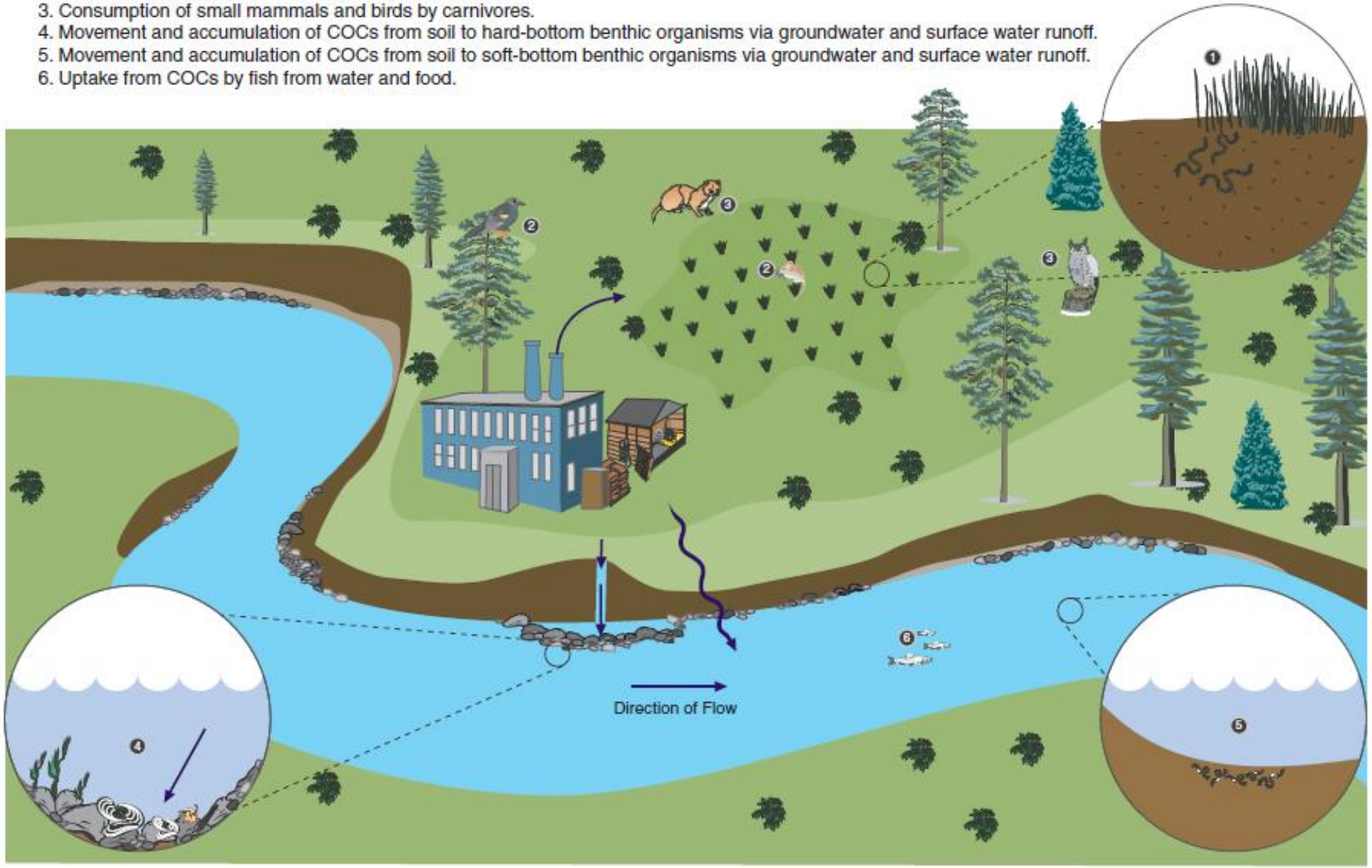
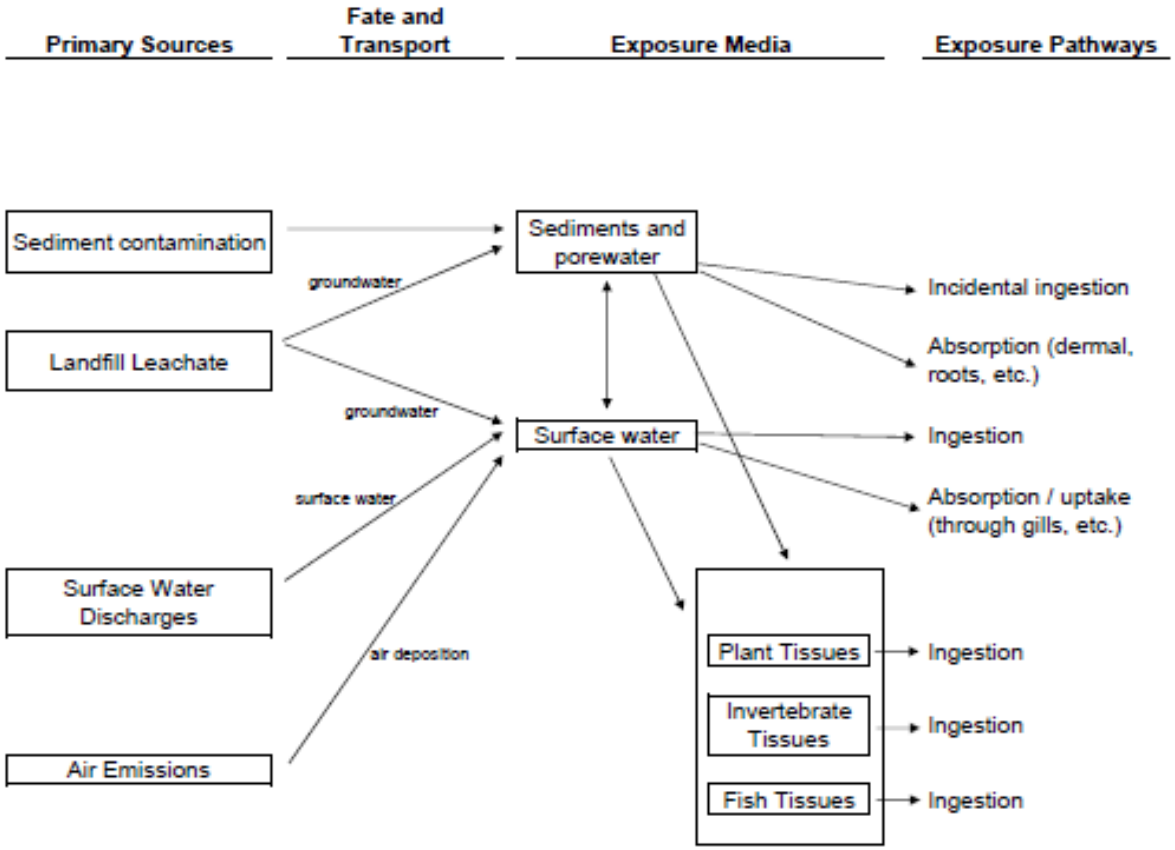


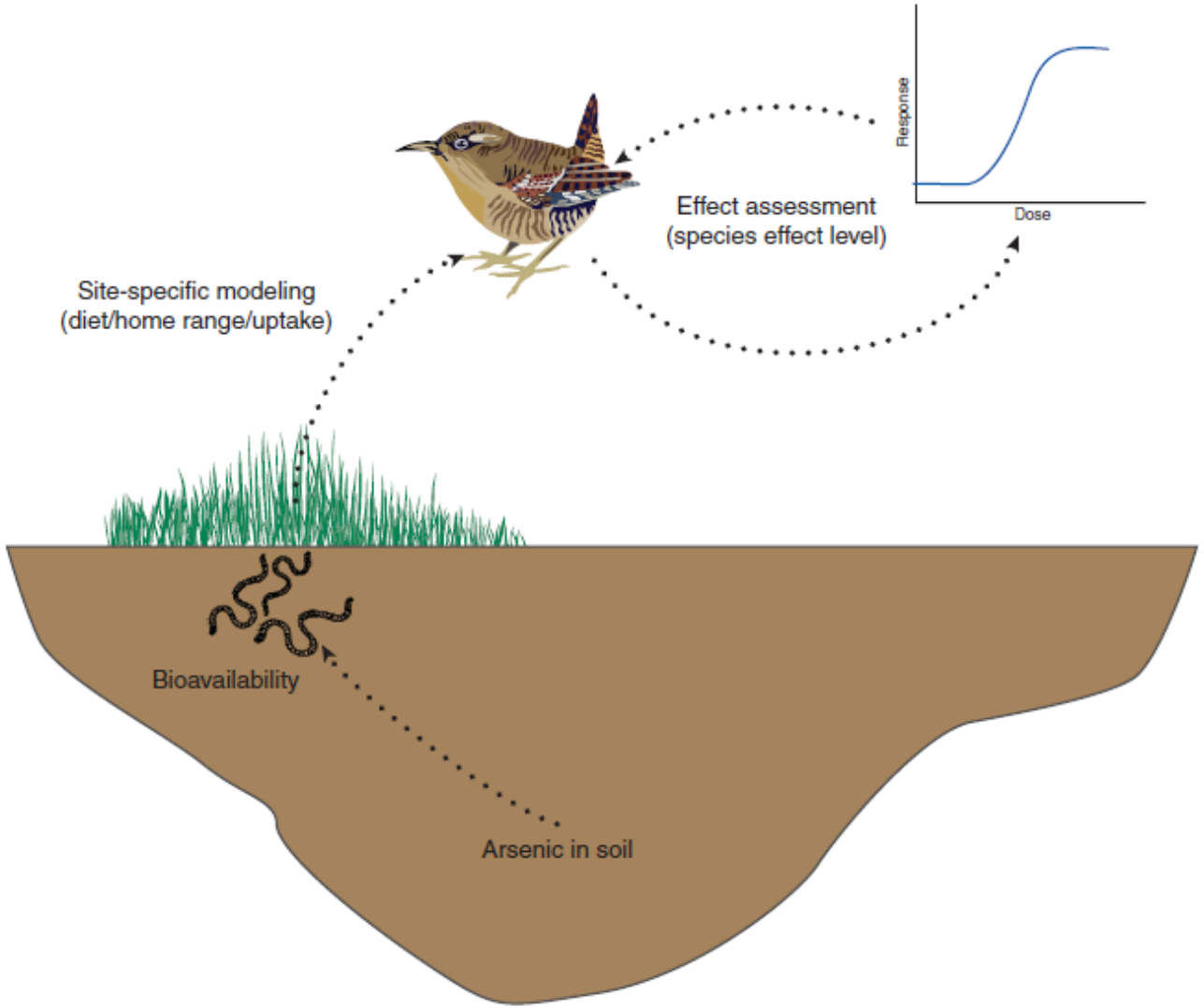
Figure 2-4. Example of Box Diagram Conceptual Site Model



Receptors of Concern							
Aquatic Plants	Benthic Infauna	Benthic Epifauna	Demersal Fish	Piscivorous Fish	Piscivorous Birds	Wading Birds & Waterfowl	Mammals
○	●	●	●	○	○	●	●
●	●	●	○	○	○	○	○
○	○	○	○	○	●	●	●
●	●	●	●	●	○	○	○
○	●	●	○	○	○	●	●
○	●	●	●	○	○	●	●
○	○	○	●	●	●	●	●

Note: solid bullets indicate pathways that are likely to be operational

Figure 2-5. Example of a Customized Conceptual Site Model



2.3. Designing and Planning the ERA

This section discusses aspects of problem formulation that are aimed at preparing for implementation of the ERA, with a focus on the tools and analyses that will be used to evaluate potential risks for each ROC and exposure pathway. The design and planning stage includes the following (for terminology and key concepts, refer back to **Section 2.1.2**):

- Establishing protection goals and (usually) associated acceptable effect levels;
- Identifying assessment endpoints – the attributes of the receptors that are to be protected (e.g., abundance or viability of a mammal population);
- Identifying measurement endpoints (the tools used to measure changes in assessment endpoints);
- Developing lines of evidence (LOEs) for each assessment endpoint, which specify how measurement endpoints will be used to evaluate potential risks; and
- Articulating the strategy for the ERA, as well as the sampling and analysis plan (SAP).

> ***Key Concept:***
>
> Relationships between assessment endpoints, measurement endpoints and lines of evidence are shown in **Figure 2-6**, and an example is provided in **Table 2-7**. In the example, LOEs are grouped around major sources of data.

Importantly, as with the earlier sections of the problem formulation, the elements in this section are inter-related and therefore developed in an iterative manner.

2.3.1. Protection Goals and Acceptable Effect Levels

For most[16] ERAs, there is a description of the type and level of protection that is intended for each receptor or receptor group at a site. This information may be used to "judge" the results of the risk assessment. A protection goal may be a narrative statement that is then operationalized through an "acceptable effect level" (AEL) that clarifies the magnitude or

[16] Not all ERAs require protection goals. Consistent with the CCME (1996) formulation of 'detailed' ERA, risks may simply be characterized, with all judgements about acceptability being made after the ERA is complete.

Figure 2-6. Conceptual Relationships between Assessment Endpoints, Measurement Endpoints and Lines of Evidence.

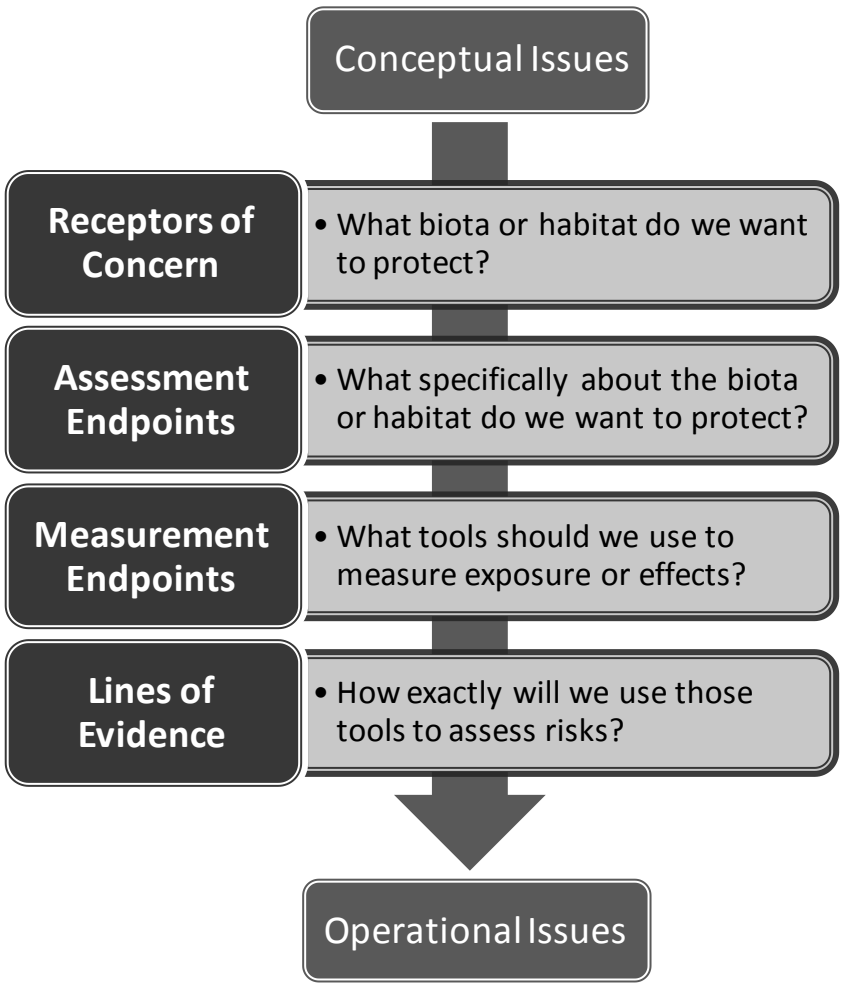


Table 2-7. Example table of assessment endpoints, measurement endpoints and lines of evidence

Receptor Group(s)	Assessment Endpoint	Lines of Evidence	
		LOE Group	**Use of measurement endpoints for specific LOEs**
Benthic invertebrates	Aquatic invertebrate community structure, and ecological function as food for fish and wildlife	LOE 1 – Sediment Chemistry	• COC concentrations – comparison of COC concentrations to CCME sediment quality guidelines, with qualitative interpretation of potential bioavailability as measured by SEM:AVS.
		LOE 2 – Benthic Community Analysis	• Benthos abundance and diversity measures (total organisms, total taxa, Simpson's index) – ANOVA and paired tests to compare water bodies adjacent to the site to reference conditions • Benthos abundance and diversity measures (total organisms, total taxa, Simpson's index) – regression of each measure on sediment COC concentrations and SEM:AVS (a measure of potential bioavailability of certain metals)
		LOE 3 – Amphipod toxicity test	• Amphipod growth – ANOVA and paired tests to compare growth between on-site and reference samples, relative to control. • Amphipod survival – ANOVA and paired tests to compare survival between on-site and reference samples, relative to control. • Amphipod growth and survival – regression of growth and survival on sediment COC concentrations and SEM:AVS (a measure of potential bioavailability of certain metals)
Birds, mammals, amphibians	Abundance and viability of local bird, mammal and amphibian populations	LOE 1 – Food chain model	• Comparison of estimated exposure to COCs (total dose via all exposure pathways) to Toxicity Reference Values relevant to effects on growth, survival, and reproduction.
		LOE 2 – Small mammal trapping	• Comparison of catch-per-unit effort (as an index of abundance) of small mammals on-site versus reference conditions
		LOE 3 – Wildlife survey	• Qualitative observations of presence/absence of particular wildlife receptors, based on a survey by a wildlife biologist.

rate of effects that would be acceptable for a specific measurement endpoint or a group of measurement endpoints[17]. A protection goal usually differs for common species (where the population-level is often of interest) relative to listed species[18] (where individual organisms may need to be protected), and may differ according to land use or the overall site management goals (**Section 2.2.1**) of the ERA[19].

Examples of narrative protection goals may include:

> ***Key Concept:***.
>
> Protection goals and associated acceptable effect levels (AELs) applied to measurement endpoints should be articulated in the problem formulation.

- Maintenance of populations and associated demographics of small mammals that are similar to background conditions;
- No adverse organism-level impacts on the western toad (a listed species);
- Low level of significant ecological effects, defined to allow small structural or functional changes that may exceed natural variability provided that such do not threaten the sustainability of receptors[20] (this is specified as a goal for commercial and industrial lands in Quebec [CEAEQ 1998])

Protection goals are not always operationalized immediately as AELs, but rather may be left in narrative form until measurement endpoints and lines of evidence are specified. Where they are applied, they are intended to provide a degree of consistency across assessments, and as such are often influenced by policy determinations rather than by technical criteria. Nevertheless, AELs may vary by ROC, by endpoint, and by site depending on several factors including:

- Is protection aimed at individual organisms, populations or communities?
- Is the ROC a common species, or a species at risk (e.g., listed as rare / endangered)?
- Are there relevant federal or provincial laws, or pertinent policy determinations, that dictate appropriate AELs?

---

[17] Given the inter-linkages between AELs and endpoints, they are typically developed at the same time. An AEL can be applied directly to the assessment endpoint if the assessment endpoint is quantitative.

[18] Refers to species that are formally designated at provincial or federal level as, for example, rare or endangered or threatened..

[19] For example, Quebec guidance (CEAEQ 1998) specifies more stringent protection goals when protection of biodiversity is an overall objective for a site.

[20] Original text: Un faible niveau de réponses écologiquement significatives, c'est-à-dire un faible changement structurel ou fonctionnel pouvant excéder la variabilité naturelle mais ne mettant pas en cause la pérennité des récepteurs

- Can appropriate AELs be inferred from methods used to derive national or provincial environmental quality guidelines?
- What effect size can be reasonably detected given natural variability?
- What effect size (at individual, population or community level) would be ecologically relevant for the particular ROC?
- What are the spatial and temporal scales at which the effect will occur?
- Would the effect be reversible?
- What are the environmental or economic consequences of a Type I error (false-positive) or Type II error (false-negative) in the risk assessment conclusions?

As implied by the range of considerations above, derivation of ecologically-meaningful AELs can be complex. However, even if AELs are not explicitly articulated, they are often implicit. For example, any wildlife ERA that uses a published toxicity reference value (TRV) to estimate a hazard quotient is assuming an AEL that is equal to the response size specified in the derivation of the TRV (see **Section 4** for further discussion of TRVs). The risk assessor should ensure to the extent possible that the AEL implicit in a TRV is compliant with the protection goal.

> ***Key Concept:***.
>
> AELs should be based on ecologically-relevant effect sizes.

For risk assessments where AELs are specified, it is preferable to base AELs on effect sizes that are defined in advance, as opposed to effect sizes that happen to be statistically significant based on hypothesis tests. Thresholds that are specified based on statistical significance in hypothesis tests are subject to large variation in ecological significance, depending on the level of statistical significance chosen, the specific study chosen, and details of the experimental design such as the range of treatments and sample sizes. An AEL based on a pre-defined effect-size facilitates application of concentration-response methods (e.g., Allard *et al.* 2010) that provide a more standardized level of protection across contaminants, receptors, and assessments.

2.3.2. Assessment Endpoints

An assessment endpoint is an explicit expression of the environmental value to be protected. An assessment endpoint must include a receptor (or receptor group – i.e., an entity to be protected) and a specific property or attribute of that receptor. For example, if the receptor is a fish community, candidate endpoint properties could include the population demographics, biomass, the genetic variability, the physical condition, or the trophic structure.

The distinction between assessment endpoints and protection goals is a subtle one; specifically, the former describes the environmental attribute of interest, whereas the latter articulates the desired state of that attribute. To distinguish an assessment endpoint from a protection goal, practitioners should avoid using assessment endpoints that express an objective or a desired state (e.g., "healthy" or "functional"), and instead apply value-neutral terminology. The following examples illustrate this point:

- "Benthic community diversity" (assessment endpoint) versus "maintenance of a diverse benthic community" (protection goal);
- "Osprey reproduction" (assessment endpoint) versus "successful osprey reproduction" (protection goal);
- "Marmot abundance" (assessment endpoint)"Self-sustaining marmot population" (protection goal).

Checkai *et al.* (2002) identify other pitfalls during endpoint identification, including assessment endpoints that:

- are too vague (e.g., "stream integrity" rather than "abundance of juvenile salmonids");
- evaluate an overly specific ecological entity (e.g., *Hyalella* growth instead of abundance of benthic fish prey). When assessment endpoints are too specific, they may be poorly aligned with the stressors of concern in terms of sensitivity and relevance;
- are difficult to operationalize (i.e., cannot be fully considered in the ERA); or
- are not sufficiently sensitive given the management goals (e.g., if the management goal is to assess potential effects on wildlife, an assessment endpoint based on "presence versus absence of wildlife" would be too coarse to be useful).

2.3.3. Measurement Endpoints

A measurement endpoint[21] is generally any measure of exposure or effects for an ROC or any measure of change in the attribute of an assessment endpoint. Measurement endpoints form the basis for lines of evidence used to estimate risks (see **Figure 2-6** and **Table 2-7** above). Examples of measurement endpoints include:

[21] The term *measurement endpoint* is preferred to 'measure of effect' because the broad definition of measurement endpoint can include not only measures of effect (measurable change in an attribute), but also measures of exposure (measures of stressor existence, bioavailability, and movement) and measures of ecosystem and receptor characteristics (characteristics that influence or mediate the relationship between exposure and effect) (Checkai et al. 2002).

- Survival and growth of giant kelp (*Macrocystis pyrifera*) gametophytes exposed to field-collected seep samples;
- Plant biomass per unit area;
- Simpson's diversity index for soil invertebrate samples;
- Abundance of mayflies, caddisflies, and stoneflies (known as EPT taxa – Ephemeroptera, Plecotera, Trichoptera) per standard grab; or
- Molar ratio of acid volatile sulfides to simultaneously extractable metals (SEM:AVS), as an indicator of potential bioavailability.

These measurement endpoints are measures of either exposure or effects, but not both. In general, to maintain the distinction between measurement endpoints and lines of evidence, these simple types of measurement endpoints are preferred. More complex formulations of measurement endpoints that attempt to incorporate both exposure and effects information (e.g., comparison of deer mice density on-site and off-site; or comparison of the daily ingested COC dose for deer mice at the site to a dose-based toxicity reference value that represents an acceptable effect level) are no longer measurement endpoints but lines of evidence. Measurement endpoints and LOEs must be developed at the same time, otherwise a measurement endpoint would be proposed without any understanding of how the information will be used.

2.3.3.1. Criteria for Selection of Measurement Endpoints

Measurement endpoints are selected in the context of particular receptor groups and assessment endpoints. Consequently, selection of measurement endpoints does not occur in isolation. Some of the criteria relevant to selection of receptors (**Section 2.2.5.3**) are therefore directly relevant to selection of measurement endpoints. Major technical criteria relevant to selection of measurement endpoints are reviewed in the context of lines of evidence in **Section 2.3.4**. Importantly, in addition to technical criteria, risk assessors also consider practical criteria such as cost, feasibility, and time constraints. Cost is best addressed separately, but should be addressed nonetheless. Ideally, the first tier of an ERA uses measurement endpoints that are both effective and cheap. Unfortunately, some tools are either effective and expensive, or less effective but inexpensive. In such cases, the risk assessor should be explicit about the trade-offs made between cost and expected effectiveness of the measurement endpoints. In many ERAs, an iterative approach may be used, whereby the measurement endpoints that offer the best value (effectiveness per unit cost) are used first, and additional measurement endpoints are used in subsequent iterations as needed. Regarding time constraints, if there is a regulatory or other driver that requires completion of an ERA within a particular timeframe, measurement endpoints for which turn-around time is short may be preferred.

2.3.3.2. Level of Organization – Organism, Population, Community

It is desirable to maximize the correspondence between assessment and measurement endpoints, such that attributes are measured that are functionally related to the environmental property of interest. It is desirable, but not necessary to align measurement and assessment endpoints across a common level of ecological organization.

> ***Key Concept:***.
>
> Relating effect sizes in individual organisms to effects at the wildlife population level is a key challenge in ERA. At the least, the risk assessor should provide qualitative judgements based on life history characteristics of the ROCs.

For example, for the assessment endpoint "passerine abundance" the measurement endpoint might be "percentage difference in density of adult breeding pairs of American robins on site X compared to reference conditions" – in this case a population-level attribute (density of breeding pairs) is applied to the local population of robins. In this example, the measurement and assessment endpoints are both expressed at the population level. However, if it is difficult to measure the number of breeding pairs, or if it is a highly variable measure, or if the measure is likely to be confounded by immigration from off-site, alternative measurement endpoints that might be considered include "mortality rate and reproductive success of robins on site X". In the latter case, two organism-level attributes (mortality and reproductive success) are assumed to be representative measures that may be extrapolated to the local population of robins. Such extrapolation could be conducted qualitatively using a narrative or quantitatively using a population model.

A key property of any measurement endpoint should be the ability to interpret the results in relation to protection goals. If the protection goal is "minimal effects to a terrestrial mammal community," it is desirable to be able to relate changes in measurement endpoints to potential effects on populations and ultimately to that community. In practice this is quite challenging; this issue has plagued ecotoxicologists for several decades due to the complex linkages and uncertainties in ecological systems, including density dependence, intraspecies sensitivity variations, and confounding habitat factors. Although ERA practices have evolved to address some of the uncertainties, such as adjustments using extrapolation factors and safety factors, there remains a significant degree of difficulty in extrapolating across levels of organization, and complexity in understanding dynamics at higher levels of organization..

Most measurement endpoints in ERAs address organism-level attributes of a population or community (Suter *et al.* 2005) such as mortality rate, reproductive success, and growth. Assessment endpoints commonly address populations or communities whereas

measurement endpoints address organism-level attributes that are believed to be linked to the population / community-based assessment endpoint (CCME 2006). Although the quantitative linkage between organism-level attributes and responses to populations or communities is seldom known with confidence, it is usually assumed that there will be no effects at the population or community level if an ERA predicts no effects at the organism level. If the ERA predicts effects to individual organisms, it is not easy to predict effect levels for populations or communities. For example, a population that is already at carrying capacity may be unaffected by a higher mortality rate among individual organisms. Conversely, a population that is barely able to sustain itself may be extirpated under any additional stress. Extrapolating from organism-level attributes to populations requires an understanding of factors controlling population dynamics. Extrapolating from populations to communities requires an understanding of community interactions (e.g., one species may increase in abundance if its associated predator decreases in abundance). Although there is a desire in the ERA community to develop methods, normally ERA practitioners describe possible community effects qualitatively, if at all.

Some measurement endpoints are easy to interpret ecologically because they address true community-level or at least population-level attributes, but they have only the power to detect very large changes. Other measurement endpoints that address organism-level attributes are more powerful at detecting change, but are less-easily extrapolated to populations and communities. It is partly for this reason that ERAs depend on multiple lines of evidence. Nevertheless, methods for evaluating effects on population and communities exist (see Suter 2007) and should be employed whenever possible.

A specific difficulty in evaluating population-level effects lies in defining the population of interest (i.e., the assessment population). From a pure biology perspective, an ecological population is defined as a group of organisms of a single species that interbreed and share a common habitat. From a risk assessment perspective, however, this definition is too broad, particularly for organisms that migrate across large areas (up to the continental scale). If assessment populations are defined across large spatial scales, then effects on local groups of individual organisms near a particular contaminated site may not have an impact on the assessment population, yet still exert local impacts that are considered unacceptable in relation to protection goals.

> ***Key Concept:***.
>
> Whenever populations are of interest, particularly for wildlife, the population of management interest (the assessment population) should be defined as clearly as possible.

A further issue with respect to defining populations is understanding the ecological context of the group identified as a local population. A small patch of forest in the middle of farmland or an urban center may play an important role (e.g., migration corridor) when

overall habitat is fragmented, whereas a similar-sized area located in an unfragmented wilderness area may be less sensitive to ecological disruption.

The general issues of spatial scale and the overall magnitude of effects are addressed in more detail in **Section 5**.

2.3.3.3. Types of Effects Used in Measurement Endpoints

For measurement endpoints that are direct measures of effects, there is general agreement that certain types of effects are more suitable than others. Specifically, effects that are measured need to be ecologically relevant, and linkable back to assessment endpoints that focus most often on population or community-level attributes.

CCME (2006) notes that effects that are measured at the organism level should be those that are critical for a species to complete a normal lifecycle and to produce viable offspring. Mortality and reproduction are the two types of effects that can be most easily related to population-level effects, but population dynamics are typically complex and there may be several direct and indirect mechanisms by which lethal and sublethal effects could impact at population level. Direct measures at the population and community levels are ideal but not often feasible or practical to obtain.

> ***Key Concept:***.
>
> Mortality and reproductive effects on individual organisms can be most easily extrapolated to effects at population level. Growth effects can be related to effects at population level in terms of biomass. Other types of effects may be sensitive to COCs, but careful consideration must be given to the ability to extrapolate effects on individual organisms to populations.

Other types of effects can be applied as surrogates for population responses but are generally more difficult to relate to population- and community-level effects. This is particularly true for effects that are not truly a measure of an adverse effect but are rather a measure of the potential for adverse response (e.g., enzyme induction).

Some guidance on the types of effects that should be used for measurement endpoints is provided in various Canadian jurisdictions (see **Table 2-8**). There are no clear rules for use of these types of effects in endpoint selection; ERA practitioners must use their judgement on a case-specific basis. When considering the advantages and disadvantages of various types of effects (**Table 2-9**), preference should be given whenever possible to types of effects that are as closely tied to assessment endpoints as possible. Endpoints that are relevant to assessment endpoints include direct measures at the population and community levels, or mortality and reproductive effects that can be directly related to population-level attributes. Among the other types of effects, growth is generally the most preferred. This does not mean that other endpoints should be excluded from consideration. If behavioral effects with likely implications at the population level

Table 2-8. Types of effects and their acceptability in various jurisdictions for use in measurement endpoint selection[22]

Types of Effects	CCME (1997a)	SAB-CS (2008)	OMOE (20011)[23]	CEAEQ (1998)	CCME (2006)	CCME (2007)	CCME (1995)
Mortality	Acceptable	Acceptable	Acceptable	Acceptable	Acceptable	Acceptable	Acceptable
Reproduction	Acceptable	Acceptable	Acceptable	Acceptable	Acceptable	Acceptable	Acceptable
Growth	Acceptable	Acceptable	Acceptable		Acceptable	Acceptable	Acceptable
Behavior	Acceptable	Acceptable	Variable			Variable	Variable
Morphology / Deformity		Acceptable	Variable				
Tumors	Acceptable		Variable				
Physiological measures such as absorption efficiency, nutrient uptake, blood volume, etc.	Acceptable		Variable			Variable	Variable
Enzyme activity	Acceptable		Not acceptable			Variable	
Histopathology (cellular changes)	Acceptable	Acceptable	Not acceptable				
Development (some measures only – e.g., sexual development)			Variable				Acceptable
Immunological response			Not acceptable				
Population-level attributes (e.g., biomass, abundance)	Acceptable	Acceptable	Acceptable	Acceptable			
Community-level attributes (e.g., diversity)		Acceptable		Acceptable			

[22] This table is simplified. Some guidance documents make exceptions on a case-by-case basis, with a key criterion being whether a particular effect is likely to affect survival, reproduction or growth. Blanks in the table do not indicate acceptable or unacceptable, rather no specific mention was made of that effect type.

[23] The OMOE (2011) guidance varies by receptor group, and there are exceptions in many cases.

Table 2-9. Advantages and disadvantages of the use of particular types of effects as measurement endpoints

Type of Effect	Key Advantages (A) and Disadvantages (D)
Direct measures at community-level, such as diversity, species richness or biomass	• A: Usually directly relevant to the assessment endpoint • D: Can be difficult to measure directly; power to detect change may be limited.
Direct measures at population level, such as abundance or biomass	• A: Usually directly relevant to the assessment endpoint • D: Can be difficult to measure directly; power to detect change may be limited, and ability to establish causal relationships with stressors may be limited.
Mortality rates	• A: Easy to measure in some cases; response times usually fast; relatively easy to relate to population level • D: Often less sensitive than other endpoints; difficult to measure in situ (e.g., for wildlife); responses may be delayed.
Reproductive endpoints (e.g., fecundity, reproductive success)	• A: Can be easy to measure depending on the specific effect; relatively easy to relate to population level depending on the specific effect; can be an indicator of other unknown effects • D: Some reproductive effects are not easy to measure (e.g., for wildlife) and have long response times for some receptors
Growth	• A: often more sensitive than mortality or reproductive endpoints; can be an indicator of other unknown effects • D: More difficult to relate to population level.
Behavior (where the behavior could be linked to mortality, such as predator avoidance, or to reproduction, such as mating frequency)	• A: often more sensitive than mortality or reproductive endpoints; can be an indicator of other unknown effects; can often be linked to reproduction and mortality. • D: Links to population and community level may be vague, effects may be subtle, and response times may be long.
All other types of endpoints (see examples in **Table 2-8**)	• A: May be more sensitive to contaminants than other endpoints • D: Difficult to relate to population and community level assessment endpoints, or may not have a net adverse effect.

(e.g., decreased predator avoidance) are observed at low concentrations, those should be considered relevant[24].

2.3.4. Lines of Evidence

2.3.4.1. LOE Definition

The ways in which measurement endpoints are organized and applied define the lines of evidence (LOEs) that will be carried through the ERA. Lines of evidence are derived from assessment and measurement endpoints (**Figure 2-6**; examples shown in **Table 2-7)**. Although the figure shows a stepwise process, in reality the lines of evidence should be developed nearly concurrently with the measurement endpoints (i.e., there is no point identifying a tool without thinking ahead to the proposed application of the results).As highlighted previously, the scope of measurement endpoints varies widely and they can be defined in a way that makes them functionally equivalent to a line of evidence. Generally, it is easier to define measurement endpoints as measures of exposure or effect so that they are clearly distinguished from lines of evidence. The expression of LOEs provides a bridge between the unprocessed data collected to inform the risk assessment (measurement endpoints) and the subsequent analysis/interpretation of those data in the analysis stage of the ERA.

Definition:

Line of evidence (LOE) – any pairing of exposure and effects measures that provides evidence for the evaluation of a specific assessment endpoint. Typically a line of evidence requires use of one or more measurement endpoints.

Key Concept:.

It is helpful to view measurement endpoints as tools, and lines of evidence as the use of those tools in one or more ways.

For example, if we measure species diversity in a soil invertebrate community (measurement endpoint), the data could be applied in several ways, including:

(1) Comparisons of mean diversity index on-site versus reference conditions (*e.g.*, using ANOVA);

(2) Comparison of a diversity index to pre-determined thresholds for soil quality based on ecological principles; and

[24] CCME (2007) notes that, for derivation of water quality guidelines, nontraditional endpoints such as behaviour can be used if ecological relevance can be demonstrated.

(3) Modeling of the diversity index versus the soil concentration of a contaminant (*e.g.*, using simple linear regression).

These would be considered as separate lines of evidence derived from the same measure of effect (diversity index). Two of the analyses measure the magnitude of potential risks, while the other focuses on establishing potential causal relationships with contamination. Each LOE carries different (but valuable) information for informing the assessment endpoint.

From a practical perspective, it may be appropriate to group closely-related LOEs. For example, as illustrated in **Table 2-7**, all LOEs that use results of an amphipod toxicity test in one way or another may be grouped together for purposes of analysis and reporting, even though that toxicity test may have more than one specific measurement endpoint (e.g., growth and survival) and there may be several specific LOEs developed that use the results of the toxicity test.

In specifying LOEs, it is important to provide a clear expression of the relationship between exposure and effects measures. For some LOEs, the relationship is obvious (e.g., comparison of soil chemistry to soil quality guidelines). In other cases, the relationship is less intuitive and requires explanation (e.g., benthic invertebrate community diversity as a function of proximity to a point source). In the latter example, the "proximity" could be a function of distance, direction, or both, and the LOE may require specification of groupings of stations, distance-based transects, or other measure of exposure.

2.3.4.2. LOE Organization

To facilitate consistency in practice, it is helpful to conceptualize the following four major categories of lines of evidence (LOEs), as follows:

- **Site-specific toxicological evidence** – Considers measurement endpoints related to studies of test organism exposures to contaminated site media under controlled conditions[25].
- **Indirect toxicology evidence** – Considers toxicological information gleaned from other sites, under an assumption that the concentration-response relationship is either similar to, or can be estimated from, the data collected at other sites.
- **Site-specific biological evidence** – Considers direct assessment of the site biological condition.
- **Indirect biological evidence** – Considers indirect assessment of biology, through extrapolation of knowledge obtained at other sites.

[25] Controlled conditions may be in the laboratory or *in situ*.

The framing of LOEs in this manner streamlines the risk characterization (**Section 5**) and is consistent with the framing of tools in the effects assessment (**Section 4**).

It is common for one or more of these LOE categories to be omitted from any given stage of a risk assessment, depending on the scope / complexity of the study, the project tiering strategy, and the objectives of the risk assessment. For example, site-specific toxicology studies are infrequently conducted for birds and mammals, and are practically non-existent for endangered species. Similarly, site-specific biological investigations are seldom conducted during a preliminary quantitative risk assessment (PQRA). Lack of representation of any of the four types is, unto itself, not a cause for criticism. However, the risk assessor should explicitly acknowledge the implications and uncertainties associated with emphasizing or omitting any of the above LOEs.

Each of these broad LOE categories carries different uncertainties and evaluation methods. Furthermore, it is common to have multiple individual LOEs within a single broad LOE category, as follows:

- Site-specific toxicity tests are commonly conducted as part of a test battery approach with multiple species, durations, and endpoints;
- Comparisons to guidelines / benchmarks may entail multiple comparisons (different jurisdictions or case studies);
- Community studies have a multitude of potential endpoints (e.g., total density and diversity, major taxa density and diversity, sensitive taxa density and diversity, diversity indices);
- Biological endpoints from other sites can be numerous in type;
- Biological and toxicological endpoints can be compared against many candidate exposure metrics (e.g., chemistry of individual COCs, chemical surrogates such as toxic equivalents (TEQ), multivariate chemistry exposure metrics (principal components), distance/direction metrics, etc.).

One of the reasons for grouping endpoints into the four broad categories is to explicitly acknowledge the partial redundancy of having multiple related endpoints. Of course, formal organization of LOEs is not important for simple ERAs where there may only be a limited number of LOEs needed to address risks, or for cases where a limited number of LOEs lead to a clear conclusion that risks are negligible.

2.3.4.3. LOE Selection

The rationale for selection of LOEs should be explicit as part of problem formulation. The LOEs that are used for an ERA (see example in **Table 2-7**) are derived from prior

consideration of a long list of potential LOEs. The criteria that are relevant for selecting LOEs may include[26]:

- Ecological relevance – To what degree is the assessment endpoint represented by the LOE?
- Sensitivity – To what degree can the LOE detect change or differences from reference conditions? Are results reported quantitatively or using broad categories such as low, moderate, high? Does the LOE typically suffer from a high degree of random error?
- Specificity – To what degree is the LOE capable of distinguishing effects of site-related COCs from other factors?
- Spatial representativeness and site specificity – Does the LOE provide information that is site-specific and at a spatial scale relevant to assessment endpoints?
- Temporal representativeness – Does the LOE capture temporal variation relevant to potential ecological risks?
- Expected data quality – Based on the practitioner's experience, what is the likelihood that the quality of data generated by this LOE will be poor and result in reduced utility of the LOE?
- Expected acceptability – Does the LOE have standard test methods or a long history of use that provide confidence that regulators will accept the results?

At the least, practitioners should provide a list of LOEs that were considered for a particular ERA, with rationale for inclusion or exclusion of each. This can be done in text or in a table. The rationale should be based on appropriate criteria such as those listed above. For complex ERAs where a phased approach to implementation is used, rationale should also be provided to justify which LOEs are proposed initially and which are deferred.

2.3.4.4. LOE Application

> ***Key Concept:***.
>
> It is important to cross-check lines of evidence against exposure pathways for the ROCs, to ensure that no exposure pathways are missed.

Because LOEs are carried forward in the ERA to evaluate risks, it is important to cross-check selection to ensure that the selected LOEs represent all receptor groups and all exposure pathways.

[26] Adapted in part from Menzie et al. (1996) and SAB-CS (2010).

Receptor groups are inherently cross-checked during preparation of the LOE table (**Table 2-7**). For cross-checking exposure pathways, the simplest approach is to add a single yes/no column to the templates in **Tables 2-5 and 2-6** to confirm yes or no that there are one or more LOEs proposed that are relevant to each exposure pathway.

Lines of evidence are evaluated and implemented as part of the ERA. The specific approach that will be used to integrate the LOEs in a weight-of-evidence framework should be described as part of the general strategy for the ERA (see next section). The level of detail that needs to be provided depends on the complexity of the ERA and the type of approach used for risk characterization. In **Section 5**, the range of options for conducting weight-of-evidence (WOE) assessments are discussed. The most detailed WOE approaches may require formal quantitative assessment of the LOEs during problem formulation (e.g., weighting or ranking of respective LOEs based on a multi-attribute assessment of each). While such approaches may be overly cumbersome for most ERAs, critically assessing LOEs for relevance prior to data analysis guards against "ad-hockery," which may be defined as gratuitous assumptions (made after the fact) that provide superficial appearance of a systematic decision-making process, but are in fact arbitrary and impossible to discern from subjective interpretations. Regardless of the level of formality used, it is important that rationale be provided for the selection of LOEs at the problem formulation stage, as discussed in the previous section. Criteria such as ecological relevance that do not change based on data collected after the problem formulation can be carried forward directly from the problem formulation to risk characterization in a WOE framework (see **Section 5**).

2.3.5. General Strategy for the ERA

At the same time that LOEs are developed, it is important to design the overall implementation strategy for the ERA. The strategy should not get into details about field methods, lab methods or data analysis methods, as those are best left to the sampling and analysis plan (see next section). The strategy focuses on big-picture issues, typically covering:

- Phasing / iteration – Will the ERA be implemented in phases? If yes, what LOEs will be pursued in which phases? Under what conditions (results) could the first phase be sufficient for the ERA to be considered completed?

> ***Key Concept:***.
>
> The general strategy for the ERA provides a high-level overview regarding the approach to be used for the ERA. The general strategy should be part of every problem formulation. On the other hand, development of the detailed sampling and analysis plan may in some cases be deferred until there is agreement on the general strategy and lines of evidence proposed for the ERA.

- Timeline – Implications of phasing and other constraints should be presented as they relate to the overall timeline expected for the ERA.
- Experimental design (see subsection below for more discussion on experimental design) – Will field studies incorporate a gradient design or comparison of the site to a reference condition? What amount of field replication will be needed in order to have adequate power to detect effect sizes of interest or to establish correlations between exposure and effects? What is the general spatial scale of sampling for each type of data? While details are listed in the sampling and analysis plan, the conceptual design should be articulated as part of the general strategy.
- Coordination with ongoing site investigation – If supplemental site investigation work is ongoing, how will that work mesh with and support the ERA? How will the site investigation data be used in the ERA?
- Approach to risk characterization – Assuming there is more than one LOE for at least some of the assessment endpoints, it is important to describe during problem formulation how the weight of evidence approach to risk characterization will be implemented. This should include details regarding

  > ***Key Concept:***.
  >
  > The weight of evidence methodology should be described in the problem formulation and implemented during risk characterization.

  - how LOEs will be summarized and integrated; and
  - how judgements about the magnitude of risks, uncertainty about risks, causation or other attributes will be made (a default table is provided in **Section 5** for this purpose).

  In short, the details of how risk characterization will be implemented should be fully understood and articulated at the problem formulation stage. Detailed discussion of risk characterization including WOE approaches is deferred to **Section 5** for organizational simplicity, but most of the content is relevant (i.e., must be considered) at the problem formulation stage.
- Transparency – How will the ERA results as a whole be presented? What mechanisms or tools will enable reviewers to understand how conclusions were drawn? What mechanisms or tools will enable reviewers to make

  > ***Key Concept:***.
  >
  > ERA practitioners must ensure that risk assessment results and conclusions are presented in a transparent manner, so that reviewers, stakeholders and decision-makers can easily understand the findings and make their own judgments.

independent evaluations of risk based on the information presented?

The general implementation strategy should be discussed with site custodians (and possibly with FCSAP expert support, regulators and stakeholders) in order to confirm that the ERA will fulfill expectations. Provided there is agreement on the strategy, a sampling and analysis plan can be prepared prior to commencement of the work.

2.3.5.1. Control-Impact versus Gradient Designs

Experimental design warrants careful consideration in ERA, because the design dictates what types of inferences can be drawn from the data collected. An important element of experimental design for the practitioner is deciding in advance how the potential effects of contamination will be evaluated. The classic "control-impact" design that is often used in ERA to compare a site to a reference site has fundamental problems because of natural variability among sites unrelated to contamination. Comparison to a reference condition (based for example on multiple reference sites) is preferable but is also confounded to some extent by natural variability among sites. In most cases, a gradient design should be considered, as it allows the practitioner to evaluate potential relationships between contamination and effects, and to understand any differences observed between areas of varying concentrations of contaminants. The following discussion provides rationale in this regard.

In a classic control-impact design, the effect of contamination would be interpreted by comparing site-related performance to control performance. For example, a practitioner may compare plant growth at a contaminated site to plant growth at a control or "reference site", assuming the two sites are identical except for the contamination. Unfortunately, no two sites are identical, so comparison of a site to a single reference site is of limited value. If multiple samples are taken for both sites, the hypothesis that the two sites have similar plant growth can be tested statistically, but any difference between the two sites cannot be taken as evidence as a contaminant-related effect, because we should expect the two sites to be innately different even in the absence of contamination. A practitioner who incorrectly assumes that a statistically significant difference in this case is related to contamination is committing pseudoreplication (Hurlbert 1984) because the data only provide evidence of variability between those two particular sites, not variability between contaminated and uncontaminated sites in general. The samples at each site are considered pseudoreplicates, not true replicates in a test for the effect of contamination.

One approach to address the problem of inherent variation among sites is to define a "reference condition" against which a contaminated site could be evaluated. In the context of a contaminated site, a reference condition would usually be one that is assumed to represent a range of conditions that would occur in the absence of site-

specific contamination. A reference condition could be established in various ways (Stoddard *et al.* 2006), the most common of which is to use multiple reference sites to establish a range of conditions that represent reference. For example, the reference condition approach based on multiple reference sites is used under the Canadian Aquatic Biomonitoring Network (CABIN)[27] to evaluate potential impacts of stressors on freshwater aquatic systems. This is a vast improvement over comparison to a single reference site, but since the contaminated site itself is not replicated the problem remains that we do not know to what extent any observed differences between the site and the reference condition are natural or related to contamination. Provided that practitioners understand this limitation, comparison of a contaminated site to reference conditions derived from multiple reference sites can be useful. Other approaches for deriving a reference condition include interpreting historical condition (if information exists for conditions at a site before contamination); extrapolating from empirical relationships relating biological indicators to contamination (e.g., from other sites); or using ecological principles to specify expected conditions in the absence of contamination (Stoddard *et al.* 2006).

Comparison of biological variables at a contaminated site to a reference condition will always be confounded by the inherent variation in biological systems. Because assessment endpoints for ERAs are often at population or community levels of organization, practitioners should expect that the population or community of interest at a site will be inherently different from the population or community associated with any other particular site or set of sites. Landis *et al.* (2011) provide detailed arguments as to why "there is no such thing as a reference site when it comes to populations and landscapes." The implication is that the sampling design for ERAs should usually focus less on testing for a difference between the site and a reference condition, and more on evaluating patterns based on gradients of contamination and other factors that are likely to drive biological variability (Landis *et al.* 2011). This is particularly true for measurement endpoints that measure populations or communities directly.

Gradient designs should aim to capture the range of COC concentrations from highest (on-site) to lowest (could be on-site or away from the site). Some gradient designs may have a directional spatial element such as distance from a point-source of contamination. Most importantly, gradient designs should control for patterns in environmental variables that may be correlated with contamination. Confounding variables often limit the ability of practitioners to make links between observed biological patterns and site-related contaminants. Inherent in a gradient design is the objective of determining whether populations and communities of interest are correlated with *and caused by* contamination. As discussed at length in Section 5, this latter aspect – establishing causality – should be a key component of any ERA.

[27] http://www.ec.gc.ca/rcba-cabin/default.asp

In short, whenever possible the study design for ERAs should aim to characterize gradients in contamination and other factors that are likely to drive responses in populations and communities. Comparison of a site to reference conditions is also useful, but conclusions based solely on such comparisons are limited when relationships to contamination and other predictor variables are not understood.

2.3.6. Sampling and Analysis Plan

A sampling and analysis plan (SAP) describes details of how the ERA will be implemented. Because it usually focuses on technical details rather than higher level strategic issues outlined in the previous section, it is common for a draft problem formulation to end after discussion of the general implementation strategy, with the SAP added later if there is general agreement that the ERA should proceed. In that case, the SAP may be added to the problem formulation prior to finalization or may be developed as a stand-alone document.

> ***Key Concept:***.
>
> The appropriate level of detail in a sampling and analysis plan varies depending on the scale and complexity of the ERA, as well as expectations of stakeholders.

The scope of a SAP will vary depending on the complexity of the ERA and the level of detail that has already been specified earlier in the problem formulation. The SAP may address all of the planned sampling and analysis details for the entire ERA, or it may be limited to plans for the first phase (or tier) of the ERA. In cases where no further field sampling is needed, the SAP will relate to analysis only.

Importantly, the SAP must demonstrate that it is fulfilling the information needs for each LOE that will be used in the ERA. The recommended way to cross-check the completeness of the SAP is to establish a checklist to make sure that the field / data requirements of each LOE are met. Because field programs are normally implemented at discrete times, there can be significant implications of accidental omissions related to field data collection. A checklist template for these purposes is provided in **Table 2-10**. While the primary benefit of such a table is for the risk assessor (i.e., to ensure the SAP is complete), it is recommended to include it in the SAP submission to demonstrate to reviewers that a cross-checking process was undertaken.

The rest of this section expands on some of the SAP requirements listed in **Table 2-10**. Prior to developing a SAP, practitioners should consult current CCME (2012) guidance on sampling contaminated sites.

Field Safety Plan – A field safety plan is important for every project involving field work. Whether it is part of the SAP or handled separately is not important, but the SAP should at least confirm that the plan is or will be in place.

Table 2-10. Example checklist for a sampling and analysis plan.

	Field Component		
	Soil Chemistry	Invertebrate Bioassay	Soil Invertebrate Community
A. Planning Checklist			
Field safety plan established?			
Logistics			
Permits and site access permissions			
Transportation and access			
Availability of major sampling gear			
Seasonality appropriate for data?			
Sampling			
Core parameters (e.g., COCs) included?			
Ancillary / supporting parameters included?			
On-site sampling locations selected?			
Reference sampling locations selected?			
Detailed field sampling methods established?			
Sample handling methods specified?			
Field QA/QC methods and objectives established?			
Laboratory Analyses			
Lab methods specified?			
Lab detection methods specified and adequate?			
Lab QA/QC methods and objectives established?			
Data Analyses and Modeling			
Data expected to be adequate to support all analyses?			

B. LOE Requirements Checklist					**LOE Data Needs Met?**
LOE 1a - compare bioassay results on-site vs reference		x			y/n
LOE 1b - regression of bioassay results vs soil chemistry	x	x			y/n
LOE 2a - compare abundance/diversity on-site vs reference			x		y/n
LOE 2b - regression of abundance/diversity vs soil chemistry	x		x		y/n

Logistics – An important but often overlooked component of a SAP is review of big-picture logistical considerations. Logistical considerations may include:

- time required to get sampling permits;
- permission for site access;
- transportation and accessibility (particularly for remote sites);
- availability of key sampling equipment;
- seasonal considerations for biological sampling (e.g., sampling for berries, mushrooms, kelp, eelgrass, leaves); and
- tide cycles (e.g., intertidal work may require a very low tide that occurs over multiple days during daylight hours).

Chemistry Sampling – Whenever chemistry samples are included as part of a SAP (water, soil, sediment, tissues), the SAP should specify, at the least, the following:

- Relevant COCs for each media;
- The form(s) of each COC to be measured in each medium;
- The list of supporting or 'conventional' parameters to be measured. The list of conventional parameters will vary by media type. The risk assessor should not assume that the list of conventional parameters collected during site investigation will suffice for ERA. Conventional parameters should include relevant indicators of potential bioavailability of COCs, which vary depending on the COC and media;
- Sampling locations and replicates;
- Sample collection methods (equipment, depth of samples, processing, volumes, jars to be used, etc.);
- Critical aspects of sample handling (filtration, storage, holding times, etc.);
- Expected lab methods including preparation (e.g., dry weight or wet weight) and reporting units; and
- Expected lab detection limits.

Typical tools for chemistry sampling (by media), typical ancillary data, cautions to be exercised in field sampling, and selected guidance documents related to chemistry sampling are identified in **Section 3** (see **Table 3-1**).

Biological and Other Sampling – As with chemistry sampling, for each other type of field sampling there should be a description of what will be sampled, how it will be sampled, and how the laboratory (if applicable) will conduct analysis of the samples. Examples of field sampling details might include mesh size for sieving benthic invertebrates; quadrat

size for evaluations of vegetative cover; or design specifications and bait for small mammal traps. Examples of lab methods to be specified might be taxonomic resolution for measures of invertebrate density; or plans for salinity adjustments for bioassays conducted using groundwater adjacent to the marine environment.

Typical tools for biological sampling, typical ancillary data, cautions to be exercised in field sampling, and selected guidance documents related to chemistry sampling are identified in **Section 3** (see **Table 3-1**).

Quality Assurance / Quality Control – QA/QC methods and expectations should always be specified in advance of sampling, so that the quality of data is ensured to the extent possible. If data quality objectives cannot be met, it may be necessary to revisit the selection of measurement endpoints and LOEs.

Specific QA/QC mechanisms typically associated with collection of environmental chemistry programs include:

- Prevention of contamination during field sampling (e.g., use of clean jars).
- Decontamination procedures between sampling stations to prevent cross-contamination, and potential use of cross-contamination swipes of sampling gear to test for cross-contamination.
- Field homogenization procedures (e.g., for bulk soil samples), and potential use of field duplicate or triplicate samples to evaluate the effectiveness of homogenization.
- Sample storage, transport and chain-of-custody, and potential use of lab travel blanks as a QA/QC mechanism.
- Lab replicates to test for measurement error (as relative percent difference).
- Other lab QA/QC mechanisms including analytical methods blanks, certified reference materials, and matrix spikes.
- Method detection limits relative to screening guidelines and relevant to usage in ERA.

In the case of toxicity labs, negative and positive controls, replication, instrument calibration, and other QA/QC mechanisms are used (see **Section 4** for more information on toxicity testing). Labs implementing invertebrate enumeration may use re-sorts and sample splitting as QA/QC mechanisms. More detailed consideration of QA/QC procedures is provided in CCME (2012).

Data Analysis and Modeling – Data analysis plans may or may not be included in a SAP, depending on the scale and complexity of the ERA and expectations of regulators, site custodians or stakeholders for a particular site. The formulation of lines of evidence describes how data will be analyzed, but there may be cases where additional details are

warranted. For example, for a complex site where a food chain model will be used to estimate wildlife exposures, it may be appropriate to outline the key aspects of the model design and assumptions.

2.3.7. Communication and Review

Prior to embarking on field data collection and analysis, it is important to try to seek review prior to implementation. Reviewers might include site custodians, FCSAP expert support, other regulators, other affected stakeholders, or peers. For complex sites, it may be important to develop communication tools (e.g., figures) that reduce the complexity in the problem formulation to something that is understandable by readers without technical expertise in ERA.

2.4. Uncertainties and Data Gaps in Problem Formulation

Uncertainties are pervasive in ERA. A focused discussion on the key uncertainties in a problem formulation is worthwhile. One major benefit of explicit discussion of uncertainties is that it may lead to identification of specific data gaps that could be addressed as part of, or parallel to the ERA. Some of the sources of uncertainties and data gaps common at the problem formulation stage include:

- **COCs** – There may be uncertainty about the list of COCs relevant to the site, which may be associated with incomplete history for the site, uncertainty about potential off-site sources of COCs, or simply chance that site investigation failed to detect a COC that is actually present at elevated concentrations. There may also be uncertainty about characteristics of the COCs related to fate, transport, and effects,
- **Transport pathways** – The CSM assumes that all relevant fate and transport pathways have been characterized. However, even well-designed site investigation work may fail to detect key pathways. For example, movement of contaminants through groundwater to the marine receiving environment may only occur under certain narrow tidal and seasonal windows; in such a case, the risk assessor will remain 'ignorant' and the CSM will not fully capture all relevant pathways.
- **Receptors of concern** – There is always uncertainty about selection of ROCs during the problem formulation. Usually the major receptor types are captured, and uncertainties are associated with the selection of surrogate species. For example, wildlife biologists may fail to recognize habitat for certain ROCs, so they may be prematurely excluded from the ERA. Alternatively, effects literature

may be limited (this is usually the case), so the surrogate species selected to represent a given receptor type may not be the most sensitive species.

- **Measurement endpoints** – Measurement endpoints are imperfect, either because of uncertainty in the measurements themselves (e.g., variability reduces the power to detect differences), or because of uncertainty about how the measurement endpoints translate into effects on assessment endpoints (e.g., what does a reduction in invertebrate growth mean at the population or community level?). While there is no particular data gap to be addressed, the risk assessor should acknowledge these uncertainties up front.
- **Site investigation** – If there are particular data gaps in site investigation that are identified during the problem formulation that could affect any aspect of the problem formulation, these should be identified and brought to the attention of site custodians and site investigators. In some cases, substantial data gaps (e.g., lack of surface soil data for a large portion of a site) may warrant a delay in finalization of the problem formulation until the data are collected.

The problem formulation will always be based on uncertain information. The risk assessor should identify the key uncertainties, specify which data gaps are most critical, and specify the assumptions made in moving forward with implementation of the ERA.

Key Concept:

There are several types of uncertainty in ERA, such as:

- Natural variability that cannot be 'reduced' (e.g., variability in COC concentrations across a site, spatial variability in the distribution of biota). Natural variability can be acknowledged, characterized and incorporated into an ERA (e.g., using probabilistic methods).
- Random measurement error associated with estimation of a parameter; such as may result from limitations in the number of observations and/or imprecision in the measurement techniques. Estimates of most parameters in an ERA are imprecise – examples include average soil concentration on a site (i.e., statistical estimation error due to limited sample sizes and lab analytical error), or average dose rates used in a food chain model (i.e., due to imprecision about all of the input parameters related to ingestion rates, COC concentrations in dietary items, etc.). The precision of estimates of these parameters can be improved through increasing sample sizes.
- Systematic measurement error (i.e., bias) resulting from inaccurate estimation or analytical techniques. For example, a mark-recapture program to estimate the abundance of a fish population may systematically underestimate the true abundance if a subset of the fish are not susceptible to the fishing gear. In some cases biases may be known and can be adjusted for, but in other cases they may be unknown.
- Structural or model uncertainty that reflects our limited understanding of the mechanisms driving risks. For example, we may fail to understand how an exposure pathway works, and therefore our empirical or mechanistic models may not reflect reality very well. Structural uncertainty can be addressed in part through the use of alternative or flexible model forms. Even where underlying processes are well known, models are deliberately developed to be simplifications of reality.
- Ignorance reflecting our failure to recognize mechanisms driving risks. For example, we may fail to recognize a relevant exposure pathway completely. True ignorance is, by its definition, unknown, and will not be captured in conceptual site models or in quantitative models used to estimate risks.

For more detail on types of uncertainties see Morgan and Henrion (1990) and Finkel (1990).

3. EXPOSURE ASSESSMENT

3.1. Introduction

3.1.1. Purpose

> ***Key Concept:***
>
> Exposure information is an input for every line of evidence in an ERA.

The general purpose of exposure assessment is to characterize the mechanisms by which receptors are exposed to COCs, and to quantify or categorize the magnitude of those exposures. Exposure and effects are matched together in one or more ways for every line of evidence (LOE) that is evaluated in an ERA. Consequently, exposure assessment is not a single step in ERA, but is carried out for every LOE. In many cases, the same exposure information is used in multiple LOEs (e.g., COC concentrations are often the exposure information that is matched to several different types of effects information). Importantly, while the details of exposure assessment are discussed in this section, they must be fully understood and articulated at the problem formulation stage in order to support design and planning of the ERA.

3.1.2. Overview of Exposure Assessment

Exposure assessment used to support any particular line of evidence generally entails the following elements (all of which need to be fully contemplated during problem formulation):

- Determine which ***type(s) of exposure measure*** will be used, among the following four broad types:
    1. *External exposure media* to which a receptor is exposed, such as surface water, porewater, sediment, soil, or food items. For example, soil invertebrates are expected to be exposed to COCs in soil. In some cases where external exposure media are the measure of exposure, an ERA can rely on site investigation data without additional data collection; however, in other cases it may be preferable to have concurrent exposure data that can be more precisely matched to effects data.
    2. *Internal exposure media* are tissues where contaminant concentrations are measured to represent exposure within the receptor itself. For example, fish tissue mercury concentrations can be used as an indicator of the exposure of

fish to mercury. In general, internal exposure media are more relevant than external exposure media for COCs that bioaccumulate or biomagnify up the food chain, and can be used whenever matching effects data are available to which the data can be compared.

> ***Key Concept:***
>
> Bioaccumulation occurs when the concentration of a COC in an organism is higher than the concentration in the surrounding environment.
>
> Biomagnification is an increase in the concentration of a COC from one trophic level to the next.

3. *Estimation of total dose*. For example, a small mammal may be exposed to COCs through ingestion of surface water, ingestion of ground insects or other food sources, and incidental ingestion of soil. Typically, estimation of total dose is implemented with a food chain model.

> ***Key Concept:***
>
> Food chain models are a series of equations used to estimate total dose to a receptor exposed to COCs via the food chain. Simple to moderately complex models can be formulated efficiently in a spreadsheet.

4. *Categorical measures of exposure* which do not explicitly rely on any information about contaminant concentrations, but instead categorize exposure in a simple manner. Common examples of categorical exposure measures are:

   - On-site versus reference condition
   - Site versus lab control
   - Spatial gradient categories such as near field, mid-field and far-field.

> ***Key Concept:***
>
> Categorical measures of exposure such as on-site versus reference are commonly used implicitly without additional data collection.

   Categorical measures of exposure are often used implicitly, but risk assessors should be explicit about their use for any LOE that depends on the categorical measure. For example, if bird densities are used as an effects measure comparing on-site and reference conditions, the implicit assumption is that exposure on-site is different from the reference condition. There may be information on COC concentrations in some media (e.g., soil) but perhaps not for other media (e.g., food item tissues). In that case, exposure for the LOE

may not be characterized as COC concentrations, but rather using the implicit categories for on-site and reference.

- Determine whether the exposure data will be ***directly measured or estimated***. Usually, concentrations of COCs in abiotic media (e.g., soil, sediment, water) are measured directly, but in some cases they are estimated (e.g., using fate and transport models). Concentrations of COCs in biotic media (i.e., tissues) are more often estimated (e.g., predicted using uptake factors), but estimation methods are uncertain so preference is for direct measurement whenever possible.

> ***Definitions:***
>
> Biotic media are biological tissues where COCCOCs may be found, whereas abiotic media are any other environmental media (e.g., soil, sediment, water, air).
>
> Uptake factors are the ratio of COC concentrations in tissue to the COC concentrations in an abiotic medium such as soil or water.

- Determine how the data will be packaged to represent exposures for various ROCs. For example, will maximum values be used or will some kind of statistical metric of the data be used to represent exposures (e.g., 95% UCLM).
- Determine what ***ancillary data*** will be collected in addition to COC concentrations, including data related to evaluation of bioavailability.
- Characterize ***uncertainties*** in exposure, evaluate the implications of uncertainty using sensitivity analysis, and, if warranted, integrate uncertainties into the exposure assessment using probabilistic methods.

The outcome of exposure assessment is information that can be matched with effects measures to provide evidence in the form of a LOE. It is critical that the risk assessor conceptualize the exposure and effects information at the same time (during problem formulation) to ensure that they can be integrated effectively and to ensure that all information and ancillary data needs are identified prior to data collection.

Section 3.2 compares direct measurement and estimation, which is an issue that applies to all of the types of exposure measures with the exception of categorical measures. **Section 3.3** explores the four types of exposure measures in detail, focusing on how data will be used to represent exposure for ROCs; and what ancillary data will be collected in addition to COC concentrations. **Section 3.4** discusses options for moving beyond typical point estimates of exposure.

3.2. Direct Measurement versus Estimation

Risk assessors must not only decide what type of exposure measure(s) are appropriate for a given line of evidence, but whether to measure or estimate exposure in each case. This section provides guidance in this regard, for abiotic media and for tissues.

> ***Key Concept:***
>
> Direct measurement of COC concentrations in any medium is preferred over estimation, particularly for detailed ERAs, because of the much lower uncertainty associated with direct measurement. However, there are cases where estimation may be suitable or may be the only feasible option.

3.2.1. Direct Measurement versus Estimation for Abiotic Media

Whenever abiotic media such as soil, water, and sediment are used as measures of exposure, data on COC concentrations must be either measured or estimated. Direct measurement is most common for abiotic media, and is generally preferred for detailed ERAs because:

- There is much less uncertainty regarding measured COC concentrations compared to estimated concentrations;
- Many informative ancillary variables cannot be practically predicted and must be measured (e.g., pH, SEM:AVS);
- There are usually significant data available for soil and other media as a result of site investigations; and
- The cost of collecting additional chemistry data in abiotic media is generally not prohibitive.

However, there are cases where measurement is not possible, not practical or not necessary, and estimation of COC concentrations is preferred. This may occur, for example, when:

- An ERA is evaluating a future scenario under which current measured values are not directly relevant; or
- Chemistry data cannot be collected safely (e.g., sediment in a river with difficult access).

- It is anticipated that the characterization of risks and/or decisions regarding risk management would be unaffected by the added accuracy provided by direct measurement.

COC concentrations in abiotic media are estimated using simple or complex models that predict the fate and transport of contaminants in the environment. A simple model would be one that does not predict transport of COCs but simply predicts concentrations in one medium from concentrations in another medium based on chemical properties. For example, the partitioning of organic compounds from water into the organic matter of sediments can be predicted based on the octanol-water partitioning coefficient (K_{OW}).

More complex models take into account the complex interactions of contaminant loadings, movement and partitioning into various media (Cowan *et al.* 1995). An example of a complex fate and transport model would be one that predicts contaminant concentrations in a section of river based on information about loadings and water flows in upstream tributaries. Development of fate and transport models can be expensive, and their relative advantages and disadvantages should be carefully considered.

3.2.2. Direct Measurement versus Estimation for Tissues

Whenever tissues are used as either a measure of internal exposure or as a food item for a higher trophic level receptor, the tissue concentrations of COCs can be either measured directly or estimated. Direct measurement is relatively common for some tissue types such as plants, invertebrates and fish, but less common for other tissue types such as mammals and birds. Whenever feasible, direct measurement is usually preferred over estimation because there is much less uncertainty regarding measured COC concentrations in tissues. However, estimation may be appropriate in some cases, including:

> ***Key Concept:***
>
> Direct measurement of COCs in biological tissues is usually preferred over estimation. Most tissues in food items of birds and mammals (e.g., plants, invertebrates, fish, small mammals) can usually be collected with moderate effort.

- When time constraints for the ERA preclude waiting for seasonal tissues (e.g., berries, tree leaves, bird eggs).
- For organisms or sites for which it is considered inappropriate to sacrifice individual organisms for purposes of obtaining data;

> ***Key Concept:***
>
> Tissue data can be used in two ways – to represent internal dose to a receptor (i.e., body burden), or to characterize COC concentrations in a food item ingested by a higher trophic level receptor.

- When an ERA is evaluating a future scenario under which current measured values are not directly relevant; or
- For purposes of generating initial risk estimates on a limited budget.

Importantly, it may be efficient to use a combination of measurement and estimation for tissues at large sites. Specifically, if a relationship can be established between, for example, soil and tissue concentrations, that relationship could then be extrapolated to other samples where only soil is available.

For cases where COC concentrations in tissues are estimated, at least three methods are available, each with advantages and disadvantages as follows:

1. Uptake factors – the ratio of the contaminant concentration in tissue to the concentration in an associated abiotic medium (e.g., water, soil or sediment). Uptake factors based on water are commonly referred to as bioconcentration factors (BCFs) or bioaccumulation factors (BAFs)[28]. Uptake factors are generally very uncertain and they should be avoided if bioaccumulation regression models (below) are available. Published uptake factors are available for a range of contaminants and tissue types (Sample *et al.* 1998, Suter *et al.* 2000, and references therein), but these should be viewed as examples only – risk assessors should seek out the most recent scientific literature as part of any detailed ERA and determine which uptake factors are applicable to a given site. Importantly, the units used in uptake factors (e.g., wet weight, dry weight, lipid normalized) must be the same as the units for the site-specific data, or must be converted to be equivalent.

   Key Concept:

   Uptake factors are simply ratios. For example:

   $$UF = \frac{Concentration_{tissues}}{Concentration_{soil}}$$

2. Bioaccumulation regression models – these models are superior to simple uptake factors for two reasons. First, they allow for inclusion of variables other than contaminant concentrations (i.e., using multiple regression approaches), which ultimately are capable of explaining more of the variation in the tissue data. Second, regression models are capable of accounting for nonlinearity in the relationships between soil and tissue concentrations. Nevertheless, uncertainties in regression models are typically high. As with simple uptake factors, summaries of bioaccumulation regression models are available for a range of contaminants and

[28] In strict terms, BAF is intended to apply to the ratio between tissue and exposure medium (e.g., water) where all exposure pathways are considered simultaneously, whereas BCF is intended to refer to an exposure condition that includes water only. In common usage, however, the term BCF is often used to refer to the quantity that is more correctly described as BAF.

tissue types (Sample *et al.* 1998, Suter *et al.* 2000, and references therein), but again these should be viewed as examples only – risk assessors should seek out the most recent scientific literature as part of any detailed ERA and determine which models are applicable to a given site.

3. Mechanistic bioaccumulation models – these models are based on details of the physiology of the organism (e.g., metabolic transformation) and the behaviour of the contaminant (e.g., solubility and partition coefficients). Mechanistic models are data-intensive and complex, and therefore can rarely be developed on a site-specific basis[29]. Moreover, such models may suffer from larger uncertainties than simple empirical models, due to cumulative uncertainties in modeling several mechanistic processes which may be poorly understood.

In practice, many exposure assessments may use a combination of measured and estimated tissue concentrations simultaneously. For example, when exposure is based on evaluation of total dose, some food item tissue concentrations may be measured while others may be estimated.

3.3. Types of Exposure Measures

The key decision in exposure assessment is determining what type of exposure measure to use for a particular line of evidence in an ERA.

This section distinguishes four broad types of exposure measures:

- External exposure media (e.g., surface water, porewater, sediment, soil, food item tissue) to which a receptor is exposed;
- Internal exposure media that describe contaminants within the receptor itself;
- Estimation of total dose (i.e., estimation of dietary intake through food chain modeling); and
- Categorical measures of exposure (e.g., on-site versus reference condition).

The decision about what type of exposure measure to use should be based on:

- The level of effort to collect the data, balanced against the need for precise information – for example, collection of tissue data for input into a food chain model may not be warranted until potential risks are first evaluated using conservative measures that require less effort to collect.
- Availability of matched effects data against which the exposure tool outputs can be compared. For example, measures of contaminants in external exposure media

[29] A few examples of mechanistic bioaccumulation models are referenced by Suter et al. (2000).

such as soil can be compared to benchmark concentrations associated with effects on plants or invertebrates. Measures of contaminants in internal exposure media such as fish tissues can be compared to Critical Body Residues (CBRs).

Importantly, a single exposure measure may be used in several LOEs. The manner in which exposure measures are used should be defined up front in the problem formulation, as there is little value in identifying a measure without also clarifying how it will be used.

3.3.1. External Exposure Media

External exposure media can include any medium to which a receptor is exposed. An example of an external exposure medium would be soil for terrestrial invertebrates. External exposure media include not only abiotic media such as soil, water, sediment and air, but also food item tissues. In the case of strongly bioaccumulative and biomagnifying substances, tissues are usually the most relevant external exposure medium for higher trophic level receptors due to the high proportion contributed to total dose.

This section focuses on how external exposure data are used to represent exposure for ROCs, and what ancillary data will be collected in addition to COC concentrations. An overview of typical methods of evaluating each external exposure medium, common ancillary parameters, and key challenges are summarized in **Table 3-1**.

Field methods of collecting soil, surface water, groundwater, sediment, porewater and tissue data are not reviewed in this section, as these methods are addressed in detail elsewhere by CCME (2012), Environment Canada (2011), Mudroch and MacKnight (1994), USEPA (2007d) and PSAMP (1996).

3.3.1.1. Soil

Soil contaminant concentrations are very commonly used as a measure of external exposure media, in particular for characterizing exposure for plants and soil invertebrates, but also for characterizing some exposure pathways for wildlife. Soil data that are typically collected as part of site investigation are rarely completely adequate for risk assessment unless risk assessors have been involved up front during site investigation. Where soil data do not meet the needs of exposure assessment as defined in this section, supplemental data collection may be warranted, particularly for detailed assessments.

Table 3-1. Sampling considerations for external exposure media

Exposure Medium	Typical Tools for Measurement / Estimation	Typical Ancillary Parameters	Cautions / Key Issues	Guidance
Soil	• Measurement of bulk soil chemistry, based on collection by trowel or auger	• Site-specific, but may include organic matter content, pH, moisture content, soil texture, cation exchange capacity.	• Sample depth • Differentiation of soil layers • Spatial design and resolution	• Health Canada (2007) • Suter *et al.* (2000) • CCME (2012)
Surface water	• Measurement of total or dissolved concentrations using typical water sampling gear	• Site-specific, but may include hardness, pH, alkalinity, acidity, temperature, dissolved oxygen, anions, cations, nutrients, conductivity, salinity, TSS, DOC	• Temporal variability including seasonality	• Environment Canada (2011) • CCME (2012) • Suter *et al.* (2000) • PSAMP (1996) • Paquin *et al.* (2003)
Sediment and sediment porewater	• Measurement of bulk sediment chemistry using grabs, divers, or cores • Measurement of sediment porewater chemistry (dissolved) by extraction from sediments, or directly (e.g., using push-point samplers)	• For sediment: organic carbon, particle size, pH, sulphides, SEM:AVS, possibly iron and manganese hydroxides • For porewater: redox, plus similar parameters to surface water	• Understanding relevance of bulk sediment versus porewater for each receptor type • Sample depth • Spatial design and resolution • Understanding and addressing oxygenation of porewater samples during collection and transport	• Environment Canada (2011) • CCME (2012) • Mudroch and MacKnight (1994) • Suter *et al.* (2000) • PSAMP (1996)
Air / Vapour	• Rare for ERA, but direct measurement and modeling are both used – see text.			
Tissues	• Direct measurement preferred • Estimation using uptake factors or models	• Lipid content • Moisture content	• Consideration of whether to test whole organism or selected tissue types • Consideration of whether to depurate, depending on how tissue data will be used.	• Suter *et al.* (2000) • CCME (2012) • Beyer and Meador (2011) • PSAMP (1996)

Defining surface soil – As a default, and consistent with the default approach used during COC selection, all soil data in the top 1.5 m can be considered as surface soils for purposes of measuring exposure for plants and soil invertebrates, as well as for higher-level receptors. However, for cases where more precision is warranted, the depth of surface soils should be defined on a site-specific basis, taking into account:

> ***Key Concept:***
>
> As a default, all soil data in the top 1.5 m can be considered as surface soils for purposes of measuring exposure, but site-specific (and receptor-specific) depths should be defined when precision in soil exposure estimates is warranted.

- The depth of bioturbation due, for example, to burrowing insects, burrowing mammals, and plant root systems. For deep-rooting plants and trees, it may be necessary to consider exposure to COCs at soil depths greater than 1.5 m, whereas for insects the depth may be much shallower.
- Applicable policy at provincial or other levels (for sites intended for divestiture) that require that a different depth be considered for surface soil exposure.
- Differentiation of soil layers for certain receptor groups – for example, some receptors may be limited to the humic layer rather than underlying mineral soil (e.g., organisms that play a role in decomposition of organic matter). If there are large differences in COC concentrations between these 2 layers (e.g., airborne mercury typically accumulates in the humic layer), it may be appropriate to use soil data from only the humic layer.
- The depth that is likely to have been affected given the sources and nature of contamination at the site. For example, for an air deposition source there may only be a shallow surface layer (e.g., top 2-5 cm) that is contaminated and that should be used exclusively for understanding particular exposure pathways such as incidental ingestion.
- Natural processes or planned activities at the site that will result in accumulation of soils, or removal of surface soils that will expose soil at depths In such cases, the soil layer relevant for current exposure for a particular receptor may not be the same as the soil layer that is relevant for future exposure.

Using soil data as an exposure measure – The key question for the risk assessor in using soil data is whether to measure exposure to soil based on single soil samples or using statistical measures, both horizontally and vertically. For plants and soil

> ***Key Concept:***
>
> Soil data representing external exposure can be characterized using the maximum concentration, the mean, an upper confidence limit on the mean, or a selected percentile, depending on the quantity of samples, receptor characteristics, and the degree of conservatism appropriate for the ERA.

invertebrates, the default for spatial characterization should be to measure exposure on a sample by sample basis and to consider use of summary statistics for each AEC (e.g., 95% UCLM and 90th percentile) in cases where there are sufficient sample sizes in each AEC (e.g., > 10). Vertically, the soil data that are used (either sample by sample, or with summary statistics) must be only the soil data that are relevant for a particular receptor group. There is no point considering deep soil data for shallow-rooting plants.

While plants and immobile soil invertebrates will be affected locally by elevated COC concentrations at a single soil sample location, the spatial scale at which potential major risk management measures would be implemented is also relevant. In other words, exposure (and risks based on a given LOE) for plants and soil invertebrates should be understood at more than one spatial scale, because the spatial scale is an important component of the magnitude of risk estimates (see **Section 5** for further discussion).

For mammals, birds and other wildlife receptors exposed to soil, incidental soil ingestion can sometimes be the most important exposure pathway. Unless acute effects are expected from exposure to high COC concentrations in a particular sampling location, exposure for these ROCs should always be based on summary statistics such as the arithmetic mean, the 95% UCLM or the 90th percentile, unless the number of soil samples is small (e.g., < 10) relative to the home range size of the organism. Interpretation of summary statistics should take into account current CCME (2012) guidance on sampling for contaminated sites. The risk assessor should also determine whether the spatial layout and density of soil samples collected during site investigation or other evaluations is adequate to support assessment of risks for each surrogate ROC.

Ancillary parameters – Ancillary parameters that are often relevant for soil include:

- Percent organic matter – organic matter is important for organic compounds that partition predominantly into lipids (i.e., have a high octanol-water partition coefficient) – in such cases, soil concentrations of COCs may be more appropriately characterized using organic carbon-normalized concentrations.
- pH – pH data are important for understanding general soil conditions, including the likely solubility, speciation and complexation of metals. At extreme pH, the data can be useful in predicting plant stress as well as presence / absence of biota.
- Moisture content – moisture data are important if soil data will be used in food chain models, because data related to incidental soil ingestion rates may be based on dry- or wet-weight concentration units.
- Cation exchange capacity (CEC) – is the maximum quantity of total cations that soil can hold. Clay and humus typically have higher CEC than sandy soils. This property can be useful in determining the relative bioavailability of metals, because lower CEC soils are more likely to release metals to biota.

- Redox potential (Eh) – an electrical measurement characterizing the transfer of electrons in soils to or from a reference electrode. Eh can be used to estimate if soil is anaerobic (low Eh) or aerobic (high Eh), which can affect the dissolution or precipitation of various metals.
- Soil texture/composition – texture (relative proportions of sand, silt and clay) and structure (aggregation of soil particles into larger secondary clusters; typically developed through the action of microbes or invertebrates) can affect contaminant dynamics in soils.

Some helpful resources include: for soils (Brady and Weil 2008, Miller and Gardiner 2003) and for soil contamination (Allen 2002, Mirsal 2009, Perzynski *et al.* 2005, Harrison 2001).

Evaluating bioavailability – Although ERAs may use soil data collected during site investigation to initially characterize contamination in soil (i.e., whole soil samples and analysis of bulk soil chemistry), where precise understanding of risks is warranted it can be appropriate to consider speciation analyses or other methods that will more precisely characterize the form of contaminants in soil. In addition, extraction techniques may be considered for characterizing the fractions that are more likely to be bioavailable (Suter *et al.* 2000; Allen 2002). Alternatively, if soil porewater is considered to be the relevant exposure medium (e.g., for plant roots), then soil porewater can be either measured or estimated from bulk soil chemistry using equilibrium partitioning models. Further discussion on these approaches is found in Suter *et al.* (2000) and in Allen (2002). Finally, studies that simulate bioavailability in the human gastrointestinal tract (referred to as bioaccessibility tests or physiologically-based extraction tests [PBET]) are now used in human health risk assessment (see early work by Ruby *et al.* 1996), and are starting to be more prevalent in ERA as well. Results from methods intended to simulate the human gut may be directly relevant to mammals that have similar anatomy and gut conditions (e.g., pH) to those of humans; alternatively, specific test protocols may be modified for other species.

3.3.1.2. Surface Water

Surface water exposures occur through direct contact (e.g., for aquatic plants, fish, or benthic epifauna) or by ingestion (e.g., for wildlife).

Defining surface water – For purposes of ERA, surface water is water that is above the sediment / water interface in any aquatic system, and can also include temporary pools or watercourses that provide aquatic habitat or drinking water. Surface water is distinguished from sediment porewater, which is water in the interstitial spaces within the sediment. Importantly, surface water rather than (or in addition to) sediment porewater may be a relevant external exposure medium for some organisms that live in the sediment

– for example, clams are buried in sediment but are exposed to surface water via their siphons which filter water directly from the sediment – surface water interface.

Using surface water data as an exposure measure – As with soil, a key challenge for the risk assessor is deciding whether to use maximum measured COC concentrations or some statistical metric over space or time, for each ROC. For sessile organisms (e.g., aquatic plants), maximum concentrations may be appropriate to represent concentrations in small areas, but statistical measures (e.g., 95% UCLM and 90th percentile) can also be used to characterize average exposures in particular areas. For mobile receptors, maximum concentrations are recommended as a default if there are few (e.g., < 10) samples in the area covered by their home range size, or if seasonal variability in COC concentrations is expected but has not been measured. In cases where sample sizes are large and seasonal variability is captured (if warranted), summary statistics can be used (e.g., 95% UCLM and 90th percentile). Importantly, where surface water data are used to represent drinking water exposure for wildlife, the risk assessor should consider the number of nearby options for drinking water, and the proportion of total exposure that is likely to come from any one source. In such cases, statistical measures of surface water COC concentrations may be based on averaging across the sources rather than averaging across pooled samples (i.e., if one drinking water source has 3 samples and another has 20 samples, a 95% UCLM based on the pooled samples will bias towards the second source).

> ***Key Concept:***
>
> Surface water exposure assessment may require use of total concentrations of contaminants, or dissolved concentrations, or both.

A second issue that must be considered by the risk assessor is whether to use dissolved concentrations, total concentrations, or both as the measure of exposure. This decision may be affected in part by regulatory requirements, but should also take into account relevance for ERA. If the surface water data will be used for more than one purpose (e.g., as external exposure media for fish, as well as drinking water for wildlife), the data should be appropriate for all purposes. Total concentrations are most relevant for ingestion pathways, whereas dissolved concentrations (see the following Bioavailability discussion for more information) are more relevant for direct contact pathways. As a default, risk assessors should use total concentrations in water as it is more conservative, but can rely on dissolved concentrations provided that rationale is given. In either case, it is important to make sure that the exposure data will be comparable to available effects data (e.g., exposure data based on dissolved concentrations should not be compared to effects data based on total concentrations).

Finally, in most cases the specific form of dissolved contaminants does not need to be quantified for ERA. However, for some sites where potential risk management costs are high and the relative toxicity of a COC is highly dependent on its form, the increased cost of speciation analysis may be worthwhile. For example, different forms of iron in

sediment porewater may differ in toxicity by several-fold, and efforts to link toxic responses in porewater bioassays to a potential causal effect of iron may require understanding of the relative concentration of each iron species in the porewater samples.

Bioavailability – Dissolved contaminants in water are not necessarily bioavailable. For example, research over the last two decades regarding metals bioavailability and mechanisms of toxicity in the aquatic environment, has led to development of the biotic ligand model (BLM; Di Toro *et al.*, 2001; Paquin *et al.*, 2003, see **Section 4.2.2** for more details). This model accounts for the roles of total suspended solids, pH, dissolved organic carbon, cations (Ca, Mg, Na, K), anions (SO_4, Cl), alkalinity, hardness, and sulfide in determining free metal ion concentrations in affecting metals bioavailability (and ultimately toxicity) in freshwater.. Many of the typical ancillary parameters listed below are used to support understanding of potential bioavailability including through use of the BLM.

Ancillary parameters – Typical ancillary parameters that are measured in surface water depending on the site and the ERA include the following:

- Hardness;
- pH (pH may also be a COC);
- Alkalinity;
- Acidity;
- Temperature;
- Dissolved oxygen;
- Anions and nutrients (e.g., chloride, bromide, fluoride, nitrite, nitrate, sulphide, sulphate, etc.);
- Cations (e.g., Ca, Mg, Na, K);
- Conductivity;
- Salinity (for sites at the interface of freshwater and marine);
- Total suspended solids; and
- Dissolved organic carbon.

This is not an exhaustive list – in general, any parameter that is expected to provide useful information for reasonable cost should be considered. Some parameters such as pH can be measured in both the lab and the field – generally lab equipment will be more reliable, but field measurements should also be taken, especially if there is an expectation of changes during storage and transport (e.g., temperature), or if the additional cost of field data collection is minimal (e.g., for pH or dissolved oxygen).

3.3.1.3. Groundwater

Defining groundwater – For purposes of ERA, groundwater is any water that is not surface water and is not considered to be within the biologically active layer of surficial soil or sediment (i.e., porewater).

Using groundwater data as an exposure measure – Groundwater should generally not be considered as an exposure medium in ERA[30]. Although stygofauna (small, aquatic organisms that live within groundwater systems, such as caves and aquifers) may be directly exposed to groundwater, their adoption as assessment endpoints is rare.

> ***Key Concept:***
>
> Groundwater is rarely an appropriate medium for characterizing current exposure conditions in ERA, but may be used as a surrogate for porewater or surface water in certain cases.

Groundwater may be applied as a surrogate for exposures to organisms that live in soil or sediment porewater or even surface water in certain cases such as:

- A preliminary or screening-level assessment based on existing data, where groundwater was collected during upland site investigations but soil or sediment porewater was not collected;
- A conservative assessment where groundwater chemistry is used to represent worst-case exposure; or
- For sites where soil or sediment porewater is very difficult to access (e.g., if a foreshore is covered by rip rap).
- For sites where a groundwater plume is migrating towards a surface water body but has not yet reached the surface water body. In this case, the groundwater may be considered somewhat representative of potential future discharges to surface water.

Ancillary parameters – Most of the same ancillary parameters for surface water should be measured in groundwater, with the addition of redox potential. A key consideration for groundwater assessment is that ancillary parameters such as redox potential and pH are likely to differ from those measured as the groundwater enters the transition zone where it interfaces with surface water. These ancillary parameters can have significant effects on the bioavailability of contaminants (e.g., metals that are dissolved in groundwater may precipitate out quickly as the water becomes oxygenated in the transition zone).

[30] See Environmental Canada (2010) federal interim groundwater quality guidelines for federal contaminated sites for exceptions and further discussion.

3.3.1.4. Sediment and Sediment Porewater

Defining sediment and porewater – For purposes of ERA, sediment is the substrate in an aquatic feature, and sediment porewater is the water found in the interstitial spaces of the sediment. Sediment and sediment porewater are the primary exposure media for benthic invertebrates, particularly benthic infauna, and also for many aquatic plants or algae. Direct exposure may also be relevant for the early life history stages for some higher-level organisms (e.g., fish eggs). Incidental ingestion of sediment is also an important exposure pathway for some higher level receptors such as bottom-dwelling fish and aquatic birds.

> ***Key Concept:***
>
> An ERA may characterize exposure using bulk sediment, sediment porewater, or both.

Defining surface sediment – As with soil, the depth of sediment that is relevant to ecological receptors should be carefully considered. As a default, and consistent with the default for COC selection, all sediment data in the top 1 m can be considered for purposes of measuring exposure. However, for cases where more precision is warranted, the depth of surface sediment should be defined on a site-specific basis, taking into account:

- The depth of bioturbation due to flora and fauna (e.g., worms, bivalves).
- Applicable policy at provincial or other levels (for sites intended for divestiture) that require that a different depth be considered for surface sediment exposure.
- Natural processes or planned activities at the site that will result in deposition, erosion or removal of surface sediments that will expose sediments at depths.

Using sediment and porewater data as exposure measures – Typically, bulk sediment is used as the initial indicator of external exposure – almost all environmental quality guidelines are based on bulk sediment and not on porewater, and so initial characterization of sediments focuses on that medium. However, porewater is often the medium in which contaminants are most likely to be biologically available, as opposed to the portion bound to particulate matter. Sediment and sediment porewater may be appropriate for use as external exposure media in ERA. Bulk sediment is recommended as the default external exposure medium for most cases because:

- Effects data are more commonly associated with sediment, so bulk sediment chemistry is more likely to contribute to lines of evidence for the ERA;
- Sediment concentrations are less likely to change on short time scales (e.g., tidal fluctuation) or even longer time scales (e.g., seasonality), with the exception of patterns of deposition and scouring; and

- Sediment sampling and analysis is relatively straightforward compared to porewater sampling and analysis, as the latter is influenced by specific extraction technique and sample handling and preservation methods.

However, porewater should be evaluated in many cases (usually in addition to bulk sediment), such as:

- When there is ongoing transport of COCs in dissolved phase to the aquatic environment via groundwater;
- Cases where COCs are likely to partition predominantly into water and not adsorb to sediments;
- When increased precision is desired in relating effects measures to bioavailable (dissolved) contaminant concentrations; and
- When effects measures that will be matched to the exposure data are based on porewater (e.g., porewater bioassays).

Risk assessors should not assume that bulk sediment alone is sufficient for any particular ERA.

Ancillary parameters and bioavailability – Ancillary parameters of importance for porewater are the same as those listed earlier for surface water and groundwater. Ancillary parameters of importance for bulk sediment typically include:

- Organic carbon content – organic carbon is the most important factor determining partitioning of organic compounds into sediments.
- Particle size (e.g., % clay, silt, sand and gravel) – because the ratio of surface area to sediment volume is higher for finer sediments, evaluation of patterns of COC concentrations in sediments can be confounded by the influence of particle size (i.e., COCs are more likely to be bound up in finer sediments).
- pH – In bulk sediment, pH is an indicator of the general environment and types of receptors that could be expected to be present.
- Sulphides – In anaerobic sediments, sulphides are normally the predominant binding phase, and measurement of SEM:AVS[31] can provide insights into the potential bioavailability of cadmium, copper, lead, nickel and zinc. The SEM:AVS model can also incorporate organic carbon (USEPA 1999).

[31] AVS is acid volatile sulfide, and SEM is simultaneously extracted metal. If SEM – AVS < 0 then it is assumed that sufficient sulfides are available to bind the SEM metals. The SEM:AVS model does not apply to oxygenated sediments. More discussion and caveats to use of SEM:AVS are found in Suter et al. (2000) and Paquin et al. (2003).

- Iron & manganese hydroxides (for metals) – In aerobic sediments, iron and manganese hydroxides can be an important binding phase. Sequential extraction techniques employ a series of chemical fractionation steps to elucidate the relative importance of various binding phases (e.g., Tessier *et al.* 1979).

3.3.1.5. Air / Vapour

Air is often *not* justified as an exposure medium in ERA. In many cases the contribution of airborne COCs to total exposure for wildlife would be negligible, in part because the volatile compounds which are most likely to be inhaled volatilize rapidly to air and are dispersed rapidly. While inhalation exposure has been shown to be unimportant for several contaminants (US EPA 2003), there are relatively few studies that have evaluated volatile organic compounds in detail. Nevertheless, toxicity data are available for several compounds and screening values for evaluating potential ecological risks via inhalation have been developed and applied (Archbold *et al.* 2007; Gallegos *et al.* 2007; Markwiese *et al.* 2008).

While air may be ruled out as an exposure medium in many ERAs, it should be considered in certain cases, such as:

- Where a site with wildlife receptors is characterized by very high concentrations of volatile organic compounds;
- Where a site with volatile organic compounds has wildlife receptors that burrow on the site;
- Where plant foliage is expected to accumulate certain contaminants (e.g., mercury, DDT) through uptake of vapours (Suter *et al.* 2000).

In such cases, as a starting point air can be sampled directly, including from existing burrows or artificial burrows (Markwiese *et al.* 2008), and compared to screening values such as those summarized or developed in existing literature (Archbold *et al.* 2007; Gallegos *et al.* 2007).

3.3.1.6. Tissues of Food Items

Analysis of contaminant concentrations in organism tissue is a relevant external exposure tool in cases where the tissues represent an important food item for an ROC. As mentioned in **Section 3.2.2**, direct measurement of tissues is preferred over estimation.

Defining food item tissues – Food item tissues include any diet items of a receptor, but do not include incidental ingestion of soil or sediment.

Key Concept:

Tissues of food items can be used in two different ways.

1. If there are effects data based on the COC concentration in a food item (e.g., CEQGs for tissue residue for protection of wildlife consumers of aquatic life), then the measured or estimated tissue concentration itself is the measure of exposure that is compared to the effects data. This use of the tissue data is as an external exposure medium.
2. In contrast, and more commonly, effects data are based on total dose. This is the case for most wildlife Toxicity Reference Values. In that case, the COC concentration in a tissue is one input to total dose (along with tissue data for other diet items, and all other relevant exposure pathways such as drinking water, incidental soil or sediment ingestion, etc.).

Using food item tissue as an exposure measure – The use of food item tissue as an external exposure measure is appropriate whenever there are matching effects data. For example, concentrations of biomagnifying substances (e.g., mercury, PCBs) in fish may be compared to CEQGs for tissue residue for protection of wildlife consumers of aquatic life (CCME 2001). In any case where food item tissues are used as an external exposure measure, the specific tissue type that is collected (e.g., muscle only or whole body) must match with the effects data to which the comparison will be made[32]. In cases where a whole organism body is used, the risk assessor should as a default not have the organism depurated (i.e., intestines voided) or washed, unless the whole organism effects data were known to be based on depurated organisms.

Ancillary parameters – Ancillary parameters that are generally important with tissue sampling are lipid content and moisture content. Lipid content is particularly important for contaminants that partition strongly into the lipid fraction (e.g., PCBs), because concentrations among tissues can only be meaningfully compared when lipid-normalized. Moisture content is important so that comparisons can be made to effects measures specified in either wet-weight or dry-weight terms.

[32] Unless there are models available that establish relationships between concentrations in various tissues and/or whole organism. If both a particular tissue type and whole body are relevant (e.g., for different purposes), the particular tissue may be submitted for analysis, as well as the remaining tissues, so that a whole body concentration can later be calculated if necessary (as a mass-weighted average).

3.3.2. Internal Exposure Media

Defining internal exposure media – Internal exposure refers to measures of contaminant concentrations within the receptor itself, which may include chemical concentrations in particular tissues where toxic effects occur (e.g., liver), or other tissues used as indicators of body burden (e.g., bone, hair or muscle tissue), or whole animals. Measures of internal exposure are commonly referred to as body burdens or residues.

Using internal (body burden) contaminant concentration as an exposure measure – Body burdens of COCs can be used as measures of internal exposure whenever there are readily available effects benchmarks to which the exposure data can be compared. To determine whether it is possible to use internal exposure measures in an ERA, the risk assessor should:

- Review information on the behaviour of the COCs in receptors to determine if internal exposure measures would be useful. Typically this information is summarized as part of the review of COC characteristics during problem formulation. Some COCs are not suitable for internal exposure analysis due to their behaviour or fate in receptors – for example, PAHs are metabolized by wildlife and therefore body burden of PAHs may not be a useful indicator of exposure.
- Review published studies that have derived effects thresholds based on body burdens. This requires a review of primary literature – some thresholds are compiled by Beyer and Meador. (2011) and Suter *et al.* (2000), but are not comprehensive.
- Review the environmental residue-effects database (ERED)[33] jointly compiled by the US Army Corps of Engineers and the US Environmental Protection Agency to determine if there are adequate data available from which effects thresholds could be derived (using methods described in **Section 4** and **Technical Module 2** appended to this document).
- Ensure that there are suitable (practical) methods of collecting the particular tissue type that would need to be matched to the effects data. If the effects data are based on whole body concentrations or common tissue types, there may be uptake factors or bioaccumulation models that would allow estimation rather than measurement of internal exposure. For measured data, the risk assessor should as

[33] The ERED database contains data from over 2,000 studies, and is available at http://el.erdc.usace.army.mil/ered/. This database is the most up-to-date comprehensive source of tissue residue effects levels. It should be supplemented with current primary literature surveys to support a particular ERA.

a default have the organism depurated (i.e., intestines voided) or washed to ensure comparability with available effects data[34].

If body burden data are used, results should be interpreted with caution. Organisms in field settings may be capable of acclimatizing or adapting to tolerate higher concentrations of COCs than would otherwise be expected. In such cases, actual risks may be lower than predicted risks. Conversely, because site-specific tissue data are generally collected from living organisms, risks may be underestimated if there are highly exposed organisms that have been eliminated from the population (e.g., through direct toxicity or reduced fitness).

Ancillary parameters – Ancillary parameters of importance for measuring internal exposures are the same as those associated with measurement of tissues for purposes of characterizing food items (**Section 3.3.1.6** above). However, any other parameters needed to support matching of the exposure data to effects data should also be considered.

3.3.3. Estimation of Total Dose

Defining total dose – Exposure is often assessed for higher level receptors (i.e., wildlife) as total dose or intake, which is the total intake of a contaminant from all exposure pathways. Total dose can be used as a measure of exposure whenever there are effects data to compare to, which may be a literature-derived dose-response relationship or a toxicity reference value.

> ***Key Concept:***
>
> Estimating exposure as total dose requires various types of data to characterize exposure for receptors. Technical Module 3 provides default values for several common wildlife receptors in Canada. These can be used as a starting point, particularly for simple ERAs or initial risk estimates. For cases where more precise estimation of risks is warranted, site-specific information should be used.

Using total dose as an exposure measure – Total dose is the most commonly used exposure measure for higher trophic level organisms (i.e., wildlife), and should always be considered for detailed risk assessments involving wildlife unless other lines of evidence are judged sufficient to draw conclusions about risks. As explained in **Section 2**, ingestion pathways – water, diet items, and incidental ingestion of soil and sediment – are usually by far the most important pathways, and inclusion of dermal exposure and inhalation pathways is rarely necessary. For each ingestion pathway, the minimum data needed to estimate total dose, and the recommended sources of data, are as follows (n.b.,

[34] In contrast, as explained in Sections 3.3.1.6 and 3.3.3, depuration is not usually appropriate as a default when tissue data are used as a food item for higher level receptors.

Technical Module 3 provides specific default values for many of the receptor characteristics listed below, for a range of common wildlife receptors in Canada):

- Ingestion rate for drinking water (typically characterized as L/day or L/kg body weight/day). For receptors not covered by Technical Module 3, water ingestion rates may be available in the primary literature or other sources (e.g., the USEPA Wildlife Exposure Factors Handbook – USEPA 1993). Allometric scaling can be used for organisms for which data are not available, using equations specified for example by Nagy (1987).
- Ingestion rate for food (typically characterized as kg food/kg body weight/day). For receptors not covered by Technical Module 3, food ingestion rates may be available in the primary literature or other sources (e.g., the USEPA Wildlife Exposure Factors Handbook – USEPA 1993). Allometric scaling can be used for organisms for which data are not available, using equations specified for example by Nagy (1987). Alternatively, equations relating food ingestion to metabolic rate can be used (USEPA 1993).
- Incidental ingestion rates for soil and sediment (typically characterized as a percentage of total food intake). For receptors not covered by Technical Module 3, incidental ingestion rates may be available in the primary literature or other sources (e.g., Beyer *et al.* 1996; see also CCME 2006 for discussion). These rates may vary depending on whether they account for soil and sediment contained in the digestive tract or trapped in fur (see bullet on contaminant concentrations below).
- Body weight of each receptor. For receptors not covered by Technical Module 3, water ingestion rates may be available in the primary literature or other sources (e.g., the USEPA Wildlife Exposure Factors Handbook – USEPA 1993).
- Diet proportions for any receptor that consumes more than one type of food. Default values are provided for some common receptors in Technical Module 3. However, diet proportions are highly site-specific and vary seasonally. For sites where precision in risk estimates is warranted, site-specific information should be collected (see Technical Module 3 for discussion)..
- Contaminant concentrations in soil, sediment, water and each food item. As discussed in **Section 3.2**, contaminant concentrations in each media are preferably measured, but can also be estimated. The specific tissues that are sampled should match consumption patterns. There are at least three considerations in this regard:
    - Whether to submit a whole animal (e.g., small mammal) for analysis, or only parts of an animal. If receptors are unlikely to consume (and digest) certain tissues such as bones or feathers, it may be appropriate to exclude those tissues from lab analyses.

- Whether to depurate (i.e., void the digestive tract) tissues prior to analysis, which applies to earthworms or filter feeders that take in volumes of soil or sediment but only digest the 'food' components. It is most conservative not to depurate (except for bioaccumulative substances), and this should be the default approach[35]. However, risk assessors should recognize that contaminants bound in the soil or sediment may not be bioavailable to higher trophic level consumers.
- Whether to wash tissues prior to analysis. This applies to any organisms such as invertebrates or particularly mammals with fur. Washing will remove soil trapped in fur, which is acceptable as long as estimated incidental soil ingestion rates account for this route of soil ingestion.

- Moisture content of soil, sediment, and food items, so that conversions can be made (as necessary) between ingestion rates and COC concentrations in food items. Whether data are reported on a dry weight or wet weight basis, harmonization of units is essential when calculating total dose. Moisture content should be measured by labs, or derived from primary literature.
- Home range size or forage range size of each receptor relative to the size of the site (or relevant portion of the site). The home range size should be estimated based on an up to date literature review, but can be adjusted based on professional judgement of a wildlife biologist (e.g., if habitat quality is low, range size may be larger). In some cases, a conservative screening assessment may assume that a receptor spends all of its time on a site, but more realistic assessments should apportion exposure between on-site and off-site (which requires data for off-site exposures). This is particularly important for large mammals or other receptors which may spend only a very small portion of time on a site.
- Other dose adjustment factors to account for partial bioavailability (or any other factor that is believed to affect actual dose) may also be used in more realistic models. Most TRVs are developed from studies conducted using highly bioavailable forms of contaminant (e.g., soluble metal salts) that may overestimate actual availability from site media. In the absence of specific information about bioavailability, risk assessors should assume 100% bioavailability, although this will typically contribute to overestimation of exposure.

[35] Depuration is also not recommended as a default when diet items are used directly as an external exposure measure, for example when comparing to CEQGs for tissue residue for protection of wildlife consumers of aquatic life. In contrast, when tissues are collected and evaluated as an internal measure of exposure for the organism itself, depuration is usually appropriate – see discussion in Section 3.3.2.

Food chain models – Simple models to calculate total dose can be formulated in a spreadsheet. An example of the equations that might be used is provided in the text box at the end of this subsection. Complex models covering several COCs, several receptors and several distinct areas of a site may warrant more elaborate set up than a series of equations in a spreadsheet. For example, to avoid repetition of formulae in a spreadsheet, it may be appropriate to host input data in a spreadsheet or data file, then use a programming language to read the data and perform all calculations (e.g., Visual Basic for Applications, if the input data are contained in Microsoft Excel), and then output the results back to the spreadsheet or a data file. Visual Basic for Applications can also call other software that is useful for particular functions (e.g., Crystal Ball for probabilistic models). Software packages designed specifically for risk assessment are also available (e.g., GoldSim). Use of more than one method of estimating total dose can be valuable for detecting errors (i.e., QA/QC check of the model).

> ***Key Concept:***
>
> Food chain models are a series of equations that can be set up in a spreadsheet, though more elaborate models warrant programming.

Key Concept:

Food chain models can be formulated in a spreadsheet using a series of equations. Following is an example set of equations:

1. **Food Ingestion Rates** – Food ingestion rates (*FI*, kg dw/kg ww/day), if not known for a given receptor, can be estimated using allometric equations such as those described in Nagy (1987) for various feeding guilds, i.e.,

   $$FI = a \times BW^{b} \quad \text{(Eq. 1)}$$

   Where:
   BW represents the organism's mean body weight (g, ww)
   a and *b* are constants specific to various groups of terrestrial vertebrates
   These dry weight food ingestion rates can be converted into wet weights (I_F, kg ww/kg ww/day) following equation 2:

   $$I_F = FI \Big/ \left(1 - moisture_{diet}\right) \quad \text{(Eq. 2)}$$

   Where:
   $Moisture_{diet}$ (unitless fraction) represents the weighted average moisture content in the diet of the animal, based on measured contents in tissues from the site or values from the literature.
2. **Soil and Sediment Ingestion Rates** – Soil and sediment ingestion rates (I_S, kg dw/kg ww/day) based on an estimated fraction of incidental ingestion during foraging activities. If not known for a given receptor, they can be derived from the food ingestion rate according to:

   $$I_S = FI \times \phi \quad \text{(Eq. 3)}$$

   Where:
   FI (kg dw/kg ww/day) is the dry food ingestion rate
   ϕ is the fraction of incidental soil or sediment ingested during feeding.
3. **Drinking Water Ingestion Rates** – Drinking water ingestion rates (I_W, L/kg ww/day), if not known for a given receptor, can be estimated using allometric equations such as those described in Nagy (1987):

   $$I_W = a \times BW^{b} \quad \text{(Eq. 4)}$$

   Where:
   BW (kg, ww) represents the organism's mean body weight
   a (L/kg*kg/day) and *b* (unitless) are constants specific to various groups of terrestrial vertebrates
4. **Dose From Food -** Intake dose of contaminants from food (D_F, mg/kg·bw/day) determined from the dietary concentration following:

   $$D_F = I_F \times \sum_{1}^{j}\left(C_{Fj} \times p_{Fj}\right) \quad \text{(Eq. 5)}$$

Where:
I_F (kg ww/kg bw/day) represents the feeding ingestion rate
C_{Fj} (mg/kg ww) represents the COC concentration in prey item *j* in the diet of the ROC
p_{Fj} (unitless) represents the proportion of prey item *j* in the diet of the predator

continued next page

5. **Dose From Soil Intake** (primarily terrestrial foragers) - The total dose from incidental ingestion of COC contaminated soil (D_S, mg/kg·bw/day) calculated using the following equation:

$$D_S = I_S \times C_S \quad \text{(Eq. 6)}$$

Where:
I_S (kg dw/kg bw/day) represents the ingestion rate of soil
C_S (mg/kg dw) represents the COC concentration in ingested soil

6. **Dose From Sediment Intake** (primarily aquatic foragers) - The total dose from incidental ingestion of COC contaminated sediment (D_{SED}, mg/kg·bw/day) calculated using the following equation:

$$D_{SED} = I_S \times C_{SED} \quad \text{(Eq. 7)}$$

Where:
I_S (kg dw/kg bw/day) represents the ingestion rate of sediment
C_{SED} (mg/kg dw) represents the COC concentration in ingested sediment

7. **Dose From Drinking Water** – The total dose from drinking water ingestion of COCs (D_W, mg/kg·bw/day) calculated using the following equation:

$$D_W = I_W \times C_W \quad \text{(Eq. 8)}$$

Where:
I_W (L/kg bw/day) represents the drinking water ingestion rate
C_W (mg/L) represents the COC concentration in the water

8. **Total Unadjusted Dose -** The unadjusted dose (D_{UT}, mg/kg ww/day) can be calculated by taking the sum of the doses for the separate media: food, soil, sediment, water:

$$D_{UT} = D_F + D_S + D_{SED} + D_W \quad \text{(Eq. 9)}$$

Where:
D_F (mg/kg wet/day) is the dose from food
D_S (mg/kg wet/day) is the dose from soil
D_{SED} (mg/kg wet/day) is the dose from sediment
D_W (mg/kg wet/day) is the dose from water

9. **Dose Adjustment Factor -** The dose adjustment factor can be calculated as a function of territory/foraging range, habitat quality, and bioavailability of the COCs.

$$DAF = FRF \times \alpha \quad \text{(Eq. 10)}$$

Where:
FRF (unitless) is the foraging range factor, which represents the surface area of the site that overlaps with the territory or foraging range of the species.
α (unitless) is the dietary uptake efficiency of a given chemical and can be thought of as the proportion of chemical that is absorbed through the intestinal tract compared to the total amount ingested. The value does not account for difference in availability between soil and different food types.

10. **Total Adjusted Dose** – The total adjusted dose (D_{AT}, mg/kg wet/day) calculated by multiplying the unadjusted dose and the dose adjustment factor:

$$D_{AT} = D_{UT} \times DAF \quad \text{(Eq. 11)}$$

Where:
D_{UT} is the unadjusted total dietary dose of a given chemical (mg/kg wet/day)

3.4. Beyond Point Estimates of Exposure

For receptors that are relatively immobile (e.g., invertebrates and plants), the assessment of exposure is typically conducted on a spatially explicit basis. This may be conducted by directly applying station-specific measurements of exposure to represent a management unit (grid cell), or by using multiple measurements to generate a modeled surface of exposures.

For wildlife (birds, mammals) and mobile aquatic and semi-aquatic organisms (fish, amphibians), exposure estimation is more challenging. Screening level risk assessments often apply the principle of the exposure point concentration (or estimated exposure concentration), which is a conservative point estimate of the chemical concentration (or dose) available from a particular medium or route of exposure. Simple models may use the maximum concentrations from each medium to represent the exposure point concentration, or some other statistical metric (e.g., 95% UCLM or 90^{th} percentile) depending on sample sizes.

Disadvantages of the simplified point estimate approach described above include:

- lack of consideration of the relative spatial positions of receptors and contaminated media (due to habitat preferences, migration patterns, etc.), which can strongly influence estimates of exposure;
- overreliance on extreme values (maxima) in the calculations of exposure point concentrations;

The point estimation approach assumes that receptors have equal and random access to all areas of an exposure unit, and occur evenly throughout the exposure unit. These conditions rarely apply in natural environments.

Point estimates of exposure can be improved by using probabilistic methods (see Sections 5.3.6 and 5.6.3), and by incorporating spatial information as discussed below.

3.4.1. Partially Spatially Explicit Approaches

Several methods are available for cases where use of summary statistics yields unacceptable uncertainty. If more spatial realism is desirable, more advanced methods can be used such as:

- Dividing drinking water exposure among several sources based on evaluation of likelihood of use.
- For purposes of estimating incidental soil ingestion, soil samples can be weighted by their spatial ‘area-of-influence’, and/or by their relative probability of use by a receptor based on evaluation of habitat preferences (i.e., less soil will be incidentally ingested in areas subject to low use). The result of this weighting may

be a spatially weighted average concentration (SWAC) that is used in the ERA for evaluating incidental ingestion. This approach would typically require overlay of soil data and habitat polygons using Geographic Information Systems (GIS).

- For purposes of estimating ingestion of contaminants through food items, the food item concentrations measured in various areas of a site can be weighted according to their relative probability of consumption based on habitat preferences (e.g., a sample of insect tissue in one area of the site would receive twice the weight of another that occurs in habitat that is half as preferable for an insectivorous receptor).
- The curve model (Freshman and Menzie 1996) may be used to describe the exposure to wildlife that forage over the contaminated site. The approach is based on rank-ordering of contamination measurements, and consideration of the home range (foraging area) of the species of interest. This approach considers the distribution of concentration measurements (both frequency and magnitude), but does not account for natural foraging patterns or habitat preferences.

All of these types of improvements are attempts to account for spatial information, and risk assessors should implement these refinements where the level of effort is justified by the increased precision in risk estimates.

3.4.2. Spatially Explicit Methods

None of the refinements in Section 3.4.1 result in a truly spatially explicit exposure model. Exposure models that are truly spatially explicit aim to simulate the spatial behaviour of individual animals on a site, in the context of the habitats and other factors that influence site use. This is the only way to realistically capture variability in exposure within a population of animals.

With advances in GIS, explicit consideration of the heterogeneous distribution of receptors, their habitats, and contamination is increasingly feasible. Tools for incorporating such spatial considerations in ecological risk assessments are more available, although they tend to be applied on large complex risk assessments.

Some models exist that may be adaptable to particular sites (e.g., the Spatially Explicit Exposure Model, and others reviewed by Loos *et al.* 2010 and Wickwire *et al.* 2011), but their flexibility is often limited and there has not yet been widespread application of such models. Use or development of spatially explicit exposure models should be considered where increased precision in risk estimates is worth the cost (see Hope *et al.* 2011 and Wickwire *et al.* 2011 for further discussion).

4. EFFECTS ASSESSMENT

4.1. Introduction

4.1.1. Purpose

The general purpose of effects assessment is to characterize the nature of effects elicited by each COC under an exposure condition that is relevant to each ROC. This characterization is often called a response profile and is required for each COC / LOE combination. Note that for some LOEs (e.g., toxicity tests of contaminant mixtures) it may only be possible to characterize the response for that mixture.

Key Concept:

Effects information is an input, along with exposure information, for every line of evidence in an ERA. Both effects and exposure data need to be reported in compatible units to facilitate integration of the results in the risk characterization stage of risk assessment.

Definitions:

Response profile is the relationship between COC concentrations or doses and ecological effects.

Toxicity reference value is broadly defined as an exposure concentration or dose that is not expected to cause an unacceptable level of effect in receptor(s) exposed to the contaminant of concern.

There are a variety of ways (these are not mutually exclusive) that effects information can be used in risk assessment:

- Develop a toxicity reference value (TRV). TRVs are commonly used in the hazard quotient (HQ) method of risk characterization (see **Section 5.3.1** for more details), where they are compared to exposure estimates.
- Develop concentration-response (or dose-response) relationships. These can be used directly to estimate effect levels for a particular exposure concentration, or they can be used to derive TRVs for specific effect levels.
- Develop a site-specific remediation objective, for a site where an initial ERA indicates that risk management is warranted, using either (a) a TRV (first bullet above) based on literature or site-specific data, or (b) a concentration-response relationship (second bullet above) based on literature or site-specific data.

Exposure and effects are matched together in one or more ways for every line of evidence (LOE) that is evaluated in an ERA. Consequently, effects assessment is not a single step in ERA, but is carried out for every LOE. Importantly, while the details of effects

assessment are discussed in this section, they must be fully understood and articulated at the problem formulation stage in order to support design and planning of the ERA.

4.1.2. Overview of Effects Assessment

Effects assessment used to develop any particular line of evidence generally entails the following elements (note that the first four elements will have been decided as part of the problem formulation):

- Determine which ***type(s) of effects assessment measure*** will be used, among the four broad types:

    1. *Site-specific toxicity studies* – Considers measurement endpoints related to studies of test organism exposures to contaminated site media under controlled conditions. This category includes toxicity tests conducted in the laboratory using site-collected media, conducted in the field (*in situ*), or a combination of both. The category includes both standardized test protocols and exploratory techniques, such as toxicity identification evaluations.

    2. *Indirect toxicity information* – Considers toxicological information derived from other sites (or laboratory studies), under an assumption that the concentration-response relationship is either similar to, or can be estimated from, the data collected at other sites. Results are extrapolated to the site of interest through consideration of contamination profiles, habitat similarities, and factors that may influence relative bioavailability (*e.g.*, chemical speciation, organic carbon or lipid content, particle size, salinity, *etc*). Indirect toxicological evidence can take many forms, ranging from generic environmental quality guidelines based on toxicity database information, to concentration-response relationships gleaned from the literature or drawn from focused studies conducted at other sites.

    3. *Site-specific biological studies* – Considers direct assessment of the site biological condition relative to the exposure metric. This category may include endpoints at the sub-organism level (*e.g.*, histopathological indicators), organism level (*e.g.*, mortality, growth, deformities, erosions, lesions, and tumors), population level (*e.g.*, numbers and proportions of indicator organisms, vital rates), and community level (*e.g.*, diversity, distribution of taxonomic groups).

    4. *Indirect biological information* – Considers indirect assessment of biology, through extrapolation of knowledge obtained at other sites. As with toxicology studies, the biological evidence must be scaled to the site condition based on consideration of exposure levels and ecological relevance. Given natural

ecological variability, indirect biological information alone would almost never be sufficient for characterizing risks as part of a detailed ERA.

- Determine ***whether the effects data will be interpreted relative to an acceptable effect level (AEL;***(e.g., to derive a TRV) or used without pre-determined AELs (e.g., to estimate actual effect sizes and leave the determination of "acceptable" or "unacceptable" to risk managers).

- Determine ***how contaminant mixtures will be considered***. While response profiles need to address each COC / ROC combination, a single response profile could address multiple COCs simultaneously when appropriate measures are used. Site-specific effects assessment measures (e.g., toxicity testing or biological surveys) allow for explicit consideration of chemical mixtures present at a site, thus integrating all interactions. Consequently, site-specific approaches are usually recommended where feasible.

- Decide which ***type of response profile*** should be developed given the nature of the available effects data:

  1. *Continuous response profile* – documents how effects (i.e., magnitude of response) vary over the range of realistic exposure levels. The profile can be used directly in risk characterization (e.g., when estimating actual effect levels associated with a particular exposure level), or can be used to derive a toxicity reference value (TRV) for a response magnitude of interest (e.g., what exposure level corresponds to a 20% adverse response). Understanding the exposure-response relationship also facilitates the interpretation of the potential effects should predicted exposure exceed a TRV in the risk characterization.

  2. *Discrete response profile* – occurs in situations where effects data are scarce (e.g., limited literature effects data for some COCs for wildlife) or when effects apply to particular exposure scenarios only (i.e., those occurring at specific locations on the site). This could arise when using a control-impact study design (e.g., when determining if a contaminated area differs from a reference condition) or a gradient design with discrete levels of impact, or when testing complex contaminant mixtures using site-specific effects

> ***Key Concept:***
>
> The response profile must consider exposure conditions consistent to those expected at the site. This not only pertains to exposure intensity, duration and spatial pattern, but also to COC-specific information (e.g., COC form, congeners versus total PCBs, etc.) and to factors that control bioavailability (see **Section 3.3**). It is preferable to include effects assessment measures that integrate site-specific exposure conditions because they are typically more realistic and have lower associated uncertainty.

measures (e.g., toxicity testing or biological surveys). In the first case (i.e., limited data), TRVs can still be derived from a discrete response profile, but they may not coincide with the desired effect magnitude or exposure condition.

- ***Develop response profiles*** for each COC/ROC combination, or as appropriate (e.g., for contaminant mixtures) if COC/ROC-specific profiles are not feasible or appropriate.
- Characterize ***uncertainties*** in effects, evaluate the implications of uncertainty using sensitivity analysis, and, if warranted, integrate uncertainties into the effects assessment (e.g., using probabilistic methods).

The outcome of effects assessment are measures that can be matched with exposure estimates to provide evidence in the form of a LOE. It is critical that the risk assessor conceptualize the exposure and effects information at the same time (during problem formulation) to ensure that they can be integrated effectively.

The remainder of this section is organized as follows:

Section 4.2 describes the main approaches for effects assessment;

Section 4.3 considers the application of those approaches to particular receptor groups;

Section 4.4 discusses options for moving beyond point estimates of toxicity; and

Section 4.5 discusses the role of safety factors in effects assessment.

4.2. Categories of Measures for Effects Assessment

This section discusses measures for effects assessment, categorized according to the four broad categories for Lines of Evidence that were introduced in **Section 4.1.2** (examples provided in **Table 4-1**; more details provided in **Section 5**):

- Site-specific toxicity studies;
- Indirect toxicity information;
- Site-specific biological studies; and
- Indirect biological information.

Table 4-1. Examples and categorization of effects assessment measures.

Endpoint Type	Source of Dose-Response Information	Line of Evidence Type	Examples of Relevant Measures
Toxicology	Site of Interest	(1) Site-Specific Toxicological	Laboratory seed germination test conducted using site soils; Caged mussel study; Amphibian metamorphosis assay using larvae harvested in site vernal pool; *In-situ* test of survival and growth for *Hyalella*; Laboratory test of early lifestage fish growth and survival.
Toxicology	Guideline	(2) Indirect Toxicological	Water quality guideline developed from most sensitive tested species; Sediment quality guideline developed from co-occurrence database (BEDS); Soil quality guideline for protection of microbial processes.
Toxicology	Literature / Compendium	(2) Indirect Toxicological	EC_x threshold from Ecotox database review; Avian dose-based toxicity reference value from Eco-SSL; Critical tissue burden from literature search; Species sensitivity distribution.
Toxicology	Other Site	(2) Indirect Toxicological	Use of threshold for reproductive success from a captive mink feeding study conducted for another site using fish harvested from that site.
Biology	Site of Interest	(3) Site-Specific Biological	Benthic community enumeration; Evaluation of salmon reproductive success and output; Small mammal survey (density, biomass, net migration); Vegetation transect or quadrat enumeration.
Biology	Literature / Compendium	(4) Indirect Biological	Literature summary of water concentrations associated with reduced richness of epibenthic invertebrates. Literature summary of relationship between average sediment COC concentrations and incidence of tumours in fish.
Biology	Other Site	(4) Indirect Biological	Reproductive study of tree swallows (using nest box assessment) at Site A that could be used to assess potential avian effects at Site B assuming some consistency of response for a standardized measure of exposure.

These categories of measures are distinguished by two factors:

- Site-specific *versus* Indirect – based on whether the measure addresses effects in exposure media (or receptors) from the site, either *in situ* or *ex situ*, or if it relies on published effects data (e.g., from published research or grey literature or from other contaminated sites).
- Toxicological *versus* Biological – this distinction essentially relates to whether a measure involves experimental manipulations to control environmental variables so that treatments differ only in their exposure to COCs (e.g., testing fish growth in site water in a laboratory or field testing in enclosures) or whether it focuses on quantifying the effects associated with naturally-occurring exposure situations (e.g., a benthic invertebrate community study or a small mammal population survey). The distinction becomes less clear in field tests with high realism, but such field tests are rarely used in practice.

The types of measures used for effects assessment are not mutually exclusive – as described in **Section 4.2.4**, risk assessors are encouraged to use more than one type of effects measure even for the same assessment endpoint. The decision about what types of effects measures to use will be based in part on:

- Level of rigour needed to inform decision-making;
- Resources required to properly use a measure;
- Availability and quality of information (e.g., for published studies);
- Confidence in likely measure performance (i.e., ease of extrapolation to assessment endpoint and associated uncertainties); and
- Availability of matching exposure data with which effects measure outputs would be combined or integrated.

Regardless of what types of measures are used for effects assessment, it is critical that they be specified in the problem formulation when the lines of evidence to be used in the ERA are developed.

4.2.1. Site-specific Toxicity Studies

> ***Key Concept:***
>
> Site-specific toxicity assessments are used to directly test whether exposure to contaminated media (e.g., water, sediment, soil) from a site elicits adverse effects in test organisms under controlled conditions.

Site-specific toxicity assessments are used to directly test whether exposure to contaminated media (e.g., water, sediment, soil) from a site elicits adverse effects in test organisms under controlled conditions. The latter is an important distinction from field studies in that by controlling environmental variables, the test

medium becomes the primary independent variable (i.e., predictor), with test endpoints being the dependent variables (i.e., outcomes).

Although all options for this type of measure involve some form of experimental manipulation to help reduce the influence of non-chemical factors on the outcome of the test, they vary in how well they mimic reality. At one end of the spectrum are standardized laboratory toxicity tests (i.e., *ex situ* exposure to site media), where site media are taken to a laboratory facility and tested under controlled conditions following a detailed protocol. These tests are by far the most commonly used and are the primary focus of this guidance. At the opposite end of the spectrum are highly customized *in situ* toxicity studies.

Laboratory tests under controlled conditions are valuable in that they can help isolate a toxic mechanism that could be obscured in a natural environment. As a result of these controls, laboratory tests tend to be more precise, though not necessarily more accurate (relevant) in terms of describing the assessment endpoint. Because the type of errors in toxicity tests differ qualitatively from those in field studies, it is not appropriate to compare the concentration-response results using only a coefficient of determination (r^2) or other purely statistical measure. Rather, assessment of uncertainty of laboratory testing must consider both numerical measures of uncertainty (e.g., inter-replicate variability) and incertitude associated with lab-to-field extrapolation.

There are several options available for cases where the degree of environmental realism of standard toxicity tests is insufficient to properly derive a response profile (e.g., when the physical test setup is not appropriate or when sample handling of the target exposure medium might increase or decrease COC bioavailability). Some examples of ways in which toxicity testing can be modified to increase environmental realism include:

- Alteration of standard protocols (e.g., physical test setup) to increase test realism in a laboratory setting (see Appendix A of **Technical Module 3** for more information). An example would be to increase the number of refreshes of overlying water to better represent a flowing environment.
- Setting up temporary testing facilities at the site (e.g., setting up a continuous flow-through setup taking water directly from an area of interest).
- Conducting an *in situ* toxicity study (e.g., in enclosures such as pens or mesocosms) (see Appendix A of **Technical Module 3** for more information).

Another type of specialized site-specific toxicity testing is toxicity identification evaluation (TIE). TIEs involve physical/chemical manipulation of a sample to try to isolate and identify toxic substances in a test medium. TIEs are applied in an iterative fashion to progressively pinpoint a specific toxicant or class of toxicants. Clear identification of the specific cause of toxicity can reduce uncertainty and increase confidence in conclusions. Information on TIEs is provided in **Technical Module 1**.

Guidance on toxicity test selection and interpretation is presented in **Technical Module 1**. This comprehensive technical module covers the following:

- An overview of toxicity testing in risk assessment, with specific emphasis on how tests are used in a weight-of-evidence approach and how they can also be used to develop a site-specific TRV (additional information on site-specific TRVs can be found in **Technical Module 2**);
- Procedures for test selection;
- Additional considerations specific to pore water;
- A summary of key information for about 75 of the most commonly used toxicity tests in North America; and
- Interpretation of toxicity test results.

Site-specific toxicity tests are considered more useful than indirect toxicity information for the following reasons (Suter *et al.* 2000):

- The site-specific bioavailability of the contaminants is considered;
- The form of the contaminant is realistic;
- Interactions among contaminants are simultaneously addressed;
- Spatial distribution of toxicity can be determined; and
- Remedial goals may be determined with higher confidence.

Key limitations include (Suter *et al.* 2000; SAB-CS 2008):

- The medium may be modified by sample collection and test preparation (particularly for sediments, but also for water and soil), which could affect contaminant form and bioavailability;
- Differences in sensitivity between the test organism and the ROC may not be known. This could be due to taxonomic or genetic differences (e.g., some strains of test organisms are known to be particularly sensitive), or to other factors like acclimation (e.g., where pre-test holding conditions affect organism sensitivity in the toxicity test for essential elements) or adaptation (e.g., where an organism's natural detoxification systems may not be working optimally due to holding in low-metals water);
- Testing scenario (e.g., duration and setup) may not fully reflect site-specific realities;
- The cause of toxicity is not known (unless a TIE or other method for establishing causal linkages is conducted);

- Apparent toxicity may be due to differences between reference and site media in factors other than contaminant concentrations (e.g., higher nutrients or substrate-based responses in reference).
- Variability of test endpoints, particularly for sublethal endpoints during chronic exposures, may reduce the statistical power to detect target effect sizes.
- High costs, particularly for chronic testing, may force trade-offs in spatial or temporal sampling coverage; and
- Effects are measured on individual organisms, which may then need to be extrapolated to or used to predict population- or community-level assessment endpoints.

Many of these limitations directly become sources of uncertainty for this type of measure. **Section 5.6** discusses approaches for addressing uncertainties.

4.2.2. Indirect Toxicity Information

Risk assessments can benefit from the consideration of the substantial body of literature available from ecotoxicological research. Access to this information has been facilitated by the internet, where one can search online data compilations or search and retrieve primary literature. Thus, for a relatively low cost compared to other types of measures, a wealth of information can be accessed to augment the effects assessment in a number of ways, including:

> ***Key Concept:***
>
> Indirect toxicity information taps into the wealth of knowledge available in published studies. Used judiciously, this can be a cost-effective source of relevant data to develop response profiles.

- Compiling preliminary effects information during problem formulation (see **Section 2.2.4** for more details).
- Identifying and sourcing published effects models (e.g., BLM; see below for more details).
- Compiling response profiles and deriving TRVs (see **Technical Module 2** for details).

Guidance on the use of indirect toxicity information in the development of TRVs is provided in **Technical Module 2**. TRVs are an important part of the response profile in that they represent a concentration or dose that is not expected to cause an unacceptable adverse effect (see **Section 2.3.1** for more discussion on acceptable effect levels). Technical Module 2 covers the following:

- Types and use of TRVs in ERA

  - Dose-based TRVs
  - Concentration-based TRVs for exposure media
  - Concentration-based TRVs for tissues
- Options for TRV Selection
- Review of published TRVs
- General considerations for TRV derivation
- Derivation of site-specific, literature-based TRVs
  - Literature review
  - Data quality and selection criteria
  - Derivation methods
  - Uncertainty and extrapolations
- Modification of existing guidelines to develop site-specific TRVs

In addition to the limitations inherent in extrapolating from the laboratory to the field (discussed in **Section 4.2.1)** for site-specific toxicity measures, use of indirect toxicological information also requires consideration of the site-specific relevance of the data. Potential sources of bias in literature toxicity data that are uncertainties for this type of measure include (Suter *et al.* 2000):

- Chemical form used in toxicity tests may be more toxic than the dominant forms found at a contaminated site;
- Contaminant interactions are rarely considered;
- Test species may not be representative of the sensitivity of ROCs at the site;
- Exposure test media may not be representative of those found at the site; and
- Laboratory test conditions may not be representative of field conditions.

The relevance of indirect toxicological information can be improved by filtering the available data to include studies that most closely match the needs of the ERA. Depending on the contaminant, one or more of the ancillary parameters listed in **Sections 3.3.1 and 3.3.2** may play a key role in determining its toxicity (e.g., by affecting bioavailability). Uncertainty can be substantially reduced by appropriately matching reported test conditions to actual exposure conditions. Although in many situations the risk assessor must perform the task of filtering (if possible and appropriate), ideally the key factors affecting bioavailability and toxicity would be understood sufficiently to support site-specific predictive modeling of toxicity.

There have been advances in supporting science that may address some of the common limitations to indirect toxicity information highlighted above. These include the Biotic Ligand Model (BLM) and the Tissue Residue Approach (TRA) for toxicity assessment. These are discussed below.

Biotic Ligand Model – Research in recent decades (e.g., Pagenkopf 1983; Meyer 1999) has led to major progress in our understanding of metals bioavailability and mechanisms of toxicity in both the aquatic (see review by Paquin *et al.* 2003) and terrestrial (see review by Allen 2002) ecosystems. The culmination of this research to date is the development of the biotic ligand model (BLM), which integrates key discoveries from several disciplines to consider a range of factors influencing metals bioavailability and ultimately, toxicity. The premise of BLM is that toxicity is related to the metal binding to an active biochemical site on the organism (i.e., the biotic ligand) and that binding is related to concentrations of free metal cations and complexing ligands in the water (or solution phase for soils). The latter compete with the biotic ligand (e.g., in fish gills or at root elongation sites for plants) for free metals and other cations in the water (or solution phase for soils), thus directly affecting toxicity by dictating metals concentrations at the target site. A major advantage of the BLM is that it explicitly considers a range of modifying factors (e.g., as competing cations) influencing the response profile of a particular endpoint.

Aquatic BLM – The aquatic BLM (see http://www.hydroqual.com/wr_blm.html for more information and free model downloads) has successfully been used for predicting acute aquatic toxicity related to copper (Santore *et al.* 2001), silver (Paquin *et al.* 1999) and zinc (Santore *et al.* 2002). The success of the BLM in accurately predicting toxicity has already led to its use in developing water quality criteria; the BLM features prominently in the USEPA's criteria for copper (USEPA 2007a). More recently, research has focused on BLM's application to chronic toxicity (De Schamphelaere *et al.* 2005, Schwartz 2007, Clifford 2010, Schroeder *et al.* 2010, Peters 2011) and metals mixtures (Kamo and Nagai 2008), which should lead to increased use of the aquatic BLM in risk assessments.

Terrestrial BLM – More recent efforts have resulted in the development and validation of BLMs specifically for terrestrial ecosystems. Thakali *et al.* initially applied a terrestrial BLM to predicting copper and nickel toxicity to barley root elongation in a number of soils (2006a), then to an expanded suite of toxicity endpoints (plants, invertebrates and microbes) across a range of non-calcareous soils from the European Union. Terrestrial BLMs have also been used to predicting cobalt toxicity to worms (Lock *et al.* 2006) and barley (Lock *et al.* 2007). These methods are likely to be refined and expanded to other metals and toxicity endpoints.

Tissue Residue Approach – Another rapidly advancing area is the tissue residue approach for toxicity assessment (TRA). A Society of Environmental Toxicology and Chemistry (SETAC) Pellston Workshop in 2007 lead to a series of "state-of-the science" papers on this subject (Adams *et al.* 2011, Escher *et al.* 2011, McCarty *et al.* 2011, McElroy et al

2011, Meador *et al.* 2011, Sappington *et al.* 2011). This approach works on the premise that whole-body or organ-specific concentrations (residues) are a better dose metric for describing toxicity to organisms than external exposure media (Escher *et al.* 2011). While this is somewhat intuitive (i.e., because contaminant bioavailability is explicitly considered in TRA), the approach is not without its challenges, largely due to difficultly correlating internal concentrations to ecotoxicological outcomes. Variability in ecotoxicological outcomes and species sensitivity is due in part to differences in toxicokinetics, which is comprised of several key processes (absorption, distribution, metabolism, and excretion [ADME]) that influence internal concentrations (Escher *et al.* 2011). Where variability is high (i.e., where internal concentrations are not proportionate to the concentration or dose at the target site), toxicokinetic modeling may be useful to derive the target dose. One of the main challenges of using the TRA will be the availability of appropriate tissue residue-response data (Sappington *et al.* 2011). Given the advancing state of the science, it would be prudent to treat this as a complementary LOE (Sappington *et al.* 2011).

4.2.3. Site-specific Biological Studies

Key Concept:

Site-specific biological studies directly assess ROC attributes in the field, thus eliminating many of the uncertainties associated with toxicological information.

Site-specific biological studies directly assess ROC attributes in the field, thus eliminating many of the uncertainties associated with toxicological information. These studies can target a range of attributes for individuals (e.g., growth, reproductive success, survival), populations (e.g., biomass, abundance, density, age structure) or communities (e.g., diversity, species composition, abundance, density, biomass), making it possible to directly estimate the assessment endpoint (Appendix D in CCME 1997, Menzie *et al.* 2008, Carlsen *et al.* 2008). Comparisons should be made to reference conditions or along gradients in exposure. Unlike toxicity studies, however, where several environmental variables are controlled to help isolate an exposure-related "signal," biological field studies can be clouded by natural variability due to the inherent complexity of natural systems. Some of this natural variability can be controlled through proper experimental design (including identification of covariates and categorical factors) and through increased sample size (either in single studies or multiple monitoring events).

Risk assessors should consider the following for deciding whether biological field studies are appropriate (Suter *et al.* 2000):

- *Scale* – These studies are usually most appropriate for ROCs with small home range sizes and that are likely to remain mostly inside the boundaries of the assessment area. However, biological field studies may also be appropriate for

highly mobile, wide-ranging ROCs, particularly when those ROCs are of particular importance to stakeholders.

- *Interpretation* – Variation in the attribute of interest must be interpretable in the context of confounding factors such as habitat heterogeneity.
- *Difficulty* – Studies can vary greatly in the cost and time needed for implementation. This needs to be balanced against the chances of obtaining useful information.
- *Appropriateness* – The study design and methods need to match the task at hand.
- *Technical expertise* – Study complexity may require specialized expertise beyond the risk assessment team.
- *Consequences of survey* – In some cases (e.g., destructive sampling of small populations or of rare species), biological studies may cause unacceptable harm.
- *Data availability* – Suitable surrogate data sets may be available (e.g., from broader environmental management initiatives; see **Section 4.2.4**).

Once a risk assessor has committed to site-specific biological effects measures in the ERA, the following additional considerations may help to design and implement the study. The design has to be worked out as part of problem formulation (see **Section 2.3**). At a minimum, risk assessors undertaking site-specific biological studies consider the following during the study design phase (and seek out more specific information relevant to their unique situation):

- *Define the question* – where possible, the focus of the study should be direct estimation of the assessment endpoint. In other cases, study objectives and how the results will be extrapolated to the assessment endpoint should be determined in advance (i.e., during the problem formulation).
- *Defining the assessment population* – this question has important implications for how effects might be interpreted (e.g., the larger the assessment population relative to the site, the more effects may be "diluted"). As a starting point, consider defining the assessment population as those organisms inhabiting the site of interest. Scale can then be adjusted based on ROC-specific considerations (see Menzie *et al.* 2008 for more discussion of assessment populations).
- *Selecting relevant attributes* – as discussed above, this should either match, or be easily extrapolated to, the assessment endpoint. Multiple attributes are recommended where practical to provide a more robust means of assessing the question (Appendix D in CCME 1997; Menzie *et al.* 2008, Carlsen *et al.* 2008).
- *Designing the study* – appropriate scientific methods (e.g., Krebs 1989; Environment Canada 2011) should be used to optimize the design to answer the

question. This will include having an understanding of the statistical methods that will eventually be used.

- *Field methods* – methods for most study types have at least been published and may even have recommended survey methods or standard protocols (e.g., PSAMP 1996; Environment Canada 2002; see SAB-CS 2008 for more references; Also, USEPA has a variety of methods posted at their Environmental Monitoring and Assessment Program website (http://www.epa.gov/emap/).
- *Data quality objectives and QA/QC* – data quality objectives (DQOs) define the specifications for the data set, quality assurance (QA) steps are the actions taken to facilitate meeting those objectives, and quality control (QC) measures are the benchmarks used for verification of data quality (e.g., Environment Canada 2011).
- *Data analysis and interpretation* – the statistical methods used should be those selected during the design phase. Interpretation should consider key uncertainties. This is often done by reporting observed effect sizes with confidence limits for each attribute (e.g., Environment Canada 2011).

Advantages of well-planned, site-specific biological studies include:

- Assessment endpoints can be directly estimated;
- They integrate exposure by accounting for complexities such as bioavailability and contaminant mixtures;
- They have a high degree of ecological relevance; and
- They are complementary to toxicity data.

Limitations of site-specific biological studies include:

- The cost and time needed to obtain robust data sets can be high, in which case such a study is warranted only when the likely value in informing management decisions is also high;
- Natural variation can make it difficult to detect contaminant-related changes, even in well planned studies;
- Measured effects may not be due to COCs, but rather due to confounding natural environmental variables or non-chemical stressors.
- Studies usually have to rely on spatial comparisons (e.g., across exposure gradients) due to the scarcity of baseline data for the site of interest. Selection of appropriate reference areas can be challenging;
- Conducting the studies may cause direct adverse effects to the target ROCs; and

Some of these limitations can translate directly into uncertainties. High natural variability can mask the detection of target effects, thus potentially resulting in a false negative conclusion (i.e., Type II error). In contrast, differences between exposure and reference conditions may result in measured effects that are not actually due to contaminant exposure, resulting in a false positive conclusion (i.e., Type I error). Statistical methods (e.g., confidence limits on effects sizes and power analysis) can be used to help understand the scope for Type II errors. Complementary use of site-specific toxicity testing can help establish causality (or lack thereof) for field studies (see **Section 5.5.2.2** and **Technical Module 4** for more discussion of causality).

4.2.4. Indirect Biological Information

This category of measures is analogous to indirect toxicity information, but emphasizes transference of appropriate biological studies from other sites (e.g., published in the literature) that could be used to help inform a response profile for the site of interest.

> ***Key Concept:***
>
> Biological studies reported in the literature can provide valuable information with which to derive a response profile for specific COC/ROC combinations. For example, Brasso and Cristol (2008) studied effects of mercury exposure on the reproductive success of tree swallows (*Tachycineta bicolor*). The authors collected several measures of mercury exposure (blood and feather total mercury in the birds, and total mercury in the insects fed to nestlings) against their primary effect measure, the number of nestlings that left the nest (i.e., fledglings). With consideration of the points listed in the text, this study could provide highly relevant data for other sites where mercury is the primary COC.

Given the resource and technical challenges of designing and implementing useful biological studies discussed in **Section 4.2.3**, the advantages of finding an appropriate study are clear. The main challenge, however, is overcoming the hurdle of establishing relevance at the site of interest. Risk assessors should consider the following when drawing inferences from studies conducted at other sites:

- *Type of contamination* – both COCs and the factors driving their bioavailability would ideally be similar between both sites. This would entail comparing data from the exposure assessment to that reported in the literature study. This is much easier for sites with one or two COCs.
- *Pattern of contamination* – this includes magnitude and spatial and temporal patterns. Ideally, the magnitude and scales of both studies would be similar.
- *Habitats and receptors* – site use patterns by receptors will vary according to available habitat types (i.e., due to their differing habitat characteristics related to

the animal's ecological needs). The configuration of high-use habitat types relative to the pattern of contamination will affect ROC exposure.

Once a study is deemed appropriate, its data would be extracted in a similar manner to that discussed for indirect toxicity information. For example, in the mercury study shown in the text box (Brasso and Cristol 2008), swallow reproductive success can be plotted against each of the mercury exposure measures (or simply the one most relevant to your risk assessment) to develop response profiles.

4.3. Receptor-Specific Considerations

> ***Key Concept:***
>
> The four types of effects assessment measures are not used equally among receptor groups in ERA. The relative frequency of use reflects current reality, which may not be ideal but often reflects limitations and challenges in application.

After having discussed each of the types of effects assessment measures, this section focuses on linking measures to receptor groups. **Table 4-2** shows the relative frequency of use of each of the major types of effects assessment measures in risk assessments[36]. It should be noted that any of these types of measures may be appropriate for a given risk assessment, so the table should be used only to provide initial insight into what is typically done. Selection ultimately depends on the specific needs of the risk assessment.

[36] This is based on the experience of the authors of this guidance document.

Table 4-2. Frequency of use of types of effects assessment measures for each receptor group.

Receptor Group	Site-specific Toxicological Studies	Indirect Toxicological Information	Site-specific Biological Studies	Indirect Biological Information
Terrestrial primary producers	moderate	moderate	moderate	low
Aquatic primary producers	high	high	moderate	low
Terrestrial invertebrates	moderate	moderate	moderate	low
Aquatic invertebrates	high	high	moderate	low
Fish	high	high	moderate	moderate
Birds and Mammals	rare	high	low	moderate
Amphibians and Reptiles	low	moderate	moderate	moderate

Categories defined as follows, based on experience and judgement of the authors:

• High – this rating was applied to both toxicological measures for aquatic ROC groups to reflect the long-term establishment, protocol development, and value for risk assessments. It was also applied to the indirect toxicological information measure for birds and mammals, due to our reliance on this measure in the face of cost and uncertainty of alternatives.

• Moderate – this was applied to both toxicological measures for terrestrial ROC groups to reflect the growing use of these measures. It was also ascribed to the indirect toxicological measure for amphibians and reptiles, mainly due to data limitations and the exclusion of these ROCs from many risk assessments (although use is increasing over time). Finally, it was applied to site-specific biological studies for most receptors to reflect the technical challenges associated with this type of measure.

• Low – this was applied to measures of indirect biological information for all ROC groups, largely reflecting the difficulty identifying studies that extrapolate well to the conditions of the site of interest (contamination pattern and relevant biology). It was also applied to site-specific biological studies for birds and mammals to reflect the cost/complexity of robust studies for discerning contaminant factors from physical or habitat factors. Note that use of reconnaissance surveys, habitat surveys, and semi-quantitative field measurements are more common in ERA, but convey greater uncertainty.

• Rare – this was applied only to site-specific toxicological studies for birds and mammals. Although they are possible to conduct, they are rarely (if ever in Canada) used due to a host of challenges including animal welfare issues (Suter et al. 2000)

4.4 Beyond Point Estimates of Toxicity

In many situations, the outcome of an effects assessment is the derivation of one or more thresholds for ecological effects. These thresholds are intended to represent the transition from an environmental exposure that does not elicit a meaningful ecological response to an exposure that conveys potential for ecological effects. Such thresholds can be developed for numerous media (soil concentration, sediment concentration, water concentration, tissue concentration, ingested dose), and are carried in the risk characterization where they are used to calculate hazard quotients.

> ***Key Concept:***
>
> Use of a point estimate, particularly if drawn from a single study, conveys high uncertainty. **Technical Module 2** provides guidance for reducing uncertainty in TRV development using relatively simple approaches applied to existing data.

A common problem encountered in ecological risk assessment is that a single threshold value is used to summarize the concentration-response relationship. In addition to the problem of over-simplifying a complex relationship, use of a point estimate is sensitive to the choice of statistical method or decision rule used to calculate the threshold. For example, use of a statistical significance criterion to discern between effect and no-effect levels of exposure can lead to substantial differences in the magnitude and/or significance of the threshold exposure level, in addition to other statistical and interpretative issues (Landis and Chapman 2011).

4.4.1. Considerations

It is desirable to move beyond the use of single point estimates for effects that commonly serve as the denominator in quotient methods. Although full quantitative integration of concentration-response relationships is not always possible, at minimum it is important for risk assessors to understand the true effect size (or range) that is represented by a TRV or other measure of effects, in part to facilitate selection of TRVs that are aligned with protection goals and AELs. Specifically, the risk assessment can be informed by consideration of:

- Effect size associated with the study that "drives" the toxicity threshold (e.g., water quality guideline, wildlife TRV dose).
- The difference between the NOAEL and LOAEL, or the steepness of the concentration-response where multiple exposure levels are tested.
- The degree to which the "most sensitive study" represents a larger number of experimental results, or alternatively represents an outlying response.

- Concordance of sensitivity for different receptor groups (such as domestic species versus wild organisms, passerines versus raptors, coldwater fish versus warm water fish).
- Concordance of short-term versus chronic test endpoints, or differences in sensitivity among various sublethal endpoints.

The above considerations cannot always be addressed in a quantitative manner. However, integrations of the relevant ranges of potential response are preferable to point-estimates.

Allard *et al.* (2010) recommended a meta-analysis approach to TRV derivation is preferred to results from single studies. This entails simultaneous consideration of numerous study results on a graph of effect size versus chemical concentration (see **Technical Module 2**). A graphical approach, while complicated by variations in endpoint type, exposure gradients, and study designs, helps to convey the variations in response at each exposure level.

4.4.2. Species Sensitivity Distributions (SSDs)

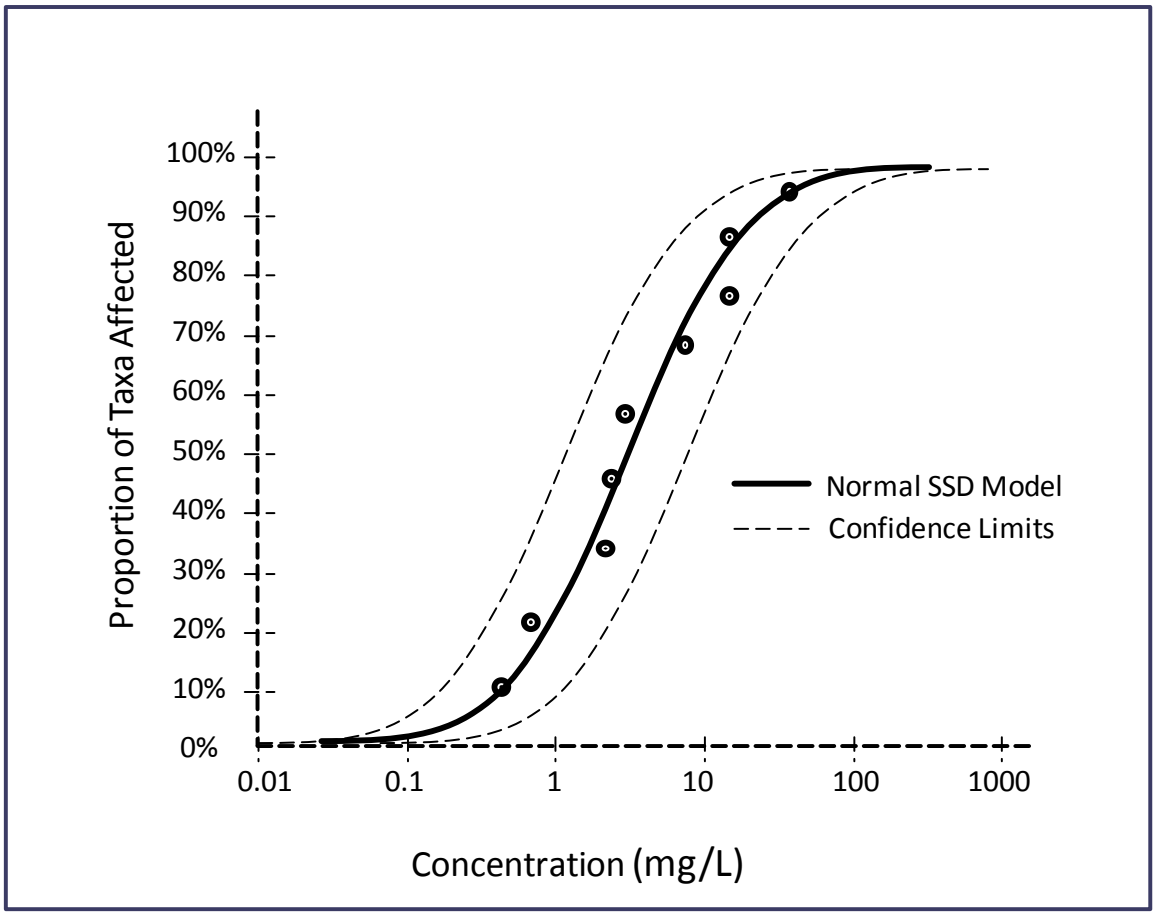
The species sensitivity distribution (SSD) concept is an example of a statistical approach to effects assessment that moves beyond the "traditional" approach to threshold development (i.e., use of the point estimate from the most sensitive study using a statistical significance criterion). For example, CCME (2007) recommends the SSD approach for water quality threshold derivations where a sufficient number, quality, and variety of toxicity test data are available.

In its usual usage, a species sensitivity distribution (SSD) is the cumulative probability distribution of some measure of toxicity of a certain chemical to a set of animal species (for more background see **Technical Module 1** of this document, SAB-CS [2008], Posthuma *et al.* [2002] and CCME [2007]). At increasing concentrations of a toxicant, the proportion of species affected (at a given level of effect, such as 20% growth impairment or 50% reduction of abundance) increases.

> ***Key Concept:***
>
> An SSD is a numerical expression of the ranges of organism sensitivity to a COC. An SSD can characterize variations among species, within species, and across taxonomic groups. Most importantly, the SSD concept conveys that individual taxa do not respond similarly to a single concentration.

A site-specific SSD is one example of a specialized application of site-specific correlation between concentration and response methods, whereby a site-specific SSD metric is related to contaminant concentration. The figure depicted below provides a hypothetical example of how the SSD concept may be applied.

The SSD approach recognizes that individual species may have highly variable sensitivities to a given COC (Kooijman 1997), and that protection of 100% of species is not necessarily required to protect the functional attributes of a community (e.g., benthic community). By combining results from multiple tests and covering a range of test organisms, it is possible to construct a distribution of sensitivities.



In the graphical example, the circles represent individual species[37] (for the purpose of this example, they may be assumed to be various freshwater epifauna). The x-axis depicts the chemical concentration (in logarithmic scale) at which a threshold response size occurs. The response type could be a 20% reduction in growth in a laboratory toxicity test, or could be a 50% reduction in species abundance in a benthic community study. The SSD entails derivation of a smoothed curve (solid line) and associated confidence limits

[37] Depending on the derivation details, individual data point may represent a single study for the species, or may be a summary metric integrating multiple studies, such as a geometric mean of multiple measurements.

(dashed lines) through the concentration-response curve. This allows an assessment of hazardous concentration at which a given proportion of species is affected (e.g., 20% of epifauna affected at 0.6 mg/L in the example).

To derive an SSD, single-species toxicity data (e.g., LC_{50} values, IC_X values, or LOAEC / NOAEC data) for many species are fit to a distribution such as the lognormal or log-logistic. From this distribution of species sensitivities, a hazardous concentration (HC_p) is identified at which a certain percentage (*p*) of all species is assumed to be affected (Postuma *et al.*, 2002). Selection by risk assessors of both the percentage of species and the effect level are in part matters of policy and may be considered as AEL determinations prior to risk characterization.

In addition to the more conventional application of SSD (application to literature data), it is also possible to apply the SSD procedure to resident biological communities. For the latter, it is necessary to identify a subset of the enumerated taxa for which there are sufficient numbers of organisms to assess potential concentration-response. Next, each retained taxon is assessed over the gradient of contamination and a benchmark level of response (such as 20% or 50% reduction in abundance) is evaluated. For each organism type, the concentration at which the threshold response is observed is documented, and the resulting concentrations are rank ordered. A hazardous concentration (HC_p) is then derived by choosing the interpolated COC concentration that matches the target percentage (*p*) of all species observed to be affected. This approach requires that the statistical power to detect the threshold level of response is considered, and as such is best suited to studies with a large number of sampling stations and a wide gradient of COC concentrations. Due to the high data demands for this approach, it is recommended only for advanced stages of risk assessment, and is less suitable where habitat variations are large relative to variations in contamination levels.

4.5. Safety Factors and Extrapolation

> ***Key Concept:***
>
> Safety factors can be useful in the development of conservative screening thresholds (such as generic environmental quality guidelines) but are not recommended for derivation of effects thresholds used in detailed risk assessments.

It is common practice in ERAs to collect effects information on an indicator organism/endpoint, and extrapolate the findings to the organisms of interest at a contaminated site. This is true for both literature-based evaluations (i.e., derivations of toxicity reference values from historical studies) and site-specific analyses (i.e., use of laboratory test species to represent potential responses in a broader array of local species).

Following historical practice in human health risk assessment, there has been frequent use of extrapolations among species and endpoints by applying various factors – so-called application, assessment, safety, or uncertainty factors (Chapman *et al.* 1998). These factors are intended to compensate for uncertainty in the effects analysis, and are conservative in nature, such that risks are not underestimated.

In past risk assessments, safety factors have been commonly applied to address several types of extrapolation in ecological risk assessment, including:

- extrapolation from test species to wild species;
- extrapolation from short-term to long-term exposures; and
- extrapolation from a significant biological effect to an insignificant magnitude or probability of effect.

Some jurisdictions have advocated the use of prescribed safety factors. Forbes and Calow (2002) summarize the commonly applied safety factors in Europe and the United States, although they note that these factors do not preclude the use of professional judgement.

This guidance does not advocate the application of safety factors in establishing toxicity reference values. Although safety factors can be useful in the development of conservative screening thresholds (such as generic environmental quality guidelines), their value diminishes greatly in quantitative risk evaluation. Some of the major disadvantages of safety factors include:

- Bias - their application is uni-directional, serving to increase risk estimates without consideration that the uncertainty may apply in both directions.
- Lack of technical basis - the standard default safety factors (commonly factors of 10) have a weak relationship to concentration-response information.
- Compounding conservatism - the application of multiple safety factors can result in predicted TRVs that are unrealistically low.
- Lack of transparency - application of safety factors buries the uncertainty such that the risk estimate is altered, but without a clear indication of the confidence (or lack thereof) in the numerical value.
- Incompatibility with newer methods - the application of arbitrary safety factors is poorly aligned with the application of methods (e.g., such as SSDs, concentration-response analysis, effect-size approaches) that are preferred for quantitative TRV development.

In recommending against the application of safety factors for derivation of effects thresholds, it is not intended that the uncertainty in effects assessment should be ignored. Rather, uncertainties should be evaluated thoroughly following guidance provided in **Section 5.6**.

5. RISK CHARACTERIZATION

5.1 Introduction

Risk characterization is the process of estimating the probability, magnitude, and extent of adverse ecological impacts based on the information obtained from the exposure and effects assessments. Risk characterization also provides the discussion of the "strengths, limitations and uncertainties arising from the data and models used to provide conclusions" (CCME 1996). Risk characterization is the stage where the various study components are integrated and interpreted in terms of overall significance for ecological risk. Risk characterization also translates complex scientific information into a format that is useful for risk managers, by conveying the ecological consequences of the risk estimates along with the associated uncertainties.

> ***Key Concept:***
>
> The risk characterization may draw together exposure and effects information for the first time, or may synthesize lines of evidence for which exposure and effects information have already been combined (see Figure 1-2) .

5.1.1. Purpose

The risk characterization merges the findings of the exposure assessment and effects assessment for each LOE, and integrates findings across multiple LOEs. As such, risk characterization techniques encompass all methods used to analyze and interpret the relationship between measures of exposure and measures of effect.

Provided that the problem formulation has been well designed, many aspects of risk characterization should be contemplated *a priori*, and integration of exposure and effect should be seamless and relatively mechanical. However, risk characterization entails more than simply merging exposure and effects information. Rather, it conveys the process by which numerous study results are evaluated to accomplish the following core objectives:

- Synthesize results from multiple measurements into a conclusion for each individual line of evidence, and synthesize conclusions from multiple lines of evidence into an overall conclusion regarding ecological risks.
- Provide a concluding narrative that presents conclusions in a clear and unambiguous manner. Where possible, conclusions are stated in plain-language, emphasizing clarity, such that risk assessment output can be used effectively by site managers in their decision making process.

- Evaluate the uncertainty in the conclusions, either qualitatively or quantitatively.
- Revisit the core questions/objectives of the study (which may have been framed as one or more study hypotheses), and provide conclusions in terms of these risk management objectives.

> ***Key Concept:***
>
> Risk assessments often have site-specific objectives (such as developing site-specific remediation standards, allocating observed effects to one or more sources, or predicting future risks under alternative management scenarios). Risk characterization summarizes the results of these study components.

The objectives summarized in the above bullets are less mechanical, and sometimes require the application of professional judgement[38]. In preparing a risk characterization, the practitioner should consider the broad assessment goals (see **Section 2.2.1.1**) to ensure that the content effectively answers the questions and addresses the hypotheses of interest.

5.1.2. Overview of Steps

The process of risk characterization includes the following steps, upon which the organization of the chapter is based:

- **Step 1: Conduct Relevance Check** – Following review of the data, a relevance check is conducted to determine whether any deviations occurred during field or lab studies that could affect the relevance of the data for supporting the LOE for which the data will be used. This step also provides an opportunity to identify adjustments that may be required to maintain the usefulness of the data for effective risk characterization (**Section 5.2)**;
- **Step 2: Interpret/Evaluate each LOE** – Selection of appropriate methods to evaluate and interpret the information generated during the risk assessment (**Section 5.3)**;
- **Step 3: Prepare Compiled Data Summary** – A summary presentation of the data for each line of evidence prior to application of detailed analyses (**Section 5.4)**;

[38] The role of professional judgement in ecological risk assessment is contentious. Application of professional judgement in interpreting or synthesizing technical information requires the practitioner to present a rationale that is transparent, clear, consistent, and reasonable, and should never be applied to circumvent or obscure sound decision-making (i.e., distortion or selective analysis of results to accommodate a desired outcome). The role of professional judgement is explored further in Step 4 of the risk characterization process.

- **Step 4: Apply Weight of Evidence Procedure** – Integration of the results of the multiple lines of evidence, using a weight-of-evidence framework established during problem formulation. Importantly, the weight of evidence procedure is inter-linked with Steps 5 to 8 below, and therefore Steps 4 to 8 are often implemented concurrently (**Section 5.5)**;
- **Step 5: Evaluate ERA Uncertainties** – Consideration of the uncertainties that affect the interpretation and/or reliability of each line of evidence (**Section 5.6)**;
- **Step 6: Consider Extrapolation / Interpolation** – Assessment of the degree to which risk conclusions drawn from a limited number of analyses can be expected to reliably translate to other conditions at the site (**Section 5.7)**;
- **Step 7: Development of Site-Specific Standards (Optional)** – development of numerical standards in site media that will be used to distinguish action levels for substances of concern (**Section 5.8)**;
- **Step 8: Summarize Risk Conclusions** – Preparation of a risk summary that characterizes risk in terms of potential magnitude of response and other key attributes (e.g., likelihood [probability], spatial extent, temporal extent, level[s] of organization potentially affected, causality, and other aspects of ecological relevance) (**Section 5.9)**; and
- **Step 9: Conduct Follow-Up Actions** – Preparation of clear recommendations and articulation of next steps for site closure, approvals, regulatory liaison, etc. (**Section 5.10)**.

Importantly, the steps in risk characterization do not infer any particular level of detail. For simple sites or sites where estimated risks are negligible, risk characterization does not need to be overly cumbersome, whereas for complex sites more detail and rigour will usually be warranted.

5.2. Step 1 – Conduct Relevance Check

As described previously, several aspects of the risk characterization are planned in the problem formulation stage including the Sampling & Analysis Plan (SAP). Strategic considerations described in **Section 2.3.5** therefore influence the way the risk characterization is conducted. Planning during the problem formulation stage must anticipate the important linkages between exposure and effects during risk characterization. Consistent with a philosophy of "beginning with the end in mind", it is expected that the PF/SAP is designed and implemented in a way that facilitates the effective integration of effects and exposure information. For this reason, strategic considerations will not be new at this stage, but rather should be revisited in light of the

findings from the implementation of the effects assessment and exposure assessment for each LOE.

5.2.1. Revisit the Overall Assessment Needs

Prior to conducting risk characterization, it is useful to revisit the key risk assessment questions posed in the problem formulation, and address the following issues in light of the data that have been collected:

- Confirm that the measures or techniques selected during the problem formulation remain the most effective and appropriate for addressing key risk assessment needs (such as evaluating causation). Where the relevance and value of some measurement endpoints can be assessed in advance, others require retrospective examination.
- Assess whether the analyses proposed in the problem formulation remain applicable to the assessment of testable hypotheses[39]. If studies could not be implemented as planned, or data quality considerations confound the application of the original methods, identify a modified approach that best meets the risk assessment needs, and explain the rationale for the modifications.
- Confirm that presentation methods defined in the problem formulation remain applicable and will provide output in a format useful to the risk manager for making decisions (i.e., data obtained are sufficient for the application of the proposed methods).

In the course of conducting a risk characterization, it is useful to revisit these issues prior to selection of specific analysis techniques, and prior to the investment of significant effort processing and synthesizing data. If the data collections were substantially compromised, it may be necessary to resample or add study components prior to proceeding with risk characterization. Although this decision may result in project delay, it is preferable to preparation of a risk assessment deliverable that does not properly address information needs for risk management.

> ***Key Concept:***
>
> In conducting the relevance check, the practitioner must determine whether the data quantity/quality are adequate to proceed, or alternatively that resampling is required to fill critical data gaps.

[39] This does not assume that classical hypothesis testing will be the only means of data analysis. Rather, the practitioner hypothesizes that certain effects may occur and attempts to determine whether or to what extent the evidence indicates effects.

5.2.2. Make Appropriate Modifications

At the time of risk characterization, the principal investigations have been completed, quality controls conducted, and a preliminary assessment of individual measurement endpoints has been completed. It is common to require revisions to the analysis plan based on a number of factors that diverge from the idealized plan specified at problem formulation stage. Accordingly, Step 1 of risk characterization serves as a relevance check to determine the consequences of such modifications, and to make appropriate adjustments if data collections did not work out as planned.

For example:

- If the problem formulation specified a gradient design in which effects measures were intended to be related to gradients in contamination, but the investigation failed to capture a useful gradient in contamination, it may not be possible to implement the statistical analyses contemplated in the problem formulation.
- If analysis of analytical data reveals data quality issues (e.g., negative control failure, interference effects, or protocol deviations), use of the data should be re-considered (e.g., the data may be given less 'weight' than originally envisioned, or eliminated entirely in the case of severe data quality failures).
- If the assumptions underlying statistical analyses are not satisfied (e.g., data distributions), then alternative methods of analysis may be needed.
- If community studies indicate significant variation in substrate or habitat type that confound analysis of contaminant-related effects, the strength of evidence from such studies may be lower than expected. In such cases, alternative statistical models may be useful in differentiating contaminant-related effects. If not, more weight may be given to other lines of evidence, or a different experimental design that controls for confounding variables may be appropriate.
- If the field data reveal new receptor groups or new exposure pathways that were not contemplated beforehand, additional analyses will be needed that were not considered in the problem formulation.

> ***Key Concept:***
>
> It is appropriate to modify the risk characterization approach based on constraints to acquisition of the data as originally planned, or if new information or methods have become available since the time the problem formulation was developed.
>
> It is not appropriate to modify a risk characterization approach simply because the results are not desirable or are unexpected.

Importantly, any modifications to the analysis plan should be considered based on whether measures and techniques planned in the problem formulation delivered usable results, but should not be made because the presence, magnitude, or type of environmental response (e.g., presence of toxicity,

patterns of community structure) differed from what the investigator suspected. Where significant changes to the analysis approach are proposed, it is important that the risk assessor document the deviations from what was expected, and provide a supporting rationale for any changes.

Modifications to the risk characterization may entail changes to specific methods to facilitate a meaningful analysis. For example, it may be necessary to consider data transformations if underlying assumptions of statistical methods (e.g., normality, stable variances, lack of high influence values [outliers]) are not met. In other cases, it may be possible to proceed with the original analyses, but with explicit acknowledgement of the reduced value of the analysis. For example, a benthic community study found to be confounded by mechanical disturbances of substrate may require a reduced weighting in the risk characterization (based on lower statistical power than was originally contemplated). In a terrestrial setting, a similar situation may arise where human disturbance of the landscape confounds the application of idealized sampling strategies intended to evaluate a soil contamination gradient.

In the broadest terms, Step 1entails the incorporation of important learning from the data collection stage, and fine-tunes the data analysis methods contemplated at problem formulation stage (and only as needed). The practitioner should not make arbitrary changes to analysis but rather link any required adjustments to the study goals and data quality objectives.

5.3. Step 2 – Interpret/Evaluate each LOE

After the measurement endpoints have been selected (during problem formulation) and applied, the investigator must apply tools to interpret the findings. These interpretations must be consistent with the informational needs of the risk manager, as outlined in the problem formulation. A proper problem formulation should have already identified how the data will be analyzed in order to support risk characterization (see **Table 2-7** as an example). The use of the data in

Key Concept:

The selection of specific methods is context-specific and cannot be prescribed. However, the following generic guidance applies:

It is desirable to retain available information (*i.e.*, hazard quotients should not be applied when concentration-response profiles are readily available and reduce uncertainty);

Given a choice between two methods of equal value for evaluating an assessment endpoint or reducing uncertainty, the simpler method is preferred;

Understanding of risk is improved by examining a measurement endpoint from multiple perspectives (i.e., multiple LOEs developed from a site-specific endpoint).

the WOE procedure (see Step 4) should guide how information should be summarized for individual LOE.

The following subsections summarize some of the common tools available to interpret individual LOE; it is not intended to be a comprehensive list nor a recommendation for the universal use of any specific tool. Furthermore, the tools are not meant to describe discrete options, and are not mutually exclusive.

All of the methods described in this section (Step 2) are tools that are applied to interpret the results from individual lines of evidence. The purpose of this section is to provide a discussion of the common procedures in their application, and to summarize the advantages and limitations of each. The methods are organized as follows:

- Hazard Quotients and other Quotient Methods – simple ratios of point estimates for both exposure and effects;
- Concentration-Response Relationships – using the mathematical relationship between site-specific exposure and response level to understand site-specific responses;
- Adjustment to Reference or Background Condition – Standardizing endpoint data to provide information on relative responses, rather than absolute responses only;
- Gradients – Patterns of responses over space (distance and direction) or over gradients in contamination;
- Multivariate Techniques – Interpretations of complex data sets through consideration of multiple factors simultaneously;
- Probabilistic Methods – Replacement of point estimates with distributions to provide more information on the range and likelihood of potential outcomes.

Note that some of the above techniques entail the replacement of point estimates with a more robust analysis of the available data. It is common to begin with point estimates during screening level risk assessments, beginning with a hazard quotient (HQ) approach. However, as HQs tend to incorporate conservative assumptions in the face of uncertainty, further evaluation is often needed following a screening assessment. In these situations, use of methods that make greater use of the range of exposure and effects information is encouraged. It is acceptable to proceed in a sequential (tiered) manner through a range of methods that replace conservative point estimates with ranges of values or distributions.

5.3.1. Hazard Quotients and Hazard Indices

The simplest tool for evaluating an LOE is a hazard quotient (HQ), which is the ratio between the exposure measure (numerator term) and a corresponding effect-based threshold (denominator term). HQs are widely applied, particularly in screening

assessments (e.g., PQRAs), due to the ease of application and the prevalence of point-estimate values for both exposure and effects. The HQ has particular value as a screening tool, which may be all that is required in some risk assessments. However, where HQs are calculated, care must be taken not to infer more information from the ratio than is warranted, and to consider the effect of uncertainty in both the numerator and denominator (**Section 5.3.1.1**).

Key Concept:

A hazard quotient is a ratio between an exposure term (dose or concentrations) and a response term:

$$HQ = \frac{Exposure}{Threshold\ Effect\ Level}$$

The exposure term for an HQ can be derived from many sources, including (see **Section 3** for details):

- A measured concentration in an environmental medium (e.g., mg/kg zinc in soil, mg/L selenium in water);
- A simulated concentration in abiotic environmental media or organism tissues using a model (which can range from a simple partitioning model to a complex mechanistic environmental fate and bioaccumulation model); or
- A modeled dose to an organism (mg/kg-day) from a food chain or trophic transfer model.

The threshold effects term can also be derived from numerous sources, including (see **Section 4** for details):

- An environmental quality guideline for abiotic media (soil, sediment, water, groundwater, etc.);
- A threshold value gleaned from a compendium of toxicological summaries;
- A threshold value obtained from an independent literature review;
- A threshold value (HC_X) from a species sensitivity distribution analysis;
- A site-specific threshold developed from interpreting the results of a toxicity or community study conducted over a range of exposure levels at the site of interest; or
- A meta-analysis of multiple sources of effects information (e.g., compilation of results from multiple studies that may cover a range of endpoints and species).

HQs may be applied for any of the four major categories of evidence. In practice, the most common HQs are derived for chemistry measurements in abiotic media (e.g., comparisons to soil, water or sediment quality guidelines), bioaccumulation endpoints (e.g., screening against tissue residue guidelines), and dose-based wildlife assessments

(i.e., dividing the estimated dose derived from a food-web model by a toxicity reference value [TRV]). However, an HQ can also be calculated for site-specific toxicology or community studies; threshold effects benchmarks can be calculated from concentration-response curves developed from site data[40], and then used for application to other stations or samples for which only chemistry data are available. Further discussion of the denominator in HQs is contained in **Section 4** (Effects Assessment).

5.3.1.1. Common Errors in Application

Although easy to derive, HQs are often misinterpreted (Allard *et al.* 2010). The most common error made is to incorrectly assume that an HQ is directly proportional to the magnitude of risk. HQs neither contain information about the specific probability that an adverse effect will occur, nor convey the magnitude of a potential adverse effect. Instead, a typical HQ is calculated using conservative assumptions, in which case the ratio indicates only whether existence of adverse effects is either possible (HQ > 1) or unlikely (HQ <1)[41]. In ecological risk assessment practice, there has been broad agreement that a hazard quotient ≤1.0 is indicative of negligible risk for the specified endpoint, because the HQ is usually calculated on the basis of conservative assumptions.

Another common error is to assume that HQs can be scaled across different COCs to provide reliable rankings of contaminant risk (Allard *et al.* 2010). However, as quotient methods are only as reliable as the values in the numerator and denominator (with associated uncertainty), the degree of hazard cannot be directly compared. The derivation methods for different COCs can result in large differences in conservatism that are masked by presentation of simple ratios. Similarly, separate HQ values for the same COC cannot be linearly scaled to risk (i.e., an HQ of 4 for APEC 1 cannot be assumed to be twice the risk of an HQ of 2 for APEC 2) because the intercept, slope and shape of the dose-response relationship is not reflected in the point estimate HQ. Reliable comparisons can only be made through detailed understanding of the underlying concentration-response relationships, safety (application) factors, and uncertainties, none of which are conveyed by an HQ.

Although a very large HQ suggests a greater "risk" than a HQ slightly greater than 1, it is not possible to draw conclusions about relative risk based on differences in HQs (e.g., HQs from 1 to 10 indicate moderate risk, while HQs above 10 indicate high risk), In

[40] Concentration-response profiles are discussed further in **Section 5.3.2**.

[41] In some cases, the threshold HQ is adjusted downward from 1.0 to 0.1, 0.2 or other values, to compensate for lack of data for background exposure. In general, these approaches are arbitrary and should be avoided (similar to arbitrary safety factors); instead, the uncertainty in total exposure estimates should be addressed explicitly.

addition, it is not true that minor changes in the HQ provide a meaningful differentiation (Allard *et al.* 2010; Ritter *et al.* 2002). For this reason, inclusion of excessive significant figures implies a level of certainty and precision that is not actually present; most HQs can be rounded to one or two significant figures.

5.3.1.2. Interpreting HQs

Because point estimates are applied in HQ derivation, the underlying uncertainty, bias, and variability in the data are masked[42]. Therefore, the interpretation of HQs requires explicit consideration of the selection of the numerator and denominator terms, including considerations of:

- Are the terms central tendencies, or conservative estimates intended to overstate risks?
- If conservative estimates are applied, are they based on "worst-case" assumptions (such as use of maximum observed concentrations and the most sensitive species)?

> ***Key Concept:***
>
> A hazard quotient is only as reliable as the information used to parameterize the numerator and denominator. As such there is no universal system for interpreting the magnitude of an HQ (beyond comparison to 1.0) and different types of HQs are not directly comparable.

- Is the effects term based on a NOAEL, LOAEL, or a threshold effect size, and was the threshold response level bounded in the study design used to derive the threshold?
- Were application factors (margins of safety) applied to the estimates to increase conservatism?
- Were thresholds derived from consideration of a broad range of studies and endpoints, or from limited data?
- Were exposures assessed through detailed profiling over space and time, or from isolated measurements?
- Were exposures estimated using uncertain models with high inherent uncertainty or conservatism?

[42] In some cases, quasi-probabilistic HQs that account for some uncertainties in the numerator or denominator may be calculated. This approach is rarely applied, but may be appropriate if reviewers or regulators prefer to evaluate the site solely using a quotient. Whereas it is always important to convey the types of uncertainty considered in a probabilistic assessment, such is particularly important for cases when probabilistic assessment is limited to the effects term or the exposure term.

In a screening-level assessment (PQRA), the standard approach is to apply conservative measures in both the numerator (upper bound estimate of exposure) and the denominator (lower-bound conservative guideline).

As a general rule, application of safety factors (application factors) in the calculation of hazard quotients is discouraged, as discussed in Section 4.5. Forbes and Calow (2002) and Chapman *et al.* (1998) discuss the pitfalls of assigning arbitrary or default safety factors in ecological risk assessments. Depiction of uncertainty in HQs is better handled through a separate uncertainty analysis that conveys the plausible range of risk estimates using different assumptions for exposure and/or effects parameters. This may be done probabilistically or through a bounding analysis.

5.3.1.3. Linkage of HQs to Spatial Units

An issue for the application of HQs relates to how they incorporate spatial variations in exposure levels. The procedures vary depending on the characteristics of the receptor under evaluation:

- For receptors with large home ranges, a single HQ can be calculated for the entire site[43]. This entails use of an exposure metric such as the arithmetic mean or the 95% upper confidence limit of the mean (UCLM) for all of the measured values for each medium, or the maximum measured concentration (Gilbert 1987). The degree of conservatism in the resulting HQ will depend on the metric used, the number of samples, and the variability among the samples.
- For sessile receptors or those with small home ranges, spatially distinct risk quotients can also be calculated depending on the spatial definition of the local population, and the probability of exceeding a hazard quotient of a given magnitude can be computed. This technique is generally applied when the single HQ method (screening assessment) yields a value above 1.0 and where the single HQ method is considered to be over-conservative.

Because these refinements still rely on the HQ as the underlying tool for evaluating risks, their primary use for highly contaminated sites may be to identify areas where more detailed evaluation of risks is warranted.

[43] If a wide-ranging ROC has specific habitat preferences that discourage use of portions of the site, the procedure described here can be modified by adjusting the exposure metric (e.g., exclusion of data from non-relevant habitats).

5.3.1.4. Hazard Indices and Multiple Substance HQs

> ***Key Concept:***
>
> A hazard index is the sum of the individual HQs for substances that have the same mechanism of toxic action.

The hazard index (HI) is a simple metric used to aggregate hazard from multiple substances. The HI is the sum of the individual hazard quotients for substances that have the same mechanism of toxic action[44]. The implicit assumption in HI calculation is that risks from multiple substances are additive when the mechanism of toxic action is similar. Because different pollutants may cause similar adverse health effects, it may be appropriate to combine hazard quotients associated with different substances.

As with the hazard quotient, aggregate exposures below a HI of 1.0 will likely not result in adverse responses. However, an HI greater than 1.0 does not necessarily suggest a likelihood of adverse effects. Furthermore, the HI cannot be translated to a probability that adverse effects will occur, and is not likely to be proportional to risk.

Combination of values through summation (hazard index approach) is not well-supported for most substances using existing toxicological data. There are two main reasons why hazard indexing is discouraged for most substances:

- Individual HQs are derived conservatively, such that summation of individual HQs compounds this conservatism; and,
- Summation of risks is appropriate only for contaminants that act via the same mechanism of action. Most contaminants exhibit different toxicological mechanisms, and so the scientific basis for calculating HIs for mixtures of contaminants is weak for most receptors.

> ***Key Concept:***
>
> Some approaches exist to integrate the hazard from groups of related substances. These approaches apply the hazard index concept, and have a mechanistic basis. Do not apply hazard indices where the evidence for common mechanism of action among substances is weak.

Where the mechanism of action is known, and relative toxicity of related substances can be quantified, approaches are available to integrate the effects of groups of related contaminants. These approaches apply to select groups of contaminants that are known or strongly believed to exert toxicity through a single mode of toxic action. For

[44] Note: Other indices have been developed by CCME to evaluate the potential effects of multiple contaminants. For example, the CCME water quality index (www.ccme.ca) evaluates water quality based on the number of contaminants exceeding CEQGs as well as the magnitude and frequency of those exceedances. The CCME water quality index scores water quality on a scale from 0 to 100 and categories scores from poor to excellent.

example, non-polar organic contaminants commonly exert direct toxicity via narcosis, a reversible state caused by non-specific interaction of lipophilic molecules with biological membranes (Escher and Hermens 2002). As a result, some guidelines have been developed that consider the cumulative effect of chemicals that act via this mechanism (Di Toro 2000a, 2000b). Furthermore, for some hydrophobic chlorinated organic substances believed to act via the aryl hydrocarbon (*Ah*) receptor (i.e., dioxins, furans, and a subset of PCB congeners), toxicity equivalence systems (TEQs) have been developed to simultaneously account for the relevant congeners once normalized to their receptor binding affinities. It is acceptable to apply these equivalence systems in the calculation of HQs, but not to add HQs developed using different systems (e.g., one should not add a total PCB HQ to a PCB TEQ HQ, or add a total dry weight PAH HQ to either an HQ derived using the narcosis model or to HQs calculated for individual PAHs). Where such systems exist, they are preferable to application of hazard indices, as the latter may not account for the mechanistic understanding of contaminant potencies in mixtures[45].

Where multiple contaminants are considered simultaneously, several assumptions may apply to derivation of the effects threshold, including:

- Concentrations of substances in the mixture are treated additively, with no assessment of relative toxic potential (e.g., total PAH threshold in sediment that does not discriminate among individual PAHs in the mixture);
- Concentrations of substances in the mixture are treated additively, and adjusted for relative potency (e.g., toxic equivalency systems for narcotic effects of PAHs in porewater, or for dioxins/furans through *Ah* receptor binding affinity); and,
- Concentrations of substances in the mixture are treated additively, but with a bioavailability correction prior to screening (e.g., molar difference between acid volatile sulphides and sum of simultaneously extractable metals).

However, most COCs do not have established methods for assessing the synergistic or antagonistic effects of interactions with other substances.

5.3.2. Using Concentration-Response Relationships

Concentration-response relationships are typically derived as part of an effects assessment, and the general methods of analysis are first described in the problem formulation. However, their use is almost always tailored to the data, which means that

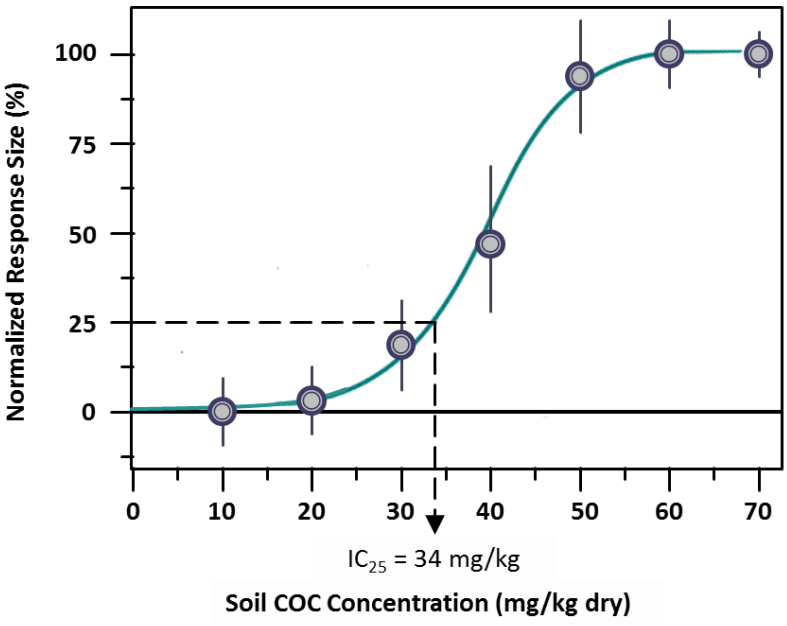
[45] Note that individual substances may have established putative effect levels that are based on empirical association (co-occurrence assessment) rather than relative potency established by mechanistic assessment. In the former case, summation of HQs is inappropriate.

details regarding analysis of concentration-response relationships are often part of risk characterization. For example, a concentration-response model that fits the data well at low concentrations but not at high concentrations may be acceptable if measured or estimated concentrations are low. Concentration-response relationships are presented here rather than in **Section 4** because of the emphasis on their use in practice.

5.3.2.1. Definition

A concentration-response relationship provides an assessment of the statistical relationship between an exposure term and a response term. Rather than provide a single threshold to describe the chemical potency of a COC, a concentration-response relationship describes the relationship between exposure and response over a range of exposure levels and effect sizes.

> ***Definition:***
>
> Concentration-response relationship – A mathematical assessment of how an exposure term relates to the observation of a biological or toxicological effect.



The above graphic depicts an example of a concentration-response relationship for a single experiment. In the example, the y-axis represents the response measure, which could be survival, growth, reproduction, or any other toxicological or biological measure at the organism, population or community level. The x-axis displays a range of exposure conditions under which the experiment was performed (in this example, seven evenly spaced treatment levels, with multiple replicates for each treatment, and variance for each treatment indicated by the error bars).

5.3.2.2. Advantages

The information contained in the above graphic is substantially greater than what is conveyed through use of a point estimate, such as the NOAEL/LOAEL, for several reasons, including:

- The response magnitude is defined at multiple exposure concentrations. For example, whereas a 50% response occurs at approximately 40 mg/kg, the dashed line shows that a 25% response occurs at approximately 34 mg/kg.
- The response curve illustrates the steepness of the response profile (total inhibition of response occurs within a factor of 3 of the concentration showing negligible response); and
- The variability among replicates is depicted, providing an indication of the variation of the response (part of the uncertainty).

The preferred procedure is to perform a direct estimate of effects / risks using concentration-response analysis. The key difference is that a quotient compares exposure to a point estimate for effects information (e.g., a TRV), whereas a true risk estimate explicitly evaluates the mathematical relationship between site-specific exposure and response level across a range of relevant exposures (i.e., mathematical evaluation of response magnitude versus chemical concentration or dose).

In the above example, the 25% effect level (34 mg/kg) could be used to derive a TRV. Alternatively, a more stringent response magnitude, such as the 10% effect level (25 mg/kg) could be applied. By plotting measured or estimated exposure on the curve, the estimated response can be understood for exposures greater than or less than these exposure values. Without the curve, it is not possible to understand the magnitude of response associated with the measured or estimated exposure.

5.3.2.3. Disadvantages

Although conceptually attractive, there are some limitations to the application of concentration-response models, including:

- In natural systems, it is rare that a clear relationship is observed between exposure and response (whether linear, sigmoidal, or other shape) – frequently, the relationship is an "interrupted" concentration response in which one or more treatments does not follow a smooth pattern[46];

[46] An idealized concentration-response relationship is shown in the figure to facilitate understanding of the approach.

- Data limitations (such as limited exposure levels, or lack of information specific to the species of interest) restrict the application of the method;
- The relationship between exposure and response can be confounded by other factors – in the example figure in the previous section, it is plausible that the response was caused by a factor that covaried with the COC concentration, such as soil pH or mean particle size;
- Concentration-response curves are challenging (and expensive) to derive on a site-specific basis due to the number of treatments and replicates required to achieve confidence in the relationship;
- Numerous mathematical functions are available to quantify the relationship, and selection of the appropriate function can be challenging, especially when used to extrapolate beyond the range of measured exposures. Formal methods of model selection based on Akaike's Information Criterion (AIC) and other criteria are available and should be considered (Burnham and Anderson 2002);
- Where concentration-response data are derived from the literature, care must be exercised to ensure that results are transferable to the context of the site. For example, if the example relationship was based on well-drained soils, but the site consisted of bog-like conditions, the relationship implied by the curve may be inapplicable to the site context. Similar considerations apply to other chemical-specific factors (e.g., metal speciation, modifying factors such as dissolved organic matter) and also to biological factors (representativeness of surrogate organism, physiological tolerance of local organisms).

Many of these disadvantages stem from data limitations. Because data limitations also affect our ability to derive point-estimate TRVs and HQs, practitioners should not be discouraged from exploring concentration-response relationships simply because data are limited – those same data limitations will carry large uncertainties regardless of what methods are used for effects assessment and risk characterization.

5.3.2.4. Application

The 'true' risk estimate (based on concentration-response profiling) is what was envisioned as 'detailed' risk characterization by CCME (1996). This is consistent with the knowledge that quotient-based methods do not provide estimates of risk, because they cannot characterize the probability and magnitude of effects. In some cases, quotient-based approaches have been applied even beyond screening-level risk assessments, due to several factors including limited data for understanding concentration-response relationships. However, there are many cases where data are adequate for supporting concentration-response analysis, and practitioners should aim to analyze concentration-

response data explicitly whenever possible for detailed risk assessments. For some measurement endpoints (e.g., aquatic toxicity tests conducted using dilution series), characterization of the concentration-response relationship is a natural outcome of the test results.

In addition to their use for risk estimation, site-specific concentration-response data have another use, which is to establish site-specific correlation or insight into causality (causality and its role in risk characterization is explored in detail in **Section 5.5**). Site-specific correlative approaches evaluate the association between contaminant concentrations and levels of response; they include formal statistical association methods and qualitative evaluations. Levels of response can be derived from any measurement endpoint used in a risk assessment, from a toxicity test endpoint (e.g., growth or reproduction in a lab bioassay) to a direct community measure (e.g., total organisms or total taxa measured in a benthic invertebrate sample).

> ***Key Concept:***
>
> Development of concentration-response curves requires understanding of the underlying statistical assumptions. Practitioners should consult toxicity test protocols and/or a biostatistician when applying statistical models.

Regardless of how concentration-response data are used, quantitative models that relate responses to any predictors should be appropriate for the data. For example, dichotomous outcomes (e.g., survival in a toxicity test) should usually be evaluated using a generalized linear model (e.g., logistic regression) that assumes the correct (binomial) error structure for the data. In addition, models that are fit to grouped data (e.g., dilution series bioassay results from more than one sample station) should use methods that account for the structured nature of the data (Pinheiro and Bates 2000; Wheeler and Bailer 2009). In short, advanced statistical methods beyond simple linear regression are often necessary, and can facilitate evaluation of concentration-response relationships while simultaneously explicitly considering the influence of categorical and continuous factors on the nature of the relationship.

5.3.3. Use of Reference or Background Condition

Many measurement endpoints pertaining to biological or toxicological parameters cannot be readily interpreted at face value (e.g., a species richness value of 12 has little meaning until placed in ecological context), but rather require comparison to a reference condition if a gradient design (**Section 5.3.4**) is not used or not feasible (see **Section 2.3.5.1** for further discussion). Accordingly, an important tool for risk characterization is the control-impact design or other comparative approaches. The general experimental design is established during problem formulation – this section focuses on application and use of the data during risk characterization.

There are several types of samples that can be used to standardize site responses, including:

> ***Key Concept:***
>
> In terms of utility for risk characterization, reference and background stations are preferred to negative control comparisons. References must be confirmed to be uncontaminated and well matched to the site conditions, preferably with sufficient sample sizes to evaluate variability and to conduct statistical tests.

- Negative Controls – Clean artificial substrate or test media used in the laboratory to evaluate test acceptability. These are not recommended for standardizing site responses, as lab conditions may be unrepresentative of the environment relevant to the site[47];
- Reference Condition – Media collected in the general vicinity of the site, but confirmed to be less contaminated relative to site media; and,
- Background Condition – Media collected from the region at stations known to exhibit a lack of incremental contamination beyond naturally occurring concentrations.

For example, if a quotient approach is used to evaluate a particular LOE, it can be useful to compare the quotients derived for on-site and compare them to quotients derived for a range of off-site conditions. Sometimes, the difference in risks between a site and a reference condition are as important to understand as the absolute magnitude of estimated risks, particularly when conservative assumptions are built into a risk assessment. For example, in some environments, natural mineralization can elevate regional background concentrations above screening values and generate "false positive" HQ values for metals that exceed 1.0.

For assessments of risks to wildlife that are based on total dose, comparative approaches are particularly useful for identifying how various exposure media are contributing to the incremental risk on-site compared to off-site.

One particular application of a control-impact design, based on application of multivariate methods, is the Reference Condition Approach (RCA; also called the Reference Envelope Approach). This procedure can be applied to both toxicity and field community studies. The procedure has been proposed as an alternative approach to overcome limitations of reference and negative control samples. These limitations include differences in non-contaminant characteristics (substrate, habitat, etc.) and low statistical power when many samples are compared to a single control or reference. The RCA for benthic invertebrate sampling (Reynoldson *et al.* 1997) selects multiple reference sites

[47] Negative control media are primarily intended to evaluate the sensitivity of test organisms to handling and manipulation; as such the substrates are often simplified or artificial (e.g., silica sand) unless the laboratory has adopted a natural substrate. Practitioners should consult with laboratories prior to testing and consider the use of additional clean controls better matched to site conditions.

from a reference database as the "control" and individual test sites as the treatment, and applies multivariate ordination methods (such as non-metric multidimensional scaling) to distinguish patterns among samples. Confidence bands (ovals) around the data (**Figure 5-1**) indicate the degree of statistical similarity of test samples in relation to the suite of references. This approach is currently the basis of the Canadian national aquatic bio-monitoring program (CABIN) (Reynoldson *et al.* 2006) and has been promoted in southern California for the interpretation of biological data (SWAMP 2009). A similar approach has been developed by Environment Canada to capture the variation within the reference toxicity endpoints in the Great Lakes region; toxicological data are first range standardized, and then Euclidean distance (a multivariate similarity measure) is applied as the distance coefficient. In applying the reference approach, consistent application of the selection criteria for reference areas is required, and an ecological relevance check is required to ensure that reference stations are appropriately matched to exposed stations in terms of key environmental variables (organic enrichment, substrate type, depth, etc.).

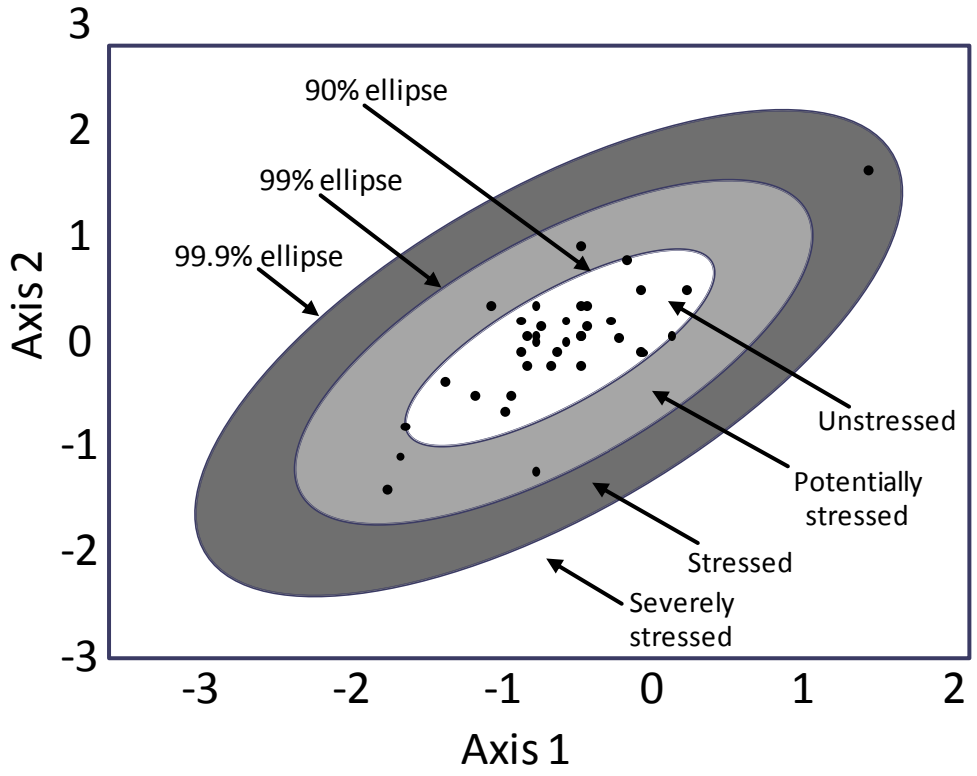
5.3.4. Gradient Designs

Given the challenges of determining reference conditions against which site conditions can be compared, experimental designs based on gradients should be considered whenever possible, and the specific design should be determined as part of problem formulation (see **Section 2.3.5** for more discussion). For example, if there is a historical point source of contaminants, it may be useful to correlate response measures to distance from that point source. Alternatively, if contaminant concentrations are known, the gradient may simply be based on categorizing spatial units according to contaminant concentrations. If a gradient design is envisaged in a risk assessment, it should consider how to best align the sampling design with the fate and transport pathways. For example, identification of spatial gradients may consider the following:

- distance or direction from a known source;
- historically observed gradients in contaminant concentrations; and
- interaction with physical factors such as water depth, salinity, and substrate type.

> ***Key Concept:***
>
> In examining potential gradients, practitioners may need to consider information other than raw COC concentrations, such as factors influencing bioavailability (organic carbon, coal particles, sulphides) or physical factors (habitat, substrate).

Figure 5-1. Depiction of the Reference Condition Approach for Invertebrate Communities (Reynoldson *et al.* 2006)



Invertebrate communities at test sites that fall within the 90% probability ellipse are considered equivalent to reference sites; within the 99% probability ellipse are possibly different; within the 99.9% probability ellipse are different; and outside the 99.9% probability ellipse are very different

In applying the gradient approach, it is desirable to provide representation of a wide range of exposure levels, ranging from exposures at or near the background condition to "worst-case" conditions found at the site. The greater the range in exposure concentrations, the better the ability to characterize a concentration-response relationship. If gradients are weak or poorly defined, additional uncertainty will be incorporated in the assessment of responses. Furthermore, if the range of exposure levels is small, natural variability may obscure a meaningful underlying relationship that would be revealed if there was greater variety of exposure conditions.

5.3.5. Multivariate Techniques

5.3.5.1. Definition

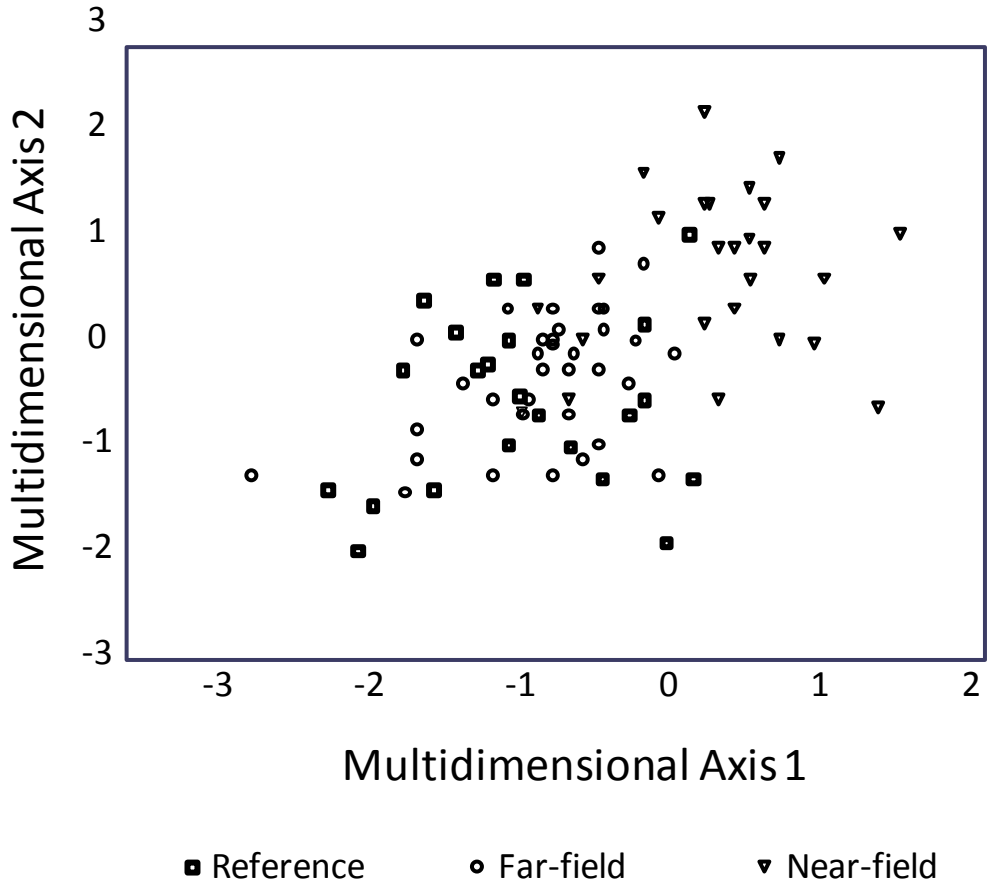
Multivariate statistical analysis refers to any of various statistical methods for analyzing more than two variables simultaneously. Assessing effects at a community or ecosystem level usually involves measuring a large number of abiotic and biotic variables. Assessing each variable individually or with many pairwise bivariate analyses can be cumbersome, difficult to interpret, and cannot detect patterns that emerge from the interactions of variables. Multivariate techniques can be used to summarize overall patterns from a large suite of variables (Bier 1999; Environment Canada 2002; Fairbrother and Bennett 2000; Sparks *et al.* 1999; USEPA 2007c). Once the number of variables has been reduced, patterns in the data can be evaluated and compared to other data (e.g., if a chemistry data set is reduced to a couple of summary variables, those variables could be correlated to toxicity data using multiple regression or similar techniques).

> ***Key Concept:***
>
> Multivariate methods are designed to simplify complex data sets with numerous individual parameters into a smaller number of variables that explains most of the variability, while being simpler to understand.

While general multivariate techniques may be discussed during problem formulation, it is often not until the data are evaluated that the details of analysis can be specified. Broad types of applications for multivariate techniques in risk characterization include, but are not necessarily limited to: ordination; clustering / discrimination; and investigating relationships between sets of variables (correspondence). The appendices of SAB-CS (2008) provide an overview of the common multivariate statistical approaches and identify potential pitfalls with their application (Landis *et al.* [2011] also discuss potential pitfalls). See Sparks *et al.* (1999) for more information on specific techniques as they have been applied to risk assessment. Because of the complexity of multivariate

approaches relative to univariate statistics, a qualified statistician with experience in biological or ecological investigations should be consulted.

- **Ordination techniques** (e.g., principal components analyses [PCA]) reduce a large set of variables into a smaller set of factors, each of which is a combination of variables that captures as much as possible of the information in the original variables. In this way, a multidimensional set of data can be reduced into a more interpretable form.
- **Clustering / discrimination techniques** identify natural groupings among sampling units (e.g., most-similar groups of sampling sites) and the parameters that contribute most to this similarity (e.g., abundances of certain species).
- **Correspondence analysis techniques** (e.g., canonical correspondence analyses [CCA]) identify the degree of covariance between sets of variables (e.g., concentrations of several chemicals versus abundances of several species), as well as identifying the variables within each set that contribute most to this covariance.



In the above example, non-metric multidimensional scaling (NMDS) has been applied to soil invertebrate community data, using abundance of major taxonomic groups as inputs. The resulting NMDS reduces the dimensionality to 2. In this example, the far-field stations (presumed to be defined using distance or concentration measures) are highly

overlapping with the reference condition, whereas the near-field stations are only partly overlapping. To interpret the axes, it would be necessary to correlate the axis scores with individual variables.

5.3.5.2. Advantages and Disadvantages

Multivariate methods reviewed above are aimed primarily at data exploration, and are usually used to reveal patterns that warrant more specific quantitative evaluation.. They distill complex data sets down to a low number of dimensions (usually two or three) that capture the main sources of variation in the data. Multivariate approaches are amenable to graphical presentation of results (e.g., cluster analysis dendograms, ordination plots) that are often intuitive relative to a large stream of univariate plots (e.g., intercorrelation matrix). These advantages must be traded off against the following drawbacks:

- Complexity in interpretation and communication of findings, and the need for thorough evaluation of the underlying assumptions of the statistical procedures applied;
- Multivariate methods are usually exploratory, and therefore cannot be defined in detail prior to the acquisition of the data (e.g., one cannot define the number of required dimensions *a priori*);
- Environmental data are prone to violations of parametric statistical frameworks (e.g., normality of distributions, independence among inputs, etc.), requiring great care in application and interpretation, or use of non-parametric techniques;
- Output of some multivariate methods cannot easily be translated to decision rules for ecological significance. For example, the axes of a PCA ordination do not have defined units and therefore differences in any dimension are challenging to interpret in terms of environmental relevance;
- Some methods are sensitive to data constraints such as missing values and non-detected concentration data;
- The meaning of each axis must be evaluated using correlations with the individual inputs.

With respect to interpretation of findings, a significant issue for risk characterization is how to score and weight the findings of ordination methods. The results of these techniques are not conducive to an IC_{20} or other effect-size based categorization. The output is useful for identifying effects (relative differences among stations, station groupings, or relative to reference), but interpretation of the ecological significance is more challenging. Determination of whether the differences among stations are ecologically meaningful requires a two stage evaluation:

- Analysis of the factors/variables that caused the observed divergence in ordination (e.g., which taxa are more or less common at extremes of each NMDS axis).
- Assessment of the functional importance of these differences in terms of community health. This step requires professional judgement, as it entails discerning between observed differences (effects, not necessarily a negative phenomenon) and impacts (adverse effects that indicate degradation of the community).

One specific application of multivariate methods (the Reference Condition Approach) was elaborated earlier in this section.

5.3.6. Probabilistic Methods

> ***Key Concept:***
>
> Probabilistic methods replace point estimates with distributions. These methods may simulate the effect of natural variations (stochasticity), uncertainty in knowledge (incertitude), or a combination of both.

Probabilistic methods acknowledge that natural ecological features are not constants, but rather are variable and complex, and that our understanding of their properties is not complete. Probabilistic models describe the state of one or more random variables as a distribution of possible values rather than fixed values (point estimates). Using probabilistic methods, important biological, chemical, physical, and environmental parameters are assumed to vary or are uncertain and therefore are specified using distributions.

Most ecological risk assessments are conducted using point estimates for exposure and effects parameters. This is acceptable for many assessments (e.g., preliminary assessments) because use of point estimates with appropriate conservatism to account for uncertainty can effectively screen numerous pathways with relatively little effort. However, for residual risks it is sometimes difficult to ascertain the influence of compounding conservatism on the risk assessment. Additionally, there are some parameters for which it is difficult to incorporate conservatism, because the degree to which a parameter is conservative depends on how it is applied.

For example, a specific dietary preference (such as consumption of fish by mink) can be increased to err on the side of overestimating exposures when applied in a forward modeling mode (because fish tend to have higher concentrations of contaminants relative to other food items). However, if the purpose of the risk assessment is to identify threshold concentrations in various dietary items, such that the total daily dose is no larger than the TRV, the situation is more complex. Specifically, the calculation of total blended ingestion rate would be biased toward the fish pathway, such that the sensitivity to changes in aquatic contamination would be exaggerated, whereas changes to terrestrial

contamination would be understated. In this situation, intentional use of a high-end point estimate for dietary preferences could result in a management decision for soil (e.g., threshold soil concentration derivation) that is contrary to the objective of conservatism. Probabilistic methods can help to resolve such problems by representing parameters as a range of plausible values rather than relying upon the relevance of a point estimate.

Probabilistic methods can be used when applying the quotient method, or when investigating concentration-response information (i.e., estimates of actual risk). In the former case, the result may be a probability distribution of quotients that allows estimation of the probability that HQ > 1. In the latter case, the result of a probabilistic assessment may be a probability distribution of effect rates – integration of this distribution provides an estimate of the *expected* risk, rather than the maximum likelihood estimate of risk. Probabilistic risk assessment explicitly acknowledges the stochastic and/or uncertain nature of model parameters, and attempts to describe the effect of multiple and linked parameter distributions. Probabilistic methods can be applied separately during the exposure assessment or effects assessment, but can also be applied during risk characterization. Some details regarding probabilistic methods are explored in **Section 5.6** in the context of evaluating uncertainties. Additional guidance and references are summarized by Suter (2007).

5.4. Step 3 – Prepare Compiled Data Summary

A relatively simple but effective risk characterization tool is to provide a simplified data summary. The intent is simply to summarize the range of endpoint data (without any sophisticated interpretation), with results for multiple endpoints organized by sampling station, habitat type, or management unit. This table can be referenced by the risk assessor (or a reviewer) during risk characterization. The compiled data provide useful reference material that may be lost in a complicated WOE process (**Section 5.5**). An example of a compiled data summary (simplified) is provided below. The data (normalized to the reference conditions and guidelines) are placed into categories of response, with no additional interpretation provided.

	Chemistry		Toxicity		Community	
Station ID	Metals	PAHs	Amphipod	Bivalve	Benthic Abundance	Taxa Richness
Endpoint	As | Cu	LPAH|HPAH|TPAH	Survival | Reburial	Survival | Development	Total | Sensitive	Number of unique taxa
NF-1	○ | ❷	❷ | ○ | ○	○ | ○	○ | ○	○ | ○	○
NF-2	❷ | ❸	❷ | ○ | ○	○ | ○	○ | ○	○ | ○	○
FF-1	❷ | ○	❸ | ○ | ❸	○ | ○	○ | ○	○ | ●	●
FF-2	❸ | ❷	❷ | ○ | ○	○ | ○	○ | ○	○ | ○	○
FF-3	❷ | ❷	❷ | ○ | ○	○ | ○	○ | ○	○ | ○	○
FF-4	○ | ○	❸ | ○ | ○	○ | ○	○ | ○	○ | ○	○

- Chemistry Data: ○ indicates below SQG; ❸ indicates above SQG, with number in symbol representing degree of exceedance.
- Toxicity Data: ○ indicates negligible to low effect size (below 20%); ○ indicates moderate effect size (20-50%); ● indicates high effect size (>50%) (all relative to reference); and
- Benthic Data: ○ indicates negligible to low effect size (below 20%); ○ indicates moderate effect size (20-50%); ● indicates high effect size (>50%) (all relative to reference).

The above table provides details on a sample by sample basis, but is nevertheless simplified in two ways. First, it bins the raw data into categories rather than reporting actual effect sizes. Second, it presents only the information on magnitude of effects, not information on causality, uncertainty, ecological relevance or any other attribute that may be relevant for evaluating LOEs. This simplified table may be most appropriate for cases where data indicate minimal risks, or where the complexity of the response profile is low.

An alternative to the above table, more applicable in cases where data show significant and/or complex indications of responses, is to present the absolute values of the endpoint responses (with numerical values and no categorization) and to also present raw information on evaluation of causality (in anticipation of supporting the WOE assessment outlined in **Section 5.5** below). An example format for a single LOE (soil invertebrate richness) is shown below. A summary of this type attempts to present information at face value without complex interpretation. In this case, results are usually not presented on a sample by sample basis but for an entire site or portion of a site.

The challenge is to present a condensed version of the field results (for simplicity or review) without introducing excessive manipulation of the data or professional judgements. Data summaries may vary in scope, but their role in facilitating review by regulators and others should not be underestimated. The two formats presented here have

advantages and disadvantages as described above – for complex ERAs use of both formats may be appropriate.

LOE	Magnitude	Uncertainty about Magnitude	Evidence for Causality	Uncertainty about Causality
Soil invertebrate community richness	Average richness 15% lower compared to reference condition	t-test not significant (p=0.22) but sample size limited; reference condition based on only 3 sites	Linear regression indicates richness weakly inversely correlated with soil zinc concentrations	Regression not significant (p=0.48) and explains little of the variation (r^2=0.08). The best predictor of richness is soil moisture (marginally significant at p=0.09).

5.5. Step 4 – Weight of Evidence Procedure

The term weight-of-evidence (WOE) is defined here to mean "any process used to aggregate information from different lines of scientific evidence to render a conclusion regarding the probability and magnitude of harm". This definition encompasses a range of practice, ranging from best professional judgement (BPJ) assessments to complex quantitative methods. A default WOE procedure is prescribed (see text box below) and will be applicable to most sites.

Key Concept:

The default procedure recommended in this guidance involves the following steps:

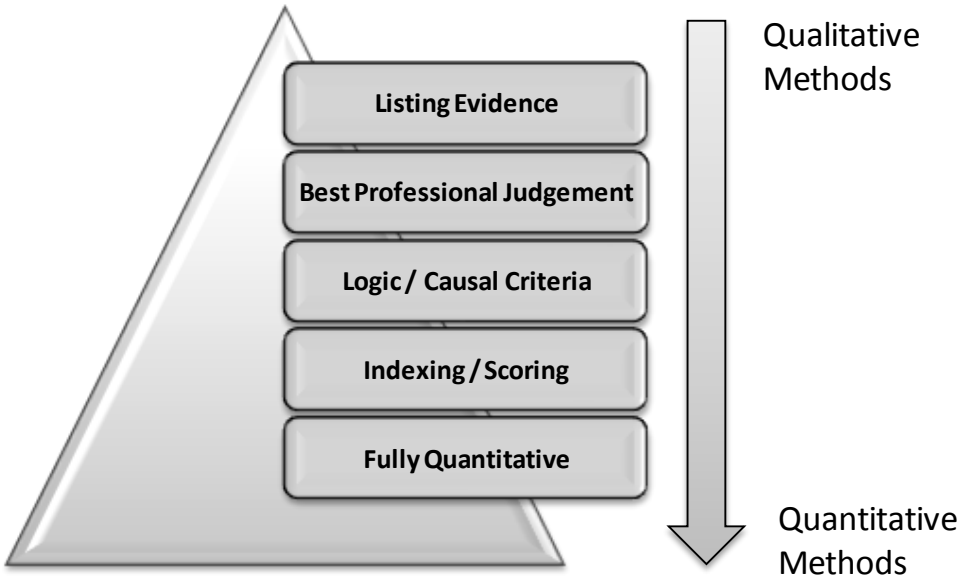
1. Summarize each LOE based on (a) magnitude of effects (including spatial extent), (b) evidence for causal relationships between contaminants and effects, and (c) ecological relevance. The methods of scoring or ranking each of these attributes should be established in advance (e.g., Table 5-1). The final LOE summary tables should be organized by assessment endpoint; examples for typical LOE are provided in Tables 5-2 (terrestrial) and 5-3 (aquatic).
2. As part of the LOE summary, evaluate uncertainty regarding magnitude of effects and evidence for causality for each LOE (uncertainty is evaluated more broadly in Step 5 following the WOE procedure, but must also be evaluated here to characterize specific uncertainty regarding magnitude and causality for each LOE).
3. For each assessment endpoint, make an integrated evaluation of findings for all of the LOEs, taking into account the degree of concordance among the various LOEs for that assessment endpoint (i.e., do the LOEs tell the same story?). The integrated evaluation should be based on a narrative rationale that clearly articulates how the overall evaluation was derived.

During the WOE step, the results for the individual lines of evidence obtained in Step 2 (and summarized in Step 3) are integrated. This provides a basic structure for all WOE assessments that provides a degree of consistency and transparency necessary for technical review of the document. The following subsections provide rationale for and details of the recommended default procedure for conducting WOE evaluations, as presented in the above text box.

5.5.1. Frame Purpose and Type of WOE

This guidance document provides a default WOE approach that builds upon the prescribed three step WOE approach described in the text box above. The default WOE approach described below is likely to be applicable for most federal sites. Other WOE approaches have been described in the literature and may be used if suited better for specific sites or specific types of ERAs. Linkov *et al.* (2009) provide a simplified, but useful, summary of the range of WOE methods available (**Figure 5-2**). The authors note that although all WOE methods may include both qualitative and quantitative considerations, the methods can be ordered by increasing degree of quantification along a continuum.

Figure 5-2. Classification of WOE Approaches in Risk Assessment (Linkov *et al.* 2009)



Method	Method Description
Listing Evidence	Presentation of individual lines of evidence without attempt at integration
Best Professional Judgement	Qualitative integration of multiple lines of evidence
Causal Criteria	Criteria-based methodology for determining cause and effect relationships
Logic	Standardized evaluation of individual lines of evidence basic on qualitative logic models
Scoring	Quantitative integration of multiple lines of evidence using simple weighting or ranking
Indexing	Integration of lines of evidence into a single measure based on empirical models
Quantification	Integrated assessment using formal decision analysis and statistical methods

An intermediate degree of quantification is likely to be appropriate for most cases (Suter and Cormier 2011), and therefore is recommended as a starting point for FCSAP sites. The most qualitative approaches and the most quantitative approaches (i.e., the extremes along the continuum depicted in **Figure 5-2**) may not often be appropriate because the former do not provide a transparent system of reaching integrated conclusions, whereas the latter can be difficult for risk managers or stakeholders to understand due to computational complexity. This does not mean that such approaches may never be applied; rather it requires that rationales be provided where the "extremes" are chosen, and that consideration should be given to the potential weaknesses of these approaches during implementation.

The following broad principles should be incorporated into the default WOE procedure or any alternative procedure that is used:

- The integration of LOEs should be performed in the context of the assessment endpoints and environmental protection goals. Specifically, the WOE needs to consider the level of organization of interest (individual, population, community) and explicitly address the linkage of the various LOEs to that level.
- The magnitudes of response observed for various measurement endpoints should be evaluated using rules that are as consistent as possible, such that various LOEs are compared using compatible decision criteria.
- The concurrence or divergence among outcomes of multiple measurement endpoints should be carefully evaluated.
- WOE determinations may be quantitative or qualitative, but should always be transparent.
- Professional judgement may be exercised, but a transparent analysis should be applied to elucidate the influence of professional judgement on the results.
- The degree of confidence in the conclusion for each endpoint is nearly as important as the conclusion itself.

Put more succinctly, when presenting the results of an assessment, the risk assessor should strive for the achievement of the following principles (TCCR; USEPA 2000):

- Transparency;
- Clarity;
- Consistency; and
- Reasonableness.

These considerations, although difficult to quantify, are mandatory aspects of the FSCAP WOE procedure.

5.5.2. Major Attributes Used to Evaluate LOEs

The application of WOE is based in part on the consideration of attributes that are used to evaluate each LOE. The recommended default WOE procedure for federal sites considers the following attributes[48]:

1. **Magnitude of response** (including effect size, probability of occurrence, spatial scale, temporal scale) and associated uncertainty.
2. **Evidence for causality** (i.e., is the observed response likely to be associated with site-related contaminants) and associated uncertainty;
3. **Ecological relevance** (i.e., to what extent does the LOE represent the assessment endpoint of interest).

Each of these is discussed below. This list of attributes emphasizes the importance of both magnitude and causality, although to some extent they can be evaluated sequentially – if there is zero magnitude of response (i.e., no effects), there is no need to look for a cause. Conversely, if a large response is measured, evaluation of causality is of critical importance (for further discussion see Technical Module 4[49]; Landis *et al.* 2011; Hull and Swanson 2006; Suter 2007; Suter *et al.* 2010).

For the default WOE procedure, uncertainty regarding the magnitude of response and evidence for causality is not specified as a stand-alone attribute *per se*, but is an integral component of the evaluation of each LOE. Uncertainty is a function of many factors including the quality of the data, the ability of the LOE to detect effect sizes of interest, the degree to which responses are specific to the stressors of interest, and the spatial and temporal representativeness of the data. Several of these factors were listed in **Section 2.3.4.3** in the context of LOE selection[50]. A thorough WOE evaluation of uncertainty must consider these factors.

Importantly, ecological relevance and some of the factors driving uncertainty are considered not only during risk characterization, but also during problem formulation (see **Section 2.3.4.3**). Specifically, these considerations may serve as criteria for the selection of measurement endpoints and LOEs. For ecological relevance, judgements made during problem formulation should be carried through to risk characterization unchanged. For example, a practitioner may judge that a lab-based bioassay has only moderate ecological relevance – that judgement should be made during problem

[48] Based in part on consideration of available WOE frameworks including those developed by Menzie et al. (1996), Hull and Swanson (2006) and Exponent (2010).

[49] Technical Module 4 was under development when this guidance document was finalized.

[50] Several factors influencing uncertainty have been identified in other WOE frameworks as formalized attributes (e.g., Menzie et al. 1996).

formulation and will not change based on results of the bioassay. In the case of uncertainty, some of the criteria listed in **Section 2.3.4.3** for selection of LOEs (e.g., anticipated data quality) ultimately become contributors to uncertainty assessment during risk characterization.

5.5.2.1. Magnitude of Response

The magnitude of any observed responses is arguably the most important attribute of an LOE. Defining the meaning of "magnitude" is an important consideration, as the term can refer to a number of characteristics including:

- **Effect size** (change or difference in the response variable) relative to AELs or levels considered potentially ecologically relevant;
- **Spatial scale** of the change or difference;
- **Temporal scale** of the change or difference; and
- **Probability** of harm suggested by the analysis[51].

Given the importance of spatial scale for most ROCs, it is usually appropriate to separate spatial scale from effect size so that the two types of information are communicated clearly. Specification of the characteristics of magnitude is a mandatory component of risk characterization; without this articulation, narrative conclusions such as "high risk" have no clear meaning. For example, if soil at a particular site is highly toxic (e.g., mortality > 50%), risks would be considered more significant if the entire site was toxic versus only one small portion of the site.

Determinations regarding magnitude may be qualitative or quantitative. If categorical assignments are used, it is preferable to constrain the number of categories to five (e.g., negligible, low, moderate, high, very high) or less and to define the terms (and decision rules for break points among categories) clearly.

5.5.2.2. Causality

An assessment of causation in an ecological risk assessment attempts to identify the cause of observed effects, and attempts to distinguish between

> ***Definition:***
>
> Causation – the act or fact of causing; the production of an effect by a cause. Causation differs from association (correlation) in that the latter does not imply a mechanistic linkage between observations.

[51] Depending on the risk assessment type, probability of harm may not be a pertinent consideration. In a retrospective condition assessment, the site conditions are already manifested, whereas a predictive risk assessment involving a population model may invoke probabilities of population decline or extinction.

associations that are coincidental (or caused by external factors) and associations that are driven by specific contaminant influences.

Ideally, causality is evaluated systematically. For example, a toxicity identification evaluation (TIE) evaluates the relationship between a cause (adjustment to a sample treatment) and an effect (modification of toxicity response magnitude) by testing each potential causal agent one at a time. However, in the absence of this type of systematic approach, wholly empirical methods can be used to provide insight into causality (i.e., circumstantial evidence), provided that a defensible underlying explanation for the response can be postulated. Causality is explored in detail in Technical Module 4 and by Suter *et al.* (2010).

5.5.2.3. Ecological Relevance

> ***Key Concept:***
>
> Ecological relevance assesses the degree to which the LOE is aligned with, or predictive of, the assessment endpoint.

A key attribute of any LOE that is considered during problem formulation and risk characterization is the ecological relevance, which is the relevance to the assessment endpoint that it is intended to address. For example, direct measures of a community (e.g., invertebrate abundance and diversity) are generally considered to be more ecologically relevant than laboratory bioassays. Thus, a direct community measure carries greater strength for this attribute than a laboratory-based measure. However, laboratory-based measures may be more precise and better able to detect responses, so would score higher for other attributes. The ecological relevance of any LOE should be evaluated during problem formulation as one of the criteria for LOE selection (see **Section 2.3.4**).

5.5.2.4. Attribute Uncertainty

Uncertainty is an integral component of risk characterization. Although consideration of uncertainty is part of the WOE procedure, Step 5 of the risk characterization process is dedicated to this issue (see **Section 5.6**) to ensure that uncertainty is rigorously addressed.

Uncertainty is the culmination of many individual factors (see Section 2.3.4.3 and Menzie *et al.* 1996). Some important categories of uncertainties include:

- **Sensitivity and specificity** – Sensitivity refers to the ability of an LOE to reliably detect a change in an environmental response despite the presence of

> ***Key Concept:***
>
> Sensitivity and specificity relate to the extent to which the LOE is sensitive to the stressor and specific to site conditions.

natural or analytical variability and uncertainty. Specificity refers to the extent to which data, media, species, environmental conditions, and habitat types used in the study design reflect the site of interest (Exponent, 2010).

- **Data Quality** – The extent to which data quality objectives (DQOs) and other recognized characteristics of high quality studies are met. LOEs that apply precise and standard methods with accepted quality assurance and quality control (QA/QC) procedures are more valued, whereas LOEs that use novel methods or imprecise data with unacceptable QA/QC would have higher uncertainty in application (Exponent, 2010). In addition, studies designed with appropriate statistical power and robust study designs are more valued.
- **Representativeness** – The degree to which the spatial and temporal nature of the data collected is reflective of real potential exposure and effects. The representativeness attribute is strongest for studies that:
    - Conduct synoptic (simultaneous) sampling of measurement endpoints;
    - Repeat sampling over multiple seasons or environmental conditions; and
    - Describe natural spatial or temporal variation through replication and characterization of stochasticity (random error).

5.5.3. Scoring or Ranking of Attributes

Results for individual attributes described in the previous section must be evaluated. Commensurate with an intermediate level of quantification in risk characterization, each attribute can be summarized using scores or ranks such as negligible-low-moderate-high, or integer scores, or continuous numerical scores. The scoring and ranking system should be defined in advance during problem formulation to facilitate transparency in interpretation of results. Examples for typical types of LOEs are provided in **Table 5-1**, using the attributes for magnitude, causality, and ecological relevance, along with associated uncertainties. For cases where more resolution is needed, or less resolution will suffice, practitioners may provide rationale for alternative approaches. Importantly, the classification of attribute performance as “negligible” or “low” or “moderate” or “high” should be consistent with protection goals and acceptable effect levels articulated during problem formulation.

Once attributes are scored, they must be considered simultaneously to support overall evaluation of several LOE for an assessment endpoint. In other words, the relative importance of magnitude, causality and ecological relevance must be weighed. This can be done quantitatively by combining some or many attributes into a common metric (e.g., Exponent 2010) or qualitatively by leaving attributes in their own units (e.g., Hull and Swanson 2006). The default approach presented here is based on leaving the major

attributes in their own units, to increase transparency. Examples for the results of scoring LOEs are provided for terrestrial and aquatic cases in **Tables 5-2 and 5-3** respectively. These example summary tables include evaluation of uncertainty (Step 5 of risk characterization) and they also include an overall evaluation of risks for each assessment endpoint (the subject of the next section below). The results shown in **Tables 5-2 and 5-3** must be generated transparently, using criteria defined in advance during problem formulation (e.g., **Table 5-1**).

5.5.4. Integrated Evaluation by Assessment Endpoint

Once individual LOEs have been characterized, the findings must be evaluated separately for each assessment endpoint (e.g., risks to wildlife are not traded off against risks to invertebrates[52]. As shown in the example **Tables 5-2 and 5-3**, the final column of those tables contains a short narrative summary of the key rationale used to make judgements about risks for each assessment endpoint. That rationale is a succinct summary, which can be elaborated in the main text of an ERA, usually as part of the narrative summary of risk conclusions that is articulated in Step 7 (**Section 5.8**).

> ***Key Concept:***
>
> Coherence refers to the concurrence of findings from different LOEs, including where they agree strongly and whether they diverge.

The most important element of integrating findings across multiple LOEs is coherence. Coherence can be defined as the degree to which components are logical and internally consistent. This does not require that all components provide the same response type; rather it means that LOEs should ideally tell a story that is logical and orderly.

Coherence assessment provides an opportunity for the risk assessor to provide a unifying explanation for the responses observed, given the information on each LOE, the uncertainty in the LOEs, and the relevance of the LOEs to the assessment endpoint.

The coherence assessment is where the logic connecting the various LOE findings should be articulated. The risk assessor should articulate overall findings for the LOE with a narrative explaining how contradictory results are reconciled. Also, the risk assessor should consider and acknowledge information and associated LOEs that were not available and therefore could not be considered in the WOE procedure.

[52] Trade-offs among valued environmental attributes may be considered later as part of risk management. In that context other factors, including human health concerns, socio-economics, legal and financial concerns, and other factors may ultimately influence site management.

Table 5-1. Example criteria for scoring attributes for major types of lines of evidence

		Rating	Type of LOE Chemistry (water, soil, sediment, tissue)	Toxicity Tests	Quantitative measures of Plant or Invertebrate or Community Abundance, Biomass, Richness	Qualitative Measures of Presence / Absence or Relative Abundance	Comparison of Dose-Based Exposure to Toxicity Reference Values (if food chain models are used)
MAGNITUDE	Degree of contamination and effect size	Negligible	Below standards/criteria/guidelines	Relative Effect Size < 10%	Relative Effect Size < 10%	Subjective evaluation based on spatial patterns	Subjective evaluation based on combined consideration of (a) HQs on-site relative to reference, and (b) for common species, likely population-level implications.
		Low	Chemistry is simply characterized as "above benchmarks" (for water, soil, sediment) or "Elevated" relative to reference or local gradient; differentiation on the basis of degree of contamination is not used.	Relative Effect Size 10-20%	Relative Effect Size 10-20%		
		Moderate		Relative Effect Size 20-50%	Relative Effect Size 20-50%		
		High		Relative Effect Size > 50%	Relative Effect Size > 50%		
	Spatial Scale for Evaluation of Magnitude		Analysis for individual samples and groups of samples across portions of the site.	Analysis for individual samples and groups of samples across portions of the site.	Analysis for individual samples and groups of samples across portions of the site.	Analysis of spatial gradients over the areas where sampling occurs.	Analysis on an area basis (probably the entire site or sampling area).
	Uncertainty About Magnitude	Low Moderate High	Subjective evaluation based on number of samples, quality and number of reference samples, and any other relevant considerations.	Subjective evaluation based on statistical significance, number of samples, number of controls & reference samples, extrapolation assumptions, and any other relevant considerations.	Subjective evaluation based on statistical significance, number of samples, number of controls & reference samples, extrapolation assumptions, and any other relevant considerations.	Subjective evaluation based on level of rigor in the measures used.	Subjective evaluation depending on uncertainty in exposure data, the type of TRV (NOAEL, LOAEL, ECx, etc.), quality of dose-response data, etc.

Table 5-1 continues on next page.

		Rating	**Chemistry (water, soil, sediment, tissue)**	**Toxicity Tests**	**Quantitative measures of Plant or Invertebrate or Community Abundance, Biomass, Richness**	**Qualitative Measures of Presence / Absence or Relative Abundance**	**Comparison of Dose-Based Exposure to Toxicity Reference Values (if food chain models are used)**
CAUSALITY	Evidence for Causality	None Weak Strong	Qualitative or quantitative evaluation of potential link between contamination and a site-related source. For tissue chemistry, spatial concordance between tissue and other media is evaluated.	Subjective evaluation based on combined consideration of study design, sample size, statistical significance, explanatory power. Rationale provided in each case.	Subjective evaluation based on combined consideration of study design, sample size, statistical significance, explanatory power. Rationale provided in each case.	Subjective evaluation based on concordance of spatial patterns with spatial patterns in chemistry.	Subjective evaluation based on concordance with chemistry data.
	Uncertainty about Causality	Low Moderate High	Subjective evaluation based on combined consideration of study design, sample sizes, and understanding of site characterization. Rationale provided in each case.	Subjective evaluation based on statistical significance, number of samples, number of controls & reference samples, extrapolation assumptions, and any other relevant considerations.	Subjective evaluation based on statistical significance, number of samples, number of controls & reference samples, extrapolation assumptions, and any other relevant considerations.	Subjective evaluation based on level of rigor in the measures used.	Subjective evaluation based on degree of concordance with chemistry data, sample sizes, etc.
ECOLOGICAL RELEVANCE	Ecological Relevance	Low Moderate High	Chemistry compared to generic environmental quality criteria or guidelines where relevance to specific receptor group is weak. Tissue chemistry, when compared to tissue-based Critical Body Residues.	Endpoints other than mortality, growth, reproduction. Endpoints for mortality, growth, reproduction.	Direct measures of plant and invertebrate communities such as abundance, biomass and	Direct measures of presence/ absence and abundance typically have high	Subjective: Usually low for simple HQs based on NOAEL/LOAEL -based TRVs. Usually moderate based on ICx - based TRVs. Can be high if results quantitatively extrapolated to match the relevant level of ecological organization (e.g. population level) and if effects are predicted using dose-response relationships.

			richness typically have high ecological relevance.	ecological relevance.	

Table 5-2. Example summary table of WOE (terrestrial ecosystem) by assessment endpoint

	Assessment Endpoint	LOE Group	Magnitude	Spatial Scale	Uncertainty about magnitude	Evidence for causal relationship between exposure and effects[3]	Uncertainty about causality	Ecological relevance	Overall Assessment
Plants	Ecological function and food and cover and wildlife	Soil chemistry	Above Benchmarks	1000 m^2	Moderate	No evidence of links between benchmarks and site-specific effects to plants, because benchmarks for site-specific COPCs are based on invertebrate data only.	High	Low	Low effects, high uncertainty - Soil and chemistry benchmarks for site-specific COPCs are not based on plants but rather are based on invertebrates. The community survey indicates there are low effects, but the cause may be fungal infection rather than site-related COPCs, and uncertainty is high.
		Community survey	Low	n/a	High	No evidence of relationships between biomass / richness and soil chemistry. Leaf spots and shoot blights that are evident on a few species are believed to be related to fungal infection, not contaminants.	High	High	
Soil Invertebrates	Diverse and abundant invertebrate community, and ecological function as food for wildlife	Soil chemistry	Above Benchmarks	1000 m^2	Moderate	Weak evidence (from literature) of links between benchmarks and effects to soil invertebrates, but application to specific sites limited by variation in toxicity modifying factors	High	Low	Low effects, moderate to high uncertainty - Although tissue concentrations of COPCs in earthworms are elevated and there is some site-specific toxicity observed, the toxicity results are not correlated with COPCs. Furthermore, the most ecologically relevant line of evidence (invertebrate abundance and richness) indicates no effects.
		Earthworm tissue bioaccumulation	Moderate	300 m^2	High	Weak evidence (from literature) that observed contaminant concentrations could be causing toxicity	Moderate	Moderate	
		Earthworm (*Eisenia foetida*) survival in laboratory toxicity test	Low	30 m^2	Moderate	No evidence of a concentration-response relationship. One sample yielded significant toxicity, but not at high contaminant concentration.	Moderate	Moderate	

Table 5-2 continues on next page.

	Assessment Endpoint	LOE Group	Magnitude	Spatial Scale	Uncertainty about magnitude	Evidence for causal relationship between exposure and effects[3]	Uncertainty about causality	Ecological relevance	Overall Assessment
Soil Invertebrates	Diverse and abundant invertebrate community, and ecological function as food for wildlife	Earthworm (*Eisenia foetida*) survival in laboratory toxicity test	Low	30 m^2	Moderate	No evidence of a concentration-response relationship. One sample yielded significant toxicity, but not at high contaminant concentration.	Moderate	Moderate	Low effects, moderate to high uncertainty - Although tissue concentrations of COPCs in earthworms are elevated and there is some site-specific toxicity observed, the toxicity results are not correlated with COPCs. Furthermore, the most ecologically relevant line of evidence (invertebrate abundance and richness) indicates no effects.
		Abundance and richness in quadrat sampling	Negligible	n/a	High	n/a	n/a	High	
Birds	Healthy and reproducing local population	Community survey	Negligible	n/a	High	n/a	n/a	High	Low effects, moderate uncertainty - hazard quotients based on species (chickens, quail) with unknown relevance to wild birds, and highly conservative exposure assumptions. Predicted effects were low, and minor individual responses (if present) are unlikely to translate to population effects.
		Food chain model (dose to reproducing females in breeding season)	Low	1000 m^2	Moderate	Literature-based dose-response relationship well established but highly variable among species	Moderate	Moderate	
Mammals	Healthy and reproducing local population	Food chain model	Negligible	1000 m^2	Moderate	n/a	n/a	Moderate	Negligible effects, moderate uncertainty - main uncertainty is TRVs based on domestic and laboratory species. No field endpoints available for confirmation.

Table 5-3. Example summary table of WOE (aquatic ecosystem) by assessment endpoint

	Assessment Endpoint	LOE Group	Magnitude	Spatial Scale	Uncertainty about magnitude	Evidence for causal relationship between exposure and effects[3]	Uncertainty about causality	Ecological relevance	Overall Assessment
Macrophytes	Ecological function as food for fish and wildlife	Sediment & surface water chemistry	Above Benchmarks	100 m^2	Moderate	No evidence of links between benchmarks and site-specific effects to macrophytes	High	Low	Negligible effects, high uncertainty - Sediment and surface water chemistry benchmarks are not based on macrophytes. The community survey indicates there are no effects, but uncertainty is high.
		Community survey	Negligible	n/a	High	n/a	n/a	High	
Benthos	Aquatic invertebrate community structure, and ecological function as food for fish and wildlife	Sediment & surface water chemistry	Above Benchmarks	100 m^2	Moderate	Weak evidence of links between benchmarks and site-specific effects to benthos	High	Low	Moderate effects, moderate uncertainty - Three of the four effects-based measures show moderate effects, with varying evidence for causal relationships to site contaminants.
		Amphipod toxicity test: survival	Moderate	100 m^2	High	Strong evidence that survival driven by COPCs (based on TIE and regressions)	Low	Moderate	
		Amphipod toxicity test: growth	Low	30 m^2	High	Weak evidence that growth related to COPCs (based on regressions)	Low	Moderate	

Table 5-3 continues on next page.

	Assessment Endpoint	LOE Group	Magnitude	Spatial Scale	Uncertainty about magnitude	Evidence for causal relationship between exposure and effects[3]	Uncertainty about causality	Ecological relevance	Overall Assessment
Benthos	Aquatic invertebrate community structure, and ecological function as food for fish and wildlife	Abundance as total organisms	Moderate	30 m^2	Moderate	No evidence of relationship between abundance and contamination	Moderate	High	Moderate effects, moderate uncertainty - Three of the four effects-based measures show moderate effects, with varying evidence for causal relationships to site contaminants.
		Richness as total taxa	Moderate	30 m^2	Moderate	Weak evidence that richness related to COPCs (based on regressions)	Moderate	High	
Fish	Abundance and viability of local fish populations	Surface water quality	Above Benchmarks	n/a	Moderate	Weak evidence of links between benchmarks and actual effects to fish	High	Moderate	Negligible to low effects, moderate uncertainty - data do not indicate effects on fish directly, but there is high uncertainty. Some effects on food sources may occur, but the spatial scale is limited and population-level impacts are unlikely.
		Relative abundance	Negligible	n/a	High	n/a	n/a	High	
		Abundance and diversity of benthos as food	Moderate	30 m^2	Moderate	No evidence that abundance of benthos is affected, but weak evidence for richness	n/a	High	
Wildlife	Abundance and viability of local bird, mammal and amphibian populations	Food chain model	Negligible	n/a	Moderate	n/a	n/a	Moderate	Negligible effects, moderate uncertainty - HQs < 1 in all cases; some uncertainty due to uncertainty in TRVs.

[1] This table may be based on a more detailed summary table that provides raw results rather than summary results.
[2] Decision rules for defining scores (e.g., negligible - low - moderate - high) need to be defined in advance for each of the attributes, during problem formulation - see Table 5-1 as a default.
[3] n/a - no need to evaluate causality where no effect exists.

For many ERAs, the diversity in measurement tools may yield divergent results, and trade-offs among contradictory LOEs will need to be made to derive an overall evaluation of risks for each assessment endpoint. In making judgements in this regard, practitioners should consider that:

- LOEs that are highly ecologically relevant should be given more emphasis when making trade-offs among LOEs, provided that uncertainties are comparable.
- If there is negligible magnitude of response and low uncertainty, there is no need to consider causality. However, if magnitude is high and/or uncertainty is great, causality becomes more important.
- If there is no evidence for causality and low uncertainty in the causality assessment, then observed responses are not related to site contaminants.

An important consideration when evaluating multiple LOEs relates to redundancy in LOEs. If any of the four major categories of evidence (see **Section 2.3.4.2**) are missing, they are effectively assigned zero weight, whereas multiple measures within a major line of evidence can result in double-counting (or more) of redundant or strongly correlated information. For example, there may be two different measures of soil invertebrate diversity, but no information at all on soil toxicity). To address this problem, the overall WOE evaluation for an assessment endpoint should take into account redundancy, by acknowledging overlap in metrics and explaining in the narrative rationale how the redundant LOEs were considered. For highly complex sites with a large number of LOEs, more formal methods may be appropriate, such as combining redundant LOEs first, prior to integrating across all LOEs relevant to an assessment endpoint.

Professional judgement – Best professional judgement (BPJ) plays a significant role in the default WOE procedure during the integrated evaluation of each assessment endpoint. Particularly for cases where individual LOEs provided contradictory results, the narrative summary must provide rationale, using professional judgement, as to how the WOE conclusions were derived. This is the primary role of BPJ. In contrast, the use of BPJ is more limited in the analysis of individual LOEs, because LOEs are evaluated based on criteria that are defined in advance during problem formulation.

The role of professional judgement is not limited to the specific default WOE procedure recommended in this guidance. Even when more formal quantitative methods are used to combine results of multiple LOEs, professional judgement is used to define the how trade-offs are made among contradictory LOEs. Nevertheless, although BPJ is a necessary and important part of WOE (Chapman *et al.* 2002), there are pitfalls of reliance on BPJ, including:

- challenges with demonstrating the reasonableness of determinations;
- lack of consistency in risk conclusions reached by different practitioners when faced with similar input (i.e., repeatability issue);

- potential for abuse by practitioners seeking to find a pre-determined outcome;
- unintended bias resulting from perception of results according to an established paradigm, rather than objective evaluation of all possible explanations.

Despite these challenges, much of the problem can be resolved through proper articulation of "good practice" in application of BPJ. Wandall (2004) argues that proper application of professional judgement in risk assessment requires that (1) risk assessors are aware of what underlying values they are relying on, (2) the values are justifiable, and (3) transparency is ensured. This requirement for transparency is the foundation of properly applied professional judgement, and translates into the following guiding principles[53]:

- All assumptions and decisions must be supported with a rationale, especially for those instances where education and training (i.e., no citations are available) were used as the basis for the professional judgement.
- Declarative and unqualified conclusions such as "the risk assessment proved that there are no adverse effects" should be avoided. Instead conclusions should reflect where professional judgement was applied in the evaluation (e.g., "The risk assessment, based on our professional judgement of ABC data, and subject to assumptions XYZ, found no evidence of adverse effects").

5.6. Step 5 – Evaluate ERA Uncertainties

There are numerous sources of uncertainty and variability in ERA. These uncertainties fall in multiple categories (see text box at end of **Section 2**). Uncertainties must be evaluated in order to determine the level of confidence associated with risk estimates and to determine to what extent additional work is warranted to reduce uncertainties.

Importantly, the level of detail and rigour needed to address uncertainty will vary depending on the complexity of the ERA and the results. If estimated risks are either extremely low or extremely high, it may be easy to demonstrate that uncertainty is unlikely to change that conclusion. On the other hand, more rigorous evaluation of uncertainty is usually warranted when estimated risks are in the range that may or may not be acceptable.

Many aspects of uncertainty can be integrated directly into WOE summary tables as shown in **Tables 5-2 and 5-3**, and discussed in **Section 5.5.2**. However, uncertainty evaluation extends beyond the assessment of uncertainties for individual attributes and

[53] Further discussion of BPJ in ERA and WOE evaluation can be found in Bay et al. (2007); WDNR (2009) and Lee and Jones-Lee (2002).

endpoints. Therefore, this section is identified as a separate step from WOE (even though uncertainties are evaluated during the WOE procedure).

Addressing uncertainties requires that the practitioner:

- Identify uncertainties in the risk assessment, and distinguish them from elements of the risk assessment where there is reasonable certainty.
- Evaluate the implications of uncertainties – for instance, could risk conclusions change if uncertainties were reduced, and how likely is it that risk management decisions may change?
- If warranted, explicitly integrate uncertainties into risk characterization methods (e.g., using probabilistic methods).
- If warranted, determine the potential value of reducing uncertainty through follow up investigations – for instance, to what extent would additional work increase accuracy and precision of risk estimates and lead to a more informed risk management decision?

5.6.1. Identifying Uncertainties

The first step in addressing uncertainties in ERA is to differentiate factors and conclusions that are known with reasonable certainty from those that are uncertain. Specific uncertainties may apply to any data, parameters, models or assumptions used in the risk assessment. The various sources of data and information related to characterizing exposure and effects (see **Sections 3 and 4**) may all be subject to uncertainty to varying degrees. For uncertainties that can be quantified using data, basic plots (e.g., boxplots) and descriptive statistics can be used to characterize the uncertainty in the data (e.g., minimum, maximum, median, mean, variance).

5.6.2. Evaluating the Implications of Uncertainties

Uncertainties are important because of their potential implications on risk estimates and ultimately on risk management decisions. The implications of specific uncertainties are most easily evaluated using sensitivity analysis to test how risk estimates change according to various "what-if" scenarios for each quantity. Sensitivity may be tested using the minimum and maximum possible values for a given quantity, or any other metrics (e.g., the 5^{th} and 95^{th} percentiles). For example, a hazard quotient could be estimated using the minimum and maximum measured COC concentration in food items, as a bounding analysis. If the hazard quotient does not differ appreciably between the two scenarios (e.g., if it was well below 1 in both cases), the risk assessor may conclude that uncertainty related to the tissue concentration is negligible. In contrast, if the hazard

quotient changes from less than 1 to greater than 1, the uncertainty in the tissue concentration may need to be explored further.

Each of the uncertain quantities used to estimate risks can be varied, independently or at the same time, to generate a range of "what-if" scenarios. Sensitivity analyses are useful for understanding which uncertainties have the most potential influence on risk estimates. The cumulative effect of multiple uncertainties can be understood to some extent using these methods. However, simultaneous consideration of the cumulative effect of multiple uncertainties is better addressed using probabilistic methods, as outlined below.

5.6.3. Integrating Uncertainties into Risk Characterization

The cumulative influence of uncertainties is best understood using probabilistic methods. As discussed earlier in **Section 5.3.6**, probabilistic methods are useful for characterizing risks because they provide accuracy and realism that is not captured when data and parameters are represented with point estimates. In the context of evaluating uncertainty, the key benefit of a probabilistic assessment is facilitating understanding of the cumulative effects of multiple uncertainties on risk estimates.

> ***Key Concept:***
>
> Probabilistic methods improve accuracy in risk characterization by capturing a more realistic range of possible outcomes than deterministic methods, and facilitate understanding of the cumulative effects of multiple uncertainties on risk estimates.

What is a probabilistic assessment? – Probabilistic methods are distinguished from deterministic methods in that exposure is characterized not as a point estimate but as a probability distribution (or frequency distribution) of possible estimates, based on the use of distributions to characterize some or all of the uncertain input quantities. For example, all of the equations in a food chain model could be based on distributions rather than point estimates for each input parameter.

When should probabilistic methods be used? – Risk assessors should consider developing probabilistic models whenever more accurate estimates of risk could be important from a risk management perspective, or to simply evaluate the cumulative effects of multiple uncertainties. Consistent with the iterative approach to ERA (**Section 1.6**), if a deterministic (point estimate) risk assessment based on conservative assumptions shows that risks are acceptable, then the increased accuracy provided by a probabilistic model is not warranted. However, if risks are identified using deterministic methods, probabilistic methods should be considered.

How to implement probabilistic methods – Uncertainties are usually modeled using numerical simulation techniques such as Monte Carlo simulation[54]. A simulation model may be run a few thousand times; each realization or trial involves random selection of a value for each uncertain quantity (according to a probability or frequency distribution). In the case of a wildlife food chain model the output from each simulation trial might be, for example, a hazard quotient or an estimate of expected mortality.

Although the implementation of simulation techniques has been greatly simplified by the availability of commercial software packages, some of the design elements require careful consideration. First, the risk assessor must decide whether the simulation model will deal with variability among individuals in a population, or only with incertitude[55] (e.g., uncertainty regarding an average individual), or with both (Hoffman and Hammonds 1994). Simultaneous consideration of inter-individual variability and incertitude may warrant a two-dimensional simulation. Conversely, a simpler model may suffice in many cases for ERA, provided that results are interpreted correctly. Second, any correlations among uncertain variables should be accounted for in simulation models, otherwise the estimated probability distribution of risks will be too wide and may be skewed. In reality, many ecological parameters are highly inter-correlated (e.g., feeding rate and growth rate of a species, or feeding rates of several species that are all a function of temperature). There are ways to account for these correlations in simulations (Haas, 1999), but this requires additional information about the form of the correlation. Even where inter-correlation structures are available, there is still uncertainty in the structure of the model itself, and it is difficult to determine the quantitative effect of the inability of our models to exactly represent natural processes.

Risks represented as a probability or frequency distribution are informative, but the risk assessor must find ways to communicate the information in a way that is easily interpreted by risk managers and stakeholders. For example, it may be useful at the risk characterization stage to report particular statistics such as the probability that an average individual would exceed a particular effects threshold. Further guidance on probabilistic exposure methods is provided by Cullen and Frey (1999), Suter *et al.* (2000), and USEPA (1997a, 1997b, 2001).

Data requirements for probabilistic models – Although any model is best when data are plentiful, risk assessors should not shy away from probabilistic analyses in cases where

[54] Analytical methods and Taylor series approximation methods of propagating uncertainties are reviewed in Cullen and Frey (1999).

[55] Incertitude is uncertainty caused by incomplete descriptions of a mechanism or process and other limitations of scientific knowledge – the term is used here to distinguish this aspect of uncertainty from natural variation and other types of uncertainty. In statistical terms, for a parameter such as body weight, variability among individuals might be characterized with a standard deviation, whereas incertitude about the mean body weight might be characterized with a standard error.

data are sparse. In general, if a probabilistic analysis is appropriate for an ERA, the advantages of implementation will outweigh the disadvantages created by data limitations, provided that limitations are explicitly described. Methods exist for using limited data to construct probability distributions (Morgan and Henrion 1990; Cullen and Frey 1999), and simple distributions (e.g., uniform, discrete, triangular) can be used in data-poor cases. In addition, sensitivity analyses usually reveal that many uncertain quantities have little impact on the cumulative uncertainty, such that precise characterization of uncertainties is not always critical for all quantities.

5.6.4. Determining the Value of Reducing Uncertainty – When to Refine Risk Estimates

Risk estimates should be refined if the benefit (more informed decision) outweighs the costs of reducing uncertainty. If preliminary risk estimates indicate the potential for adverse effects, the underlying conservative assumptions and uncertainties should be critically evaluated using approaches outlined above (e.g., sensitivity analysis). The practitioner (and client) must make a decision to: (a) further refine the exposure or effects assessments to reflect site-specific conditions, or (b) conclude that risk is unacceptable or unresolvable and that remediation or other risk management options should be considered. A matrix based on varying levels of estimated risk and uncertainty (based on Pearsons and Hopley 1999) can be a useful way to conceptualize interpretation of uncertainties:

	Low Magnitude of Risk	*High Magnitude of Risk*
Low Uncertainty in Risk Estimate	Low Precaution	Moderate Precaution
High Uncertainty in Risk Estimate	Moderate Precaution	High Precaution

Refinement of risk estimates for the "high" category of precaution is recommended; the "medium" category of precaution may also indicate a need to reduce uncertainty as necessary to support management actions. This refinement may involve one or more of the following strategies:

- Reduce parameter uncertainty by gathering additional data. Supplemental data collection should be targeted to deal with the underlying cause of the parameter

uncertainty (e.g., address spatial coverage, improve analytical detection limits, collect bioavailability information, evaluate cause and effect mechanisms).

- Reduce structural (model) uncertainty by adopting a more appropriate model and any additional data needed to support that model[56]. Risk assessment should be an iterative process where new data may require reassessment of previous approaches or conclusions. This iterative process allows risk assessment to be a dynamic process well suited to ecological study, and does not indicate a failure of the initial screening risk estimate.
- Provide risk managers with multiple risk scenarios for consideration as a series of risk estimates with different assumptions and descriptions of uncertainty.

Several other strategies are often employed; however, they do not directly reduce parameter or model uncertainty. For example:

- Professional judgement is often used to fill in gaps in model structure. This may reduce uncertainty, but it may not, and there is no objective way to know. Conservative assumptions are often used as part this strategy; although it does not reduce uncertainty, it ensures that the majority of the uncertainty errs on the side of caution. The challenge in using conservative assumptions lies in balancing conservatism and ecological realism relative to site management needs.
- Increase the number and types of lines of evidence considered in a weight of evidence approach. This strategy does not reduce the uncertainty in any single line of evidence, but does reduce overall uncertainty in the conclusions of the risk assessment because the limitations of one line of evidence are frequently balanced by the strengths of another.

5.7. Step 6 – Extrapolation / Interpolation

This aspect of uncertainty in risk assessments warrants its own step in the risk characterization process because it addresses the issue of transferability of the ERA findings over time, space, and/or alternate site use scenarios. This is particularly relevant to site managers because management decisions may require confidence that risk narratives remain applicable even if some of the underlying assumptions change. For example, where a site is divested or otherwise changes in ownership or land use, it is undesirable to repeat the entire ERA process.

By design, risk assessments focus resources on a narrow subset of the potential receptors, spatial locations, and measurement endpoints. In concentrating on a narrow and focused

[56] Additional model complexity may not reduce uncertainty, and often increases uncertainty. The benefits of additional model complexity should be evaluated on a case by case basis.

set of risk hypotheses, there is danger in not "seeing the forest for the trees." Therefore, near the end of the risk assessment, it is prudent to conduct a reality check to assess how representative the risk assessment is expected to be in terms of the broad site management goals.

Conceptually, the extrapolation/interpolation assessment entails a broadening of the scope of the risk conclusions from the detailed findings (e.g., specific risk estimates for representative organisms and exposure scenarios) to the broadly defined assessment endpoints. Due to practical constraints, ERAs have limitations in the spatial and temporal domains considered, and in the degree to which combinations of chemical, physical, and biological components are explicitly evaluated. The extrapolation/interpolation assessment serves as a reality check for the relevance of the study results to the valued ecosystem components, and provides context for the overall findings.

Some specific issues to be addressed at this stage include:

- *Can results for one receptor be extended to other species at the site?* For example, if a mallard duck was selected as a receptor of concern, can we assert that risks to other dabbling ducks, other waterfowl, or other omnivorous birds in general are expected to be lower than those for the mallard? In some cases the ROC is selected based on its presumed sensitivity to relevant COCs and pathways of exposure. However, in other cases, other considerations may dictate ROC selection (data availability, standardized methods for assessment); in these cases, the risk assessor should qualitatively evaluate the degree of protectiveness afforded other species not rigorously evaluated in the risk assessment.
- *Are thresholds for individual COCs protective of the entire contaminant mixture?* Where a site-specific standard has been developed for an individual substance of concern, and that substance serves as a surrogate for other COCs, there is an implicit assumption that the other COCs will not increase relative to the individual (indicator or surrogate) substance.
- *Are the study conclusions dependent on an assumption of fixed site use, or would the results also apply to site redevelopment or restoration?*
- *Can conclusions or quantitative relationships based on limited sample sizes be extended to other spatial units, other habitats, other depths, or physical conditions?* The underlying assumption is that exposure-effect relationships observed at sampled areas will remain applicable when extended to other unsampled portions of the site. However, if the unsampled areas are substantially different in terms of factors that may influence COC bioavailability, or represent habitat conditions not evaluated in the risk assessment, there is uncertainty in extrapolating study findings. The specific issue of deriving site-specific standards or benchmarks, which implicitly assumes transferability of quantitative relationships, is considered further in Step 7.

The risk assessor should specify constraints or caveats to the extension of study findings across space, time, habitat type, or biological assemblage. Note that the requirement for extrapolating to new conditions (or predicting future responses) is closely linked to the assessment objectives identified during the problem formulation.

5.8. Step 7 (Optional) – Site-Specific Remediation Standards

Where significant ecological effects are observed over some or all of a contaminated site, it may be appropriate to develop site-specific remediation standards. These values are often also referred to as site-specific target levels (SSTLs) and represent concentrations in environmental media that, once achieved, will meet the environmental protection goals for the site. This step is listed as optional because formal development of site-specific remediation standards may not be required, depending on the type of assessment and the risk management needs. For example, if risk characterization is conducted using a parcel-based or spatially explicit evaluation of risks (i.e., grid cells evaluated individually for acceptability of risks), then development of numerical target levels for specific substances would not be required.

CCME (1996) provides a framework for the development of site-specific environmental remediation objectives. Under the framework, where a risk-based approach is applied, risk assessment procedures can be used to establish remediation objectives on a site-specific basis, as discussed in the following subsections.

5.8.1. Considerations for Developing Site Specific Remediation Standards

5.8.1.1. Appropriate Site Media

Most risk assessments evaluate more than one site medium (e.g., soil, sediment, tissue, surface water, groundwater, porewater). In many cases, one exposure medium can be identified as the "driver" (i.e., dominate the magnitude of risk estimates) by strongly influencing the environmental exposures. The practitioner should ensure that the choice of a medium for development of a site-specific standard sufficiently addressed the risk pathways of relevance and does not leave other important pathways unaddressed. For example, if a wildlife risk assessment determined that metals uptake through soil-based pathways and drinking water were both important, it could be necessary to: (1) develop standards for both pathways; or (2) develop standards for one pathway with explicit acknowledgement that the other pathway remains.

Another consideration is the degree to which the site media can effectively be used as a means of developing a remediation and/or monitoring plan. Soil and sediment are commonly applied media because they represent sinks for contaminants, are relatively immobile, and can easily be sampled. In contrast, tissues or organisms are not commonly used because the organisms may be mobile, availability of tissues may be seasonal, and monitoring of post-remediation results may not be practical.

5.8.1.2. Appropriate Contaminants

The contaminant(s) that have been identified as the dominant sources of risk must be identified. This may be a simple decision, or quite complex, depending on the nature of the contaminant mixtures and relative risks estimated for each COPC. Some important considerations are:

- **Cumulative risks** – for related substances, a surrogate compound or integrated value may be useful. For example, total PAH may be used as an exposure measure if the composition of component PAHs across the site is stable.
- **Practical considerations** – the parameter adopted should be relatively easy to measure. For example, a total polychlorinated biphenyl (PCB) measurement may be preferred to a TEQ value derived from a weighted sum of polychlorinated biphenyl congeners, due to difference in analytical costs.
- **Degree of causation –** The contaminant should be identified to have a strong correspondence to environmental response, and ideally have strong evidence of causation. Where multiple COPCs are present and causation has not been determined, development of a site-specific standard requires an assumption that the indicator COPC is an effective surrogate for the effects of the entire mixture.

5.8.1.3. Contamination Pathways

In developing a site-specific standard, it is important to consider the pathways by which risk occurs, and the assumptions required for a standards-based remediation to be effective. For example:

- Will the site be re-contaminated by influence of either on-site or off-site (background) contributions?
- Are residual concentrations likely to attenuate over time, or increase through chemical reaction?

5.8.1.4. Spatial Scale

The application of a site-specific standard requires consideration of the spatial domain relevant to the receptors. For mobile receptors, weighted averaging of exposures can be incorporated in the development of standards. For sessile receptors, the spatial scale at which monitoring of risks will be conducted needs to be addressed (e.g., depth of soil/sediment, resolution of lateral COPC characterization).

The scale of relevance will strongly influence the methods used to apply the standard. For a sessile receptor, the standard may be a "not to exceed" threshold, whereas an averaging procedure could be applied to migratory organisms. For wildlife, an area-based average is often applied, depending on the home range of the receptor relative to the size of the site.

5.8.1.5. Modifying Factors

Where site conditions are variable, it may be appropriate to adjust site-specific standards on a location by location basis. This could account for bioavailability or toxicity differences that could be relevant across small spatial scales. For example, values of soil or sediment organic carbon content may be variable, and adjustment to account for bioavailability differences may be appropriate if the site-specific standard was developed on a dry-weight basis. Alternatively, if the risk assessment data were amenable, the standard could be developed on an OC-normalized basis. Other modifying factors include pH and salinity in aqueous samples.

5.8.1.6. Approval and Application

Regulatory review will be required for any site-specific standard, and such may entail consideration of:

- Management checks for consistency with law or policy considerations;
- Socioeconomic factors; and
- Technical constraints.

Furthermore, removal or remediation actions defined using site-specific standards typically require a clear linkage to a risk management plan, including long-term monitoring. For this reason, development of site-specific remedial standards is often performed in parallel with the risk management process, as described in Step 9 below.

5.8.2. Methods for the Development of Site-Specific Remediation Standards

Site-specific standard development relies on underlying concentration-response relationships. In some situations, it is possible to directly adopt a toxicity reference value developed in the effects assessment stage. However, it is common that additional data synthesis or modeling is required to develop a site-specific standard, particularly once the considerations discussed in Section 5.8.1 are taken into account. For example:

- conversion of tissue-based TRVs to soil or sediment media may require bioaccumulation models or equations for back-calculation purposes; or
- concentration-response relationships from multiple lines of evidence may need to be synthesized (simplified) to yield a single threshold for management purposes.

Site-specific standard development may be complicated by the numerous COPCs and the range of responses observed, but typically entails the following steps:

1. Identification of a level of harm considered acceptable based on the risk characterization findings. This could be quantitative (e.g., soil concentration associated with a hazard quotient of 1.0 for a wildlife species) or could be qualitative (e.g., low risk as determined from a sediment quality WOE assessment);
2. Plotting the degree of harm (response) versus COPC concentration, either graphically or using a mathematical relationship (such as regression analysis); and
3. Resolving the uncertainty associated with an impact relationship between response and the exposure measure. For example, it must be determined whether it is acceptable to have a "smoothed" target concentration considered protective of a receptor even if an individual station exhibited a significant ecological response.
4. Conversion of the target concentration to the desired units, scale, and media of interest (as outlined in Section 5.8.1). The target concentration must be clearly defined in terms of spatial application (e.g., spatially weighted threshold, or a maximum not to be exceeded at any location), the parameter details (e.g., dry weight sediment versus organic carbon normalized, fillet tissue versus whole body) and the conditions/assumptions required for applicability of the target concentration.

5.9. Step 8 – Summarize Risk Conclusions

Following the technical application of a risk assessment, it is important to summarize results in a manner that is clear, accurate, concise, and meaningful to the risk manager. A risk narrative is often provided for this purpose. This risk narrative may be combined

with the summary of estimated risks for each assessment endpoint that concluded at the end of the WOE procedure (Step 4, **Section 5.5**), or it may be presented separately. Risk assessors must provide an opinion regarding their results generated with respect to confidence, uncertainty and significance of impacts. As described by Exponent (2010):

A full narrative is analogous to writing the results and discussion sections of scientific papers and is intended to help other reviewers or risk managers understand how the risk assessor reached their conclusions based on the evidence in hand. The narrative can be used to help reach agreements, identify disagreements, and identify aspects of the risk assessment that require additional clarity.

Although the WOE procedure has already articulated findings for each assessment endpoint based on consideration of magnitude of effects (including spatial and temporal scales), evidence for causality, ecological relevance and uncertainty for individual LOEs, the risk narrative should integrate that information into a form that is useful for decision-makers. Specific goals of the risk narrative may be to:

- Present in lay language the key rationales used to draw overall conclusions for each assessment endpoint during the WOE procedure;
- Summarize overall confidence in the specific findings, in light of the ecological relevance of the various LOEs and the strength of evidence implicating site-related contaminants as the cause of any observed effects;
- Summarize confidence that overall the risk assessment methods are relevant and that the findings can be extrapolated to the general conditions at the site (both now and under foreseeable future conditions);
- Summarize the extent to which key uncertainties may affect risk conclusions, and whether further work to refine those uncertainties may be warranted;
- Clarify the spatial and temporal scales at which effects are observed, or provide separate summaries of risk conclusions for different spatial or temporal units;
- Summarize the potential for cumulative impacts of site-related contaminants and other stressors.

5.10. Step 9 – Conduct Follow-up Actions

The final step in risk characterization is to link the study findings to the risk management process. Risk communication is an important aspect of the overall risk management process, and therefore it is helpful to frame the path forward at the conclusion of the risk assessment process. This may entail a summary of recommendations and a clear articulation of next steps for site closure, approvals, regulatory liaison, etc. Details may

not be included in the ERA if risk management considerations are addressed as a separate deliverable.

Depending on the outcome, and provided that the scope of the ERA includes provision of recommendations for next steps, recommendations for site management may include:

a) No further action required – the rationale for the decision should be succinctly summarized;

b) Additional investigation or risk assessment – if the residual uncertainty in the risk assessment is large, a decision could be made to refine the assumptions and reduce uncertainties. Where iteration is contemplated, the advantages and limitations of follow up studies should be assessed;

c) Implement risk management strategies – no physical actions are deemed necessary, but management activities may still be required (administrative controls, monitoring program); or

d) Remediation (often to site-specific standards that were developed during the risk assessment) – considerations for conceptual remedial design may be articulated.

The evaluation of potential follow-up actions should reconsider the overall assessment goals in light of the conclusions of the ERA. In some cases, as part of an adaptive management approach, the focus for management may shift to one of the other quadrants of the overall assessment framework (Section 2.2.1.1). If monitoring has been implemented, the results of monitoring must be assessed to determine what to do next. If the results of a past management action have failed to result in expected environmental improvements, then assessment of causation may become more important. Alternatively, if environmental improvements have been substantial, the requirements for long-term monitoring may be reevaluated. Step 9 provides an opportunity for risk managers to conduct a check of the site management recommendations against the broad site management goals, adjust the course of the investigation as appropriate, and update the conceptual model of the site to reflect recent information.

6. REFERENCES

Note to Reader:

For an overview of additional, general references regarding Ecological Risk Assessment information refer to Section 1.7.3.

Adams, W.J., R. Blust, U. Borgmann, K.V. Brix, D.K. DeForest, A.S. Green, J.S. Meyer, J.C. McGeer, P.R. Paquin, P.S. Rainbow, and C.M. Wood. 2011. Utility of tissue residues for predicting effects of metals on aquatic organisms. *Int. Environ. Assess. Manage.* 7:75-98.

Allard, P., A. Fairbrother, B.K. Hope, R.N. Hull, M.S. Johnson, L. Kapustka, G. Mann, B. McDonald, and B.E. Sample. 2010. Recommendations for the development and application of wildlife toxicity reference values. *Int. Environ. Assess. Manage.* 6:28–37.

Allen, H.E. 2002. *Bioavailability of Metals in Terrestrial Ecosystems: Importance of Partitioning for Bioavailability to Invertebrates, Microbes, And Plants*. Metals in the Environment Series. SETAC Publication.

American Society for Testing and Materials (ASTM). 2008. *Standard Guide for Developing Conceptual Site Models for Contaminated Sites*. Designation: E 1689-95 (Reapproved 2008). PA, United States.

American Society for Testing and Materials (ASTM). 2011. *Standard Terminology Relating to Biological Effects and Environmental Fate*. Designation: E 943-08. PA, United States.Barnthouse, L.W. 2008. The strengths of the ecological risk assessment process: Linking science to decision making. *Int. Environ. Assess. Manage.* 4:299-305.

Archbold, J.A., R.N. Hull, and M. Diamond. 2007. Potential importance of inhalation exposures for wildlife using screening level ecological risk assessment. *Human and Ecological Risk Assessment* 13: 870-883.

Bay S., W. Berry, P. Chapman, R. Fairey, T. Gries, E. Long, D. MacDonald and S.B. Weisberg. Evaluating consistency of best professional judgment in the application of a multiple lines of evidence sediment quality triad. *Int. Environ. Assess. Manage.* 3(4):491–497.

BC Ministry of Environment (BC) 1997. *Checklist for the Tier-1 Ecological Risk Assessment of Contaminated Sites in British Columbia*. Prepared for British Columbia Ministry of Environment, Victoria. Institute of Environmental Toxicology and Chemistry, Western Washington University, Bellingham, WA;

Ecological Planning & Toxicology Inc., Corvalis, OR; EVS Environment Consultants, North Vancouver, B.C. March 31, 1997.

Bier, V. 1999. Challenges to the acceptance of probabilistic risk analysis. *Risk Anal.* 19(4):703-10

Beyer, W.N., G.H. Heinz, and A.W. Redmon-Norwood [eds]. 1996. *Environmental Contaminants in Wildlife: Interpreting Tissue Concentrations*. CRC Press, Boca Raton, Florida.

Brady, N.C., and R.R. Weil. 2008. *The nature and properties of soils*. 14th ed. Prentice Hall, Upper Saddle River.

Beyer, W.N., and J.P. Meador. 2011. *Environmental Contaminants in Wildlife: Interpreting Tissue Concentrations*, 2nd Edition. CRC Press, Boca Raton, Florida.

Brasso, R.L. and D.A. Cristol. 2008. Effects of mercury exposure on the reproductive success of tree swallows (*Tachycineta bicolor*). *Ecotoxicology* 17:133-141.

Burnham, K.P. and D.R. Anderson. 2002. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. 2nd Edition. Springer, New York.

California State Water Resources Board. 1997. *Chemistry, Toxicity and Benthic Community Conditions in Sediments of Selected Southern California Bays and Estuaries.* California State Water Resources Board. May 1997.

CCME (Canadian Council of Ministers of the Environment). 1995. *A Protocol for the Derivation of Sediment Quality Guidelines for the Protection of Aquatic Life*. CCME. Winnipeg, MB.

CCME. 1996. *A Framework for Ecological Risk Assessment: General Guidance*. CCME. Winnipeg, MB.

CCME. 1996b. *Guidance Manual for Developing Site-Specific Soil Quality Remediation Standards for Contaminated Sites in Canada*. National Contaminated Sites Remediation Program. March 1996.

CCME. 1997a. *A Framework for Ecological Risk Assessment: Technical Appendices*. CCME. Winnipeg, MB.

CCME. 1997b. *Guidance Document on the Management of Contaminated Sites in Canada*. CCME. Winnipeg, MB.

CCME. 1999a. *Protocol for the Derivation of Canadian Sediment Quality Guidelines for the Protection of Aquatic Life*. CCME. Winnipeg, MB.

CCME. 1999b. *Protocol for the Derivation of Canadian Tissue Residue Guidelines for the Protection of Wildlife that Consume Aquatic Biota*. CCME. Winnipeg, MB.

CCME. 2001. *Canadian Tissue Residue Guidelines for the Protection of Wildlife Consumers of Aquatic Biota: Polychlorinated Biphenyls (PCBs)*. Updated. In: Canadian Environmental Quality Guidelines. CCME. Winnipeg MB.

CCME. 2002. *Canadian Soil Quality Guidelines for the Protection of Environmental and Human Health: Dioxins and Furans*. In: Canadian Environmental Quality Guidelines. CCME. Winnipeg, MB.

CCME. 2006. *A Protocol for the Derivation of Environmental and Human Health Soil Quality Guidelines*. CCME. Winnipeg, MB.

CCME. 2007. *A Protocol for the Derivation of Water Quality Guidelines for the Protection of Aquatic Life 2007*. In: Canadian environmental quality guidelines, 1999, Canadian Council of Ministers of the Environment, 1999, Winnipeg MB.

CCME. 2008a. *National Classification System For Contaminated Sites: Guidance Document*. CCME. Winnipeg, MB.

CCME. 2008b. Canada-wide standard for petroleum hydrocarbons (PHC) in soil – Technical Supplement. CCME. Winnipeg, MB.

CCME. 2012 (in prep). Guidance Manual for Contaminated Site Characterization in Support of Environmental and Human Health Risk Assessment. Volume 1 Guidance Manual, Volume 2 Checklists, Volume 3 Suggested Operating Procedures. CCME, Winnipeg.

Carlsen, T.M., P.F. Chapman, S. Brassfield, N. Elmegaard, A. Hoffman, W. Landis, S. J. Moe, D. Nacci, H.M. Noel, and J. Spromburg. 2008. Empiric Approaches to Population-Level Ecologic Risk Assessment. In: L.W. Barnthouse, W.R. Munns, Jr., and M.T. Sorensen (eds), *Population-level Ecological Risk Assessment*. SETAC Press, Boca Raton, USA.

CEAEQ (Centre d'expertise en analyse environnementale du Québec). 1998. *Procédure d'Évaluation du Risque Écotoxicologique pour la Réhabilitation des Terrains Contaminés*.. Ministère de l'Environnement et de la Faune, gouvernement du Québec, 139 p.

Chapman, P.M., A. Fairbrother, and D. Brown. 1998. A critical evaluation of safety (uncertainty) factors for ecological risk assessment. *Environ. Toxicol. Chem.* 17(1):99–108.

Chapman, P.M., B.G. McDonald, and G.S. Lawrence. 2002. Weight-of-evidence issues and frameworks for sediment quality (and other) assessments. *Human Ecol. Risk Assess* 8:1489-1515.

Checkai, R.T., M.S. Johnson, and M.S. Hawkins. *Selection of Assessment and Measurement Endpoints for Ecological Risk Assessment*. February 2002. U.S.

Army Environmental Center, Aberdeen Proving Ground, Maryland. SFIM-AEC-ER-TR-2002018.

Clifford, M., and J.C. McGeer. 2010. Development of a biotic ligand model to predict the acute toxicity of cadmium to Daphnia pulex. *Aquatic Toxicology* 98:1-7.

Cormier, S.M. and G.W. Suter II. A Framework for Fully Integrating Environmental Assessment. 2008. *Environ. Manage.* 42:543–556.

Cowan, C.E., D. Mackay, T.C.J. Feijtel, D. Van de Meent, A. Di Guardo, J. Davies, and N. Mackay. 1995. The Multi-Media Fate Model: A Vital Tool for Predicting fhe Fate of Chemicals. SETAC Press, Pensacola, Florida.

CSMWG (Contaminated Sites Management Working Group). 1999. *A Federal Approach to Contaminated Sites*. Federal Committee on Environmental Management Systems, Government of Canada. www.ec.gc.ca/etad/csmwg/index_e_html.

Cullen, A.C., and H.C. Frey. 1999. *Probabilistic Techniques in Exposure Assessment: A Handbook for Dealing with Variability and Uncertainty in Models and Inputs*. Plenum Press, New York.

De Schamphelaere, K.A.C., S. Lofts, and C. R. Janssen. 2005. Bioavailability models for predicting acute and chronic toxicity of zinc to algae, daphnids, and fish in natural surface waters. *Environ. Toxicol. Chem.* 24:1190–1197.

Di Toro, D., H.E. Allen, H.L. Bergman, J.S. Meyer, P.R. Paquin and R.C. Santore. 2001. Biotic ligand model of acute toxicity of metals. 1. Technical basis. *Environ. Toxicol. Chem.* 20:2383-2396.

Di Toro, D. M. and J.A. McGrath. 2000b. Technical basis for narcotic chemicals and Polycyclic Aromatic Hydrocarbon criteria. II. Mixtures and sediments. *Environ. Tox. Chem.* 19:1971-1982.

Di Toro, D. M., McGrath, J. A. and Hansen, D. J. 2000a. Technical basis for narcotic chemicals and Polycyclic Aromatic Hydrocarbon criteria. I. Water and tissue. *Environ. Tox. Chem.* 19:1951-1970.

Efron, B., and B. Gong. 1983. A leisurely look at the bootstrap, the jackknife and cross-validation. *Amer. Stat.* 37:36-48.

Environment Canada. 2010. Guidance Document on Federal Interim Groundwater Quality Guidelines for Federal Contaminated Sites. Draft, May 2010.

Environment Canada. 2011. *Metal Mining Guidance Document for Aquatic Environmental Effects Monitoring*. National EEM Office. Ottawa, ON.

Environment Canada and Ontario Ministry of the Environment 2008. *Canada-Ontario decision-making framework: for assessment of great lakes contaminated sediment*. ISBN 978-0-662-46148-7.

Escher, B. I. and Hermens, J. L. M., 2002. Modes of action in ecotoxicology: their role in body burdens, species sensitivity, QSARs, and mixture effects. *Environ. Sci. Technol.* 36:4201-4217.

Escher, B.I., R. Ashauer, S. Dyer, J.L.M. Hermens, J.-H. Lee, H.A. Leslie, P. Mayer, J.P. Meador, and M.S.J. Warne. 2011. Crucial role of mechanisms and modes of toxic action for understanding tissue residue toxicity and internal effect concentrations of organic chemicals. *Int. Environ. Assess. Manage.*7:28-49

Exponent, Inc. 2010. Weight of Evidence Approach. Draft appendix for addition to British Columbia technical guidance for Detailed Ecological Risk Assessment (DERA). March, 2010.

Fairbrother, A. and R.S. Bennett. 2000. Multivariate statistical applications for addressing multiple stresses in ecological risk assessments. Pages 69-115 in Ferenc, S.A. and J.A. Foran (eds.). *Multiple Stressors in Ecological Risk and Impact Assessment: Approaches in Risk Estimation*. SETAC Press, Pensacola, FL.

Finkel, A.M. 1990. *Confronting Uncertainty in Risk Management: A Guide for Decision Makers*. Center for Risk Management, Resources for the Future, Washington, D.C.

Forbes, V.E. and P. Calow. 2002. Extrapolation in ecological risk assessment: balancing pragmatism and precaution in chemical controls legislation. *BioScience* 52(3): 249-257.

Freshman, J.S. and C.A. Menzie. 1996. Two wildlife exposure models to assess impacts at the individual and population levels and the efficacy of remedial actions. *Human Health and Ecological Risk Assessment* 2(3):481-498.

Gallegos, P., J. Lutz, J. Markwiese, R. Ryti, and R. Mirenda. 2007. Wildlife ecological screening levels for inhalation of volatile organic chemicals. *Environ. Toxicol. Chem*. 26(6): 1299-1303.

Gilbert, R.O. 1987. *Statistical Methods for Environmental Pollution Monitoring*. Van Norstrand Reinhold, New York, New York.

Haas, C.N. 1999. On modeling correlated random variables in risk assessment. *Risk Anal.* 19:1205-1214.

Hardin, J.W., and J.M. Hilbe. 2007. *Generalized Linear Models and Extensions*, 2nd Edition. Strata Press, January 2007.

Harrison, R M. 2001. Pollution: causes, effects and control. Cambridge, Royal Society of Chemistry.

Hill, R.A., P.M. Chapman, G.S. Mann, and G.S. Lawrence. 2000. Level of detail in ecological risk assessment. *Mar. Poll. Bull.* 40(6):471-477.

Hope, B.K., W.T. Wickwire and M.S. Johnson. 2011. The need for increased acceptance and use of spatially explicit wildlife exposure models. *Int. Environ. Assess. Manage.* 7(2):156-157.

Hoffman, F.O. and J.S. Hammonds. 1994. Propagation of uncertainty in risk assessments: The need to distinguish between uncertainty due to lack of knowledge and uncertainty due to variability. *Risk Anal*. 14:707-712.

Hull, R.N., and S. Swanson. 2006. Sequential analysis of lines of evidence – an advanced weight of evidence approach for ecological risk assessment. *Int. Environ. Assess. Manage.* 2(4):302-311.

Hurlbert, S.H. 1984. Pseudoreplication and the design of ecological field experiments. Ecological Monographs 54(2): 187-211.

Kamo, M., and T. Nagai. 2008. An application of the biotic ligand model to predict the toxic effects of metal mixtures. *Environmental Toxicology and Chemistry* 27:1479-1487.

Krebs, C.J. 1989. *Ecological Methodology*. Harper-Collins, New York.

Kooijman, S.A.L.M. 1987. A safety factor for LC_{50} values allowing for differences in sensitivity among species. *Water Res.* 21:269–276.

Landis, W.G, and P.M. Chapman. 2011. Well Past Time to Stop Using NOELs and LOELs. *Int. Environ. Assess. Manage.* 7(4):vi–viii.

Landis, W.G., R.M. Sofield and M. Yu. 2011. Introduction to Environmental Toxicology: Molecular Substructures to Ecological Landscapes. Fourth Edition, CRC Press.

Lee, G. F. and A. Jones-Lee. 2002. *Appropriate Use of Chemical Information in a Best Professional Judgment Triad Weight of Evidence Evaluation of Sediment Quality*. Report of G. Fred Lee & Associates, El Macero, CA. Published in part in the Proceedings of Sediment Quality Assessment (SQA5), Aquatic Ecosystem Health and Management Society, Chicago, IL.

Linkov, I., D. Loney, S. Cormier, F.K. Satterstrom and T. Bridges. 2009. Weight-of-evidence evaluation in environmental assessment: Review of qualitative and quantitative approaches. *Sci. Tot. Environ.* 407(19):5199-5205.

Lock, K., K.A.C. De Schamphelaere, S. Becaus, P. Criel, H. Van Eeckhout, and C.R. Janssen. 2006. Development and validation of an acute biotic ligand model (BLM) predicting cobalt toxicity in soil to the potworm Enchytraeus albidus. *Soil Biology & Biochemistry* 38:1924-1932.

Lock, K., K.A.C. De Schamphelaere, S. Becaus, P. Criel, H. Van Eeckhout, and C.R. Janssen. 2007. Development and validation of a terrestrial biotic ligand model predicting the effect of cobalt on root growth of barley (Hordeum vulgare). *Environmental Pollution* 147:626-633.

Loos, M., A.M. Schipper, U. Schlink, K. Strebel, and A.M.J. Ragas. 2010. Receptor-oriented approaches in wildlife and human exposure modeling: a comparative study. *Environmental Modelling and Software* 25:369-382.

MacDonald, D.D., T. Berger, K. Wood, J. Brown, T. Johnsen, M.L. Haines, K. Brydges, M.J. MacDonald, S.L. Smith, and D.P. Shaw. 1999. *A Compendium of Environmental Quality Benchmarks*. Report prepared for Environment Canada.

McCarty, L.S., P.F. Landrum, S.N. Luoma, J.P. Meador, A.A. Merten, B.K. Shephard, and A.P. van Wezel. 2011. Advancing environmental toxicology through chemical dosimetry: External exposures versus tissue residues. *Int. Environ. Assess. Manage.* 7:7-27.

McElroy, A.E., M.G. Barron, N. Beckvar, S.B.K. Driscoll, J.P. Meador, T.F. Parkerton, T.G. Preuss, and J.A. Steevens. 2011. A review of the tissue residue approach for organic and organometallic compounds in aquatic organisms. *Int. Environ. Assess. Manage.* 7:50-74.

Meador, J.P., W.J. Adams, B.I. Escher, L.S. McCarty, A.E. McElroy, and K.G. Sappington. 2011. The tissue residue approach for toxicity assessment: Findings and critical reviews from a Society of Environmental Toxicology and Chemistry Pellston Workshop. *Int. Environ. Assess. Manage.*7:2-6.

Markwiese, J.T., B. Tiller, R.T. Ryti and R. Bauer. 2008. Using artificial burrows to evaluate inhalation risks to burrowing mammals. *IEAM* 4(4): 425-430.

Menzie, C., N. Bettinger, A. Fritz, L. Kapustka, H. Regan, V. Møller, and H. Noel. 2008. Population Protection Goals. In: L.W. Barnthouse, W.R. Munns, Jr., and M.T. Sorensen (eds.), *Population-level Ecological Risk Assessment*. SETAC Press, Boca Raton, USA.

Menzie, C.A., M.H. Henning, J. Cura, K. Finkelstein, J. Gentile, J. Maughan, D. Mitchell, S. Petron, B. Potocki, S. Svirsky, and P. Tyler. 1996. Special report of the Massachusetts weight-of-evidence workgroup: A weight-of-evidence approach for evaluating ecological risks. *Huanm Ecol. Risk Assess.* 2(2):277–304.

Menzie, CA., L.M. Ziccardi, Y.W. Lowney, A. Fairbrother, S.S. Shock, J.S. Tsuji, D. Hamai, D. Proctor, E. Henry, S.H. Su, M.W. Kierski, M.E. McArdle, and L.J. Yost. 2009. Importance of Considering the Framework Principles in Risk Assessment for Metals. *Environ. Sci. Technol*. 43:8478-8482.

Meyer, J. 1999. A mechanistic explanation for the ln(LC50) vs. ln(hardness) adjustment equation for metals. *Environ. Sci. Technol*. 33:908-912.

Miller, R.W., and D.T. Gardiner. 2003. *Soils in Our Environment*. 10th Edition. Prentice Hall, Englewood Cliffs.

Mirsal, I. 2009. *Soil Pollution: Origin, Monitoring and Remediation.* Springer-Verlag New York.

Morgan, M.M., and M. Henrion. 1990. *Uncertainty – A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. Cambridge University Press, New York.

Mudroch, A, and S.D. MacKnight [Eds]. 2000. *Handbook of Techniques for Aquatic Sediments Sampling*, 2nd edition. Lewis Publishers – CRC Press, Boca Raton, Florida.

Nagy, K.A. 1987. Field metabolic rate and food requirement scaling in mammals and birds. *Ecological Monographs* 57(2):111-128.

NRC (National Research Council). 2009. *Science and Decisions: Advancing Risk Assessment*. National Academies Press, Washington, D.C.

Newman, M.C. 2008. "What exactly are you inferring?" – A closer look at hypothesis testing. *Environ. Toxicol. Chem.* 27(5):1013-1019.

Newman, M.C., D.R. Ownby, L.C.A. Mezin, D.C. Powell, T.R.L. Christensen, S.B. Lerberg, and B.-A. Anderson. 2000. Applying species sensitivity distributions in ecological risk assessment: Assumptions of distribution type and sufficient numbers of species. *Environ. Toxicol. Chem.* 19:508-515.

OMOE (Ontario Ministry of Environment). 20011. *Rationale for the Development and Application of Soil and Groundwater Standards for use at Contaminated Sites in Ontario*. Standards Development Branch, Ontario Ministry of Environment. (http://www.ene.gov.on.ca/stdprodconsume/groups/lr/@ene/@resources/documents/resource/stdprod_086518.pdf).

OMOE. 2005. *Procedures for the Use of Risk Assessment under Part XV.1 of the Environmental Protection Act*. Standards Development Branch, Ontario Ministry of Environment.

Ozretich, R.J., S.P. Ferraro, J.O. Lamberson, and F.A. Cole. 2000. Test of ΣPAH model at a creosote-contaminated site, Elliot Bay, Washington, USA. *Environ. Toxicol. Chem*. 19:2378-2389.

Pagenkopf, G. 1983. Gill surface interaction model for trace-metal toxicity to fishes: Role of complexation, pH, and water hardness. *Environ. Sci. Technol*. 17:342-347.

Paquin, P.R., K. Farley, R.C. Santore, C.D. Kavvdas, K.G. Mooney, R.P. Windfield, K.B. Wu, and D.M. Di Toro. 2003. *Metals in Aquatic Systems: A Review of Exposure, Bioaccumulation, and Toxicity Models*. Metals in the Environment Series. SETAC Publication.

Pearsons, T.N., and C.W. Hopley**.** 1999**.** A practical approach for assessing. ecological risks associated with fish stocking programs. *Fisheries* 24(9):16-23.

Peters, A., S. Lofts, G. Merrington, B. Brown, W. Stubblefield, and K. Harlow. 2011. Development of biotic ligand models for chronic manganese toxicity to fish, invertebrates, and algae. *Environmental Toxicology and Chemistry* 30:2407-2415.

Pierzynski, G.M., J.T. Sims, G.F. Vance. 2005. *Soils and Environmental Quality*, 3rd *Edition*. Lewis Publishers, Boca Raton.

Pinheiro, J.C., and D.M. Bates. 2000. *Mixed-Effects Models in S and S-Plus*. Springer, New York.

Posthuma, L, G.W. Suter, and T.P. Traas (Eds.). 2002. *Species Sensitivity Distributions in Ecotoxicology*. Boca Raton: Lewis Publishers.

PSAMP (Puget Sound Assessment and Monitoring Program). 1996. *Puget Sound Protocols and Guidelines*. Available at: http://www.psparchives.com/our_work/science/protocols.htm

Québec (Ministère du Développement durable, de l'Environnement et des Parcs). 2000. *Valeurs de Référence Interimaires pour les Récepteurs Terrestres*. Centre d'expertise en analyse environnementale du Québec. Ministère du Développement durable, de l'Environnement et des Parcs, 129 p.

Reynoldson, T.B., D.R. Rosenberg, K.E. Day, , R.H. Norris, and V.H. Resh. 1997. The reference condition: a comparison of multimetric and multivariate approaches to assess water-quality impairment using benthic macroinvertebrates. *J. N. Am. Benth. Soc.* 16:833-852.

Reynoldson, T.B., C. Logan, T. Pascoe and S.P. Thompson. 2006. *CABIN (Canadian Aquatic Biomonitoring Network) Invertebrate Biomonitoring Field and Laboratory Manual*. National Water Research Institute, Environment Canada.

Ritter, L., K. Solomon, P. Sibley, K. Hall, P. Keen, G. Mattu, and B. Linton. 2002. Sources, pathways and relative risks of contaminants in surface water and groundwater: a perspective prepared for the Walkerton inquiry. *J. Toxicol. Environ. Health A*. 65:1142.

Ruby, M.V., A. Davis, R. Schoof, S. Eberle, and C.M. Sellstone. 1996. Estimation of Lead and Arsenic Bioavailability Using a Physiologically Based Extraction Test. *Environmental Science & Technology* 30:422-430.

Sample, B.E., J.J. Beauchamp, R.A. Efroymson and G.W. Suter II. 1998. *Development and Validation of Bioaccumulation Models for Small Mammals*. Report prepared for the US Department of Energy, Office of Environmental Management.

Santore, R., D. Di Toro, P.R. Paquin, H.E. Allen, and J.S. Meyer. 2001. A biotic ligand model of the acute toxicity of metals. 2. Application to acute copper toxicity in freshwater fish and *Daphnia*. *Environ. Toxicol. Chem*. 20:2397-2402.

Santore, R.C., R. Mathew, P. Paquin, and D. Di Toro. 2002. Application of the biotic ligand model to predicting zinc toxicity to rainbow trout, fathead minnow, and *Daphnia magna*. *Comp. Biochem. Physiol. Part C* 133:271-285.

Sappington, K.G., T.S. Bridges, S.P. Bradbury, R.J. Erickson, A.J. Hendriks, R.P. Lanno, J.P. Meador, D.R. Mount, M.H. Salazar, and D.J. Spry. 2011. Application of the tissue residue approach in ecological risk assessment. *Int. Environ. Assess. Manage.* 7:116-140.

SAB-CS (Science Advisory Board for Contaminated Sites in British Columbia). 2008. *Guidance for Detailed Ecological Risk Assessments (DERA) in British Columbia*. British Columbia, Canada.

Schwartz, M.L., and B. Vigneault. 2007. Development and validation of a chronic copper biotic ligand model for Ceriodaphnia dubia. *Aquatic Toxicology* 84:247-254.

Sparks, T.H., W.A. Scott, and R.T. Clarke, R.T. 1999. Traditional multivariate techniques: potential for use in ecotoxicology. *Environmental Toxicology and Chemistry* 18:128-137.

Stoddard, J.L., D.P. Larsen, C.P. Hawkins, R.K. Johnson and R.H. Norris. 2006. Setting expectations for the ecological condition of streams: the concept of reference condition. Ecological Applications 16(4): 1267-1276.

Surface Water Ambient Monitoring Program (SWAMP). 2009. *Final Technical Report - Recommendations for the Development and Maintenance of a Reference Condition Management Program (RCMP) to Support Biological Assessment of California's Wadeable Streams*. March 2009.

Suter, G.W. II. 1996. *Guide for Developing Conceptual Models for Ecological Risk Assessments*. ES/ER/TM-186. Office of Environmental Management, US Department of Energy, Oak Ridge, TN, USA.

Suter, G.W. 2007. *Ecological Risk Assessment*. 2nd edition. CRC Press, Boca Raton, Florida.

Suter, G.W., R.A. Efroymson, B.E. Sample and D.S. Jones. 2000. *Ecological Risk Assessment for Contaminated Sites*. Lewis Publishers.

Suter, G.W., S.B. Norton and A. Fairbrother. 2005. Individuals versus organisms versus populations in the definition of ecological assessment endpoints. *Int. Environ. Assess. and Manage.* 1(4):397-400.

Suter, G.W. II, S.B. Norton and S.M. Cormier. 2010. The science and philosophy of a method for assessing environmental causes. *Human and Ecological Risk Assessment* 16:19-34.

Suter, GW II, and S.M. Cormier. 2011. Why and how to combine evidence in environmental assessments: Weighing evidence and building cases. *Science of the Total Environment* 409: 1406-1417.

Swartz, R.C., D.W. Schults, R.J. Ozretich, J.O. Lamberson, F.A. Cole, , T.H. De Witt, M.S. Redmond, and S.P. Ferraro. 1995. ΣPAH: A model to predict the toxicity of polynuclear aromatic hydrocarbon mixtures in field-collected sediments, *Environ. Toxicol. Chem.* 14:1977-1987.

Tessier A, P.G.C Campbell and M. Bisson. 1979 Sequential extraction procedure for the speciation of particulate trace metals. *Analytical Chemistry* 51(7):845-851.

United States Department of Energy. 1998. *Risk-Based Corrective Action Guide*. U.S. Department of Energy, Office of Environmental Policy & Assistance, RCRA/CERCLA Division, Washington, D.C. November 1998.

Thakali, S., H.E. Allen, D.M. Di Toro, A.A. Ponizovsky, C.P. Rooney, F.-J. Zhao, and S.P. McGrath. 2006a. A Terrestrial Biotic Ligand Model. 1. Development and Application to Cu and Ni Toxicities to Barley Root Elongation in Soils. *Environmental Science & Technology* 40:7085-7093.

Thakali, S., H.E. Allen, D.M. Di Toro, A.A. Ponizovsky, C.P. Rooney, F.-J. Zhao, S.P. McGrath, P. Criel, H. Van Eeckhout, C.R. Janssen, K. Oorts, and E. Smolders. 2006b. Terrestrial Biotic Ligand Model. 2. Application to Ni and Cu Toxicities to Plants, Invertebrates, and Microbes in Soil. *Environmental Science & Technology* 40:7094-7100.

USEPA (United States Environmental Protection Agency). 1992. *Framework for Ecological Risk Assessment*. Office of Research and Development, US Environmental Protection Agency, Washington, DC. EPA/630/R92/001.

USEPA. 1993. *Wildlife Exposure Factors Handbook*. Vol.1. EPA/600/R-93/187a. Office of Research and Development, Washington, DC.

USEPA. 1997a. *Policy for Use of Probabilistic Analysis in Risk Assessment*. May 5, 1997. Available online at: http://www.epa.gov/osa/spc/pdfs/probpol.pdf

USEPA. 1997b. *Guiding Principles for Monte Carlo Analysis*. Risk Assessment Forum, Washington, DC. EPA/630/R-97/001.

USEPA. 1998. *Guidelines for Ecological Risk Assessment*. Environmental Protection Agency, Office of Research and Development, Washington, DC. EPA/630/R-95/002F.

USEPA. 1999. *Integrated Approach to Assessing the Bioavailability and Toxicity of Metals in Surface Waters and Sediments*. Presented to the EPA Science Advisory Board. April 1999. USEPA Office of Water, Washington, DC.

USEPA. 2000. *Risk Characterization Handbook.* U.S. Environmental Protection Agency, Office of Science Policy, Washington, DC. 60pp. EPA 100-B-00-002.

USEPA. 2001. *Risk Assessment Guidance for Superfund (RAGS), Volume III - Part A: Process for Conducting Probabilistic Risk Assessment*. Office of Emergency and Remedial Response, U.S. EPA, Washington, D.C. EPA 540-R-02-002. Available at http://www.epa.gov/oswer/riskassessment/rags3adt/index.htm

US EPA (United States Environmental Protection Agency). 2003. Evaluation of dermal contact and inhalation exposure pathways for the purpose of setting Eco-SSLs. Attachment 1-3 of the Guidance for Developing Ecological Soil Screening Levels. Prepared November 2003.

USEPA. 2006. Guidance on systematic planning using the data quality objectives process. Washington, DC. EPA/240/B-06/001.

USEPA. 2007a. *Aquatic Life Ambient Freshwater Quality Criteria – Copper*. Office of Water, Washington, DC. EPA-822-R-07-001.

USEPA. 2007b. *Framework for Metals Risk Assessment*. Office of the Science Advisor. Risk Assessment Forum. March 2007. EPA 120/R-07/001.

USEPA. 2007c. *Statistical Primer: Multivariate Methods*. Last updated on December 20th, 2007. Available at: http://www.epa.gov/bioindicators/statprimer/index.html

USEPA. 2007d. *Operating Procedure for Porewater Sampling*. Region 4, Science and Ecosystem Support Division, Athens, Georgia. SOP # SESDPROC-513-R0

USEPA. 2011. *Integrating Ecological Assessment and Decision-Making at EPA: A Path Forwar*d. EPA/100/R-10/004. Risk Assessment Forum, Washington, DC.

Wandall, B. 2004. Values in science and risk assessment. *Toxicol. Lett.* 152:265-272.

WDNR (Wisconsin Department of Natural Resources). 2009. *Guidance for Evaluating Intake Structures Using Best Professional Judgment (BPJ)*. Bureau of Watershed Management Program Guidance, Wastewater Policy Management Team. Developed under Section 316(b) of the *Clean Water Act* and Section 283.31(6) of Wisconsin Statutes. February 2, 2009.

Wheeler, M.W., and A.J. Bailer. 2009. Benchmark dose estimation incorporating multiple data sources. *Risk Analysis* 29(2):249-256.

Wickwire, T., M.S. Johnson, B.K. Hope and M.S. Greenberg. 2011. Spatially explicit ecological exposure models: a rationale for and path toward their increased acceptance and use. *Int. Environ. Assess. Manage.* 7(2):158-168.

www.ec.gc.ca

Additional information can be obtained at:

Environment Canada
Inquiry Centre
10 Wellington Street, 23rd Floor
Gatineau QC K1A 0H3
Telephone: 1-800-668-6767 (in Canada only) or 819-997-2800
Fax: 819-994-1412
TTY: 819-994-0736
Email: enviroinfo@ec.gc.ca