



Environment
Canada

Environnement
Canada



METAL MINING TECHNICAL GUIDANCE FOR ENVIRONMENTAL EFFECTS MONITORING

2012

Canada

ISBN 978-1-100-20496-3

Cat. no.: En14-61/2012E-PDF

Information contained in this publication may be reproduced, in part or in whole, and by any means, for personal or public non-commercial purposes, without charge or further permission, unless otherwise specified.

You are asked to:

- Exercise due diligence in ensuring the accuracy of the materials reproduced;
- Indicate both the complete title of the materials reproduced, as well as the author organization; and
- Indicate that the reproduction is a copy of an official work that is published by the Government of Canada and that the reproduction has not been produced in affiliation with or with the endorsement of the Government of Canada.

Commercial reproduction and distribution is prohibited except with written permission from the Government of Canada's copyright administrator, Public Works and Government Services of Canada (PWGSC). For more information, please contact PWGSC at 613-996-6886 or at droitdauteur.copyright@tpsgc-pwgsc.gc.ca.

Photos: © Environment Canada

© Her Majesty the Queen in Right of Canada represented by the Minister of the Environment, 2012

Aussi disponible en français

DISCLAIMER

The objective of this document is to provide guidance to mines on how to meet the environmental effects monitoring regulatory requirements under the *Metal Mining Effluent Regulations* (MMER). This is not a legal interpretation of the MMER. For the Regulations, refer to the MMER, available at <http://laws-lois.justice.gc.ca/eng/regulations/SOR-2002-222/FullText.html>.

ACKNOWLEDGEMENTS

The National Environmental Effects Monitoring (EEM) Office would like to thank the many people who contributed to the updating of this technical guidance document. The content was greatly improved by contributions from the members of the EEM National Team and Science Committee. The quality of the document was vastly improved through the efforts of Environment Canada's editing team and individual members of the National EEM Office

List of Acronyms

AAS: atomic absorption spectrophotometry
AES: Auger electron spectrometry
AETE Program: Aquatic Effects Technology Evaluation Program
ANCOVA: analysis of covariance
ANOVA: analysis of variance
AQUAMIN: Assessment of the Aquatic Effects of Mining in Canada
APHA: American Public Health Association
AVS: acide volatile sulphides
EPS: Environmental Protection Service of Environment Canada
ASPT: average score per taxon
ASTM: American Society for Testing and Materials
ATW: Aquatic Toxicity Workshop
AWWA: American Water Works Association
BACI: before/after control-impact
BAR: B.A.R. Environmental Inc.
B-C Index: Bray-Curtis Index
BEAK: Beak International Inc.
BL: biotic ligand
BLM: biotic ligand model
BMWP: biological monitoring working party
CABIN: Canadian Aquatic Biomonitoring Network
CAEAL: Canadian Association for Environmental Analytical Laboratories
CALA: Canadian Association for Laboratory Accreditation
CALK: combined alkaline stream
CBR: critical body residues
CCME: Canadian Council of Ministers of the Environment
CES: critical effect size
CETTP: Complex Effluent Toxicity Testing Program
C-I: control-impact
COSEWIC: Committee on the Status of Endangered Wildlife in Canada
CPUE: catch per unit effort
CVAAS: cold vapour atomic absorption spectrometry
CVAFS: cold vapour atomic fluorescence spectrometry

D.L.: detection limit
DDW: double distilled water
df: degrees of freedom
DGT: diffusive gradient thin film
DOC: dissolved organic carbon
DQOs: data quality objectives
EC: Environment Canada
EC₂₅: 25% effect concentration
EC₅₀: 50% effect concentration
EDA: effect directed analysis
EDTA: Ethylenediaminetetraacetic acid
EEM: environmental effects monitoring
ELAP: Environmental Laboratory Approval Program
ESG: ESG International Inc.
EU: European Union
exp.: exposure
FDP: final discharge point
FF: far-field
FRAP: Fraser River Action Plan
GC: gas chromatography
GFAAS: graphite furnace atomic absorption spectrometry
GLP: good laboratory practice
GIME: gel integrated minielectrode
GM-IC₂₅: geometric mean of all IC₂₅s
GPS: global positioning system
GSI: gonadosomatic index
HALW: hardness-adjusted laboratory water
HFPLM: hollow fibre permeation liquid membrane
HPLC: high performance liquid chromatography
HSB: hyper-saline brine
IC₂₅: 25% inhibition concentration
IC₅₀: 50% inhibition concentration
ICP-AES: inductively coupled atomic absorption spectrophotometry
ICP-MS: inductively coupled plasma mass spectrometry
ID: internal diameter
INRS: Institut national de la recherche scientifique
INAA: instrumental neutron activation analysis
IOC: investigation of cause
IOS: investigation of solutions
IARC: International Agency for Research on Cancer
ISO/IEC: International Organization for Standardization
LC₅₀: median lethal concentration
LCL: lower control limit
LIMS: laboratory information management system
LOE: lines of evidence

LPL: lowest practical level
LSI: liver somatic index
LT₂₅: time to 25% mortality
LT₅₀: time to 50% mortality
LWL: lower warning limit
MC-I: multiple control-impact
MDDEP: Ministère du Développement durable, de l'Environnement et des Parcs du Québec
MDL: method detection limit
MFO: mixed function oxygenase
MG: multiple gradient
MME: metal mine effluent
MMER: *Metal Mining Effluent Regulations*
MOE: Ministry of the Environment
MS: mass spectrometry
MS: mean square
MSE: municipal sewage effluent
MSI: mantle somatic index
MT: metallothionein
NABS: North American Benthological Society
NAMC: The North American Metals Council
NF: near-field
NOM: natural organic matter
NRBS: Northern River Basins Study
NSERC: Natural Sciences and Engineering Research Council
OECD: Organisation for Economic Co-operation and Development
PLC: Public Liaison Committee
PME: pulp mill effluent
PPER: *Pulp and Paper Effluent Regulations*
QA/QC: quality assurance / quality control
R²: coefficient of variation
RCA: reference condition approach
ref.: reference
RG: radial gradient
RISS: Regulatory Information Submission System
SD: standard deviation
SE: standard error
SEC: size exclusion chromatography
sem: simultaneous extracted metals
SEM: scanning electron microscopy
SETAC: Society of Environmental Toxicology and Chemistry
SG: simple gradient
SIMS: secondary ion mass spectrometry
SOP: standard operating procedure
SPE: solid phase extraction

SQT: sediment quality triad
SRM: standard reference material
SS: sum of squares
TER: toxicity emission rate
TIE: toxicity identification evaluation
TKN: Total Kjeldahl Nitrogen
TRE: toxicity reduction evaluation
TOC: total organic carbon
TSRI: Toxic Substances Research Initiative
U.S. EPA: United States Environmental Protection Agency
UCL: upper control limit
UWL: upper warning limit
UV: ultraviolet
VECs: valued ecosystem components
WAWW: whole-animal wet weight
WEF: Water Environment Federation
WHO: World Health Organization
WMI: Whitehorse Mining Initiative
WQG: water quality guideline
XPS: X-ray photoelectron spectroscopy
XRF: X-ray fluorescence
XAFS: X-ray fluorescence spectroscopy
XRM: X-ray microanalysis
YOY: young of the year

TABLE OF CONTENTS

Chapter

- 1. OVERVIEW OF THE METAL MINING ENVIRONMENTAL EFFECTS MONITORING PROGRAM**
- 2. STUDY DESIGN, SITE CHARACTERIZATION AND GENERAL QUALITY ASSURANCE AND QUALITY CONTROL**
- 3. EFFECTS ON FISH AND FISHERIES RESOURCES**
- 4. EFFECTS ON FISH HABITAT: BENTHIC INVERTEBRATE COMMUNITY SURVEY**
- 5. EFFLUENT CHARACTERIZATION AND WATER QUALITY MONITORING**
- 6. SUBLETHAL TOXICITY TESTING**
- 7. SEDIMENT MONITORING**
- 8. DATA ASSESSMENT AND INTERPRETATION**
- 9. ALTERNATIVE MONITORING METHODS**
- 10. INFORMATION MANAGEMENT AND INTERPRETATIVE REPORTS**
- 11. PUBLIC INVOLVEMENT IN METAL MINING ENVIRONMENTAL EFFECTS MONITORING**
- 12. INVESTIGATION OF CAUSE**
- 13. REPORT ON HISTORICAL INFORMATION**

Table of Contents

1.	Overview of the Metal Mining Environmental Effects Monitoring Program	1-1
1.1	Purpose of the Guidance Document	1-1
1.2	The <i>Metal Mining Effluent Regulations</i>	1-2
1.3	Description of Environmental Effects Monitoring Studies	1-3
1.3.1	Effluent and Water Quality Monitoring Studies.....	1-3
1.3.1.1	Effluent Characterization.....	1-3
1.3.1.2	Sublethal Toxicity Testing.....	1-4
1.3.1.3	Water Quality Monitoring	1-4
1.3.2	Biological Monitoring Studies.....	1-4
1.3.2.1	Defining and Confirming Effects.....	1-5
1.3.2.2	Historical Information.....	1-6
1.3.2.3	Biological Monitoring Studies to Assess Effects	1-6
1.3.2.3.1	Fish Population Survey	1-6
1.3.2.3.2	Benthic Invertebrate Community Survey	1-7
1.3.2.3.3	Fish Tissue Survey.....	1-8
1.3.2.4	Biological Monitoring Studies to Investigate Effects.....	1-8
1.3.2.4.1	Magnitude and Geographic Extent	1-8
1.3.2.4.2	Investigation of Cause	1-8
1.4	Steps in Conducting and Reporting Environmental Effects Monitoring Studies.....	1-8
1.4.1	Conduct and Submit Results for Effluent Characterization, Sublethal Toxicity Testing and Water Quality Monitoring	1-9
1.4.2	Submit Study Design	1-9
1.4.2.1	Study Design for Biological Monitoring Studies to Assess Effects ..	1-10
1.4.2.2	Study Design for Biological Monitoring to Investigate Effects	1-11
1.4.3	Conduct Biological Monitoring Study.....	1-11
1.4.4	Conduct Data Assessment.....	1-11
1.4.5	Submit Interpretative Report	1-12
1.4.5.1	Interpretative Report for Biological Monitoring Studies to Assess Effects.....	1-12
1.4.5.2	Interpretative Report for Biological Monitoring Studies to Investigate Effects.....	1-12
1.5	Recognized Closed Mines	1-12
1.6	Identifying a Path through the Metal Mining Environmental Effects Monitoring Program.....	1-13
1.6.1	Critical Effect Sizes	1-13
1.6.2	Magnitude of Confirmed Effects	1-14

1.6.3	Decision Process for the Metal Mining Environmental Effects Monitoring Program.....	1-14
1.6.3.1	Assessing Effects	1-17
1.6.3.2	Investigating Confirmed Effects	1-17
1.6.3.2.1	Level of Effort for Investigating Effects.....	1-18
1.6.3.3	Timing of Studies to Assess Effects and Magnitude and Geographic Extent and to Investigate Cause.....	1-18
1.7	References.....	1-20

List of Tables

Table 1-1:	Effect indicators and endpoints for the fish population survey.....	1-7
Table 1-2:	Effect indicators and endpoints for the benthic invertebrate community survey	1-7
Table 1-3:	Critical effect sizes for metal mining environmental effects monitoring program	1-14
Table 1-4:	Evaluation of magnitude of confirmed effects in two consecutive phases ..	1-14

List of Figures

Figure 1-1:	Decision tree for the metal mining EEM program.....	1-16

1. Overview of the Metal Mining Environmental Effects Monitoring Program

1.1 Purpose of the Guidance Document

In 1996, Environment Canada undertook an assessment of the aquatic effects of mining in Canada (AQUAMIN 1996¹), which provided recommendations regarding the review and amendments of the *Metal Mining Liquid Effluent Regulations* (currently titled the *Metal Mining Effluent Regulations* [MMER]) and the design of a national Environmental Effects Monitoring (EEM) program for metal mining. The MMER under the *Fisheries Act* direct metal mines to conduct EEM as a condition governing the authority to deposit effluent (MMER, Part 2, section 7). EEM is a science-based performance measurement tool used to assess the adequacy of the regulations. Although this guidance document is not a legal document, it is intended to provide guidance for mines in meeting their EEM requirements and conducting EEM studies. For the regulatory EEM requirements, refer to section 7 and Schedule 5 of the MMER, located on the following website: <http://laws-lois.justice.gc.ca/eng/regulations/SOR-2002-222/FullText.html>. This guidance document replaces the 2002 version.

The MMER prescribes that EEM studies be performed using documented and validated methods, and that their results be interpreted and reported in accordance with generally accepted standards of good scientific practice (MMER, Part 2, subsection 7(3)). The methodologies recommended in this guidance document are based on generally accepted standards of good scientific practice, and incorporate improvements based on program experience, input from multi-stakeholder working groups, and external research initiatives responding to EEM needs. As well, the Metal Mining EEM Review Team, which comprised a group of experts from government, industry, and environmental and Aboriginal groups, was established by Environment Canada to examine the experiences and results of the EEM program from the first phase of metal mining EEM studies and to provide recommendations to Environment Canada for improving the program. The final report, *Review Team Report from Metal Mining Environmental Effects Monitoring Program (August 2007)*, is available on the EEM website at www.ec.gc.ca/eem. This document also reflects the changes to EEM requirements established by the 2006 and 2012 MMER amendments.

It should be emphasized that the methodologies provided in this guidance document do not constitute an exhaustive list of the possible means of conducting EEM. It is assumed that each study leader has sufficient knowledge to apply these recommendations using generally accepted standards of good scientific practice and is able to determine if unique conditions exist that would warrant modification of the generic study designs, while ensuring that regulatory requirements are met. For a scientific discussion on sound

¹ AQUAMIN. 1996. Assessment of the Aquatic Effects of Mining in Canada. Environment Canada.

science, refer to Chapter 12 on investigation of Cause. Mines are encouraged to contact the Environment Canada regional EEM coordinators² for EEM-related questions.

This first chapter provides an overview of the metal mining EEM program, including a decision tree to assist mines in identifying an appropriate path, based on their respective situation, as they move through the EEM program. Additional information and documents are available on the EEM website (www.ec.gc.ca/eem).

1.2 The *Metal Mining Effluent Regulations*

The MMER permit the deposit of mine effluent if the effluent pH is within a defined range, if the concentrations of the MMER deleterious substances in the effluent do not exceed authorized limits, and if the effluent is demonstrated to be non-acutely lethal to rainbow trout. These discharge limits were established to be minimum national standards based on best available technology economically achievable at the time that the MMER were promulgated. To assess the adequacy of the effluent regulations for protecting the aquatic environment, the MMER include EEM requirements to evaluate the potential effects of effluents on fish, fish habitat and the use of fisheries resources.

Regulations Amending the MMER were published in the *Canada Gazette*, Part II, in October 2006. The purpose of these amendments was to clarify the regulatory requirements by addressing matters, related to the interpretation and clarity of the regulatory text, which had emerged from the implementation of the Regulations. Additional amendments also occurred at later dates and on several occasions; however, these did not pertain to the EEM requirements.

Additional amendments to the MMER were published in the *Canada Gazette*, Part II, in February 2012. The following changes were made to improve the EEM provisions of the MMER:

- modifications to the definition of an “effect on fish tissue” in order to be consistent with the Health Canada fish consumption guidelines and to clarify that the concentration of total mercury in tissue of fish from the exposure area must be statistically different from and higher than its concentration in fish tissue from the reference area;
- addition of selenium and electrical conductivity to the list of parameters required for effluent characterization and water quality monitoring;
- exemption for mines, other than uranium mines, from monitoring radium 226 as part of the water quality monitoring, if 10 consecutive test results showed that radium 226 levels are less than 10% of the authorized monthly mean concentration (see subsection 13(2) of the Regulations);
- change to the time frame for the submission of interpretative reports for mines with effects on the fish population, fish tissue and benthic invertebrate community from 24 to 36 months;

² Contact information for regional EEM coordinators is available on the EEM website: www.ec.gc.ca/eem.

- change to the time frame for the submission of interpretative reports for magnitude and geographic extent of effects and for investigation of cause of effects from 24 to 36 months; and
- minor changes to the wording for consistency within Schedule 5.

1.3 Description of Environmental Effects Monitoring Studies

EEM studies are designed to detect and measure changes in aquatic ecosystems (i.e., receiving environments). The metal mining EEM program is an iterative system of monitoring and interpretation phases that is used to assess the effectiveness of environmental management measures, by evaluating the effects of effluents on fish, fish habitat and the use of fisheries resources by humans.

EEM goes beyond the end-of-pipe measurement of chemicals in effluent to examine the effectiveness of environmental protection measures directly in aquatic ecosystems. Long-term effects are assessed using regular cyclical monitoring and interpretation phases designed to investigate the impacts on the same parameters and locations. In this way, both a spatial and temporal characterization of potential effects to assess changes in receiving environments are obtained.

EEM studies consist of:

- effluent and water quality monitoring studies comprising effluent characterization, sublethal toxicity testing and water quality monitoring (MMER, Schedule 5, Part 1); and
- biological monitoring studies in the aquatic receiving environment to determine if mine effluent is having an effect on fish, fish habitat or the use of fisheries resources (MMER, Schedule 5, Part 2).

1.3.1 Effluent and Water Quality Monitoring Studies

1.3.1.1 Effluent Characterization

Effluent characterization is conducted by analyzing a sample of effluent and recording the hardness, electrical conductivity and alkalinity, as well as the concentrations of aluminum, cadmium, iron, molybdenum, selenium, ammonia and nitrate (MMER, Schedule 5, subsection 4(1)). Mercury in effluent is also analyzed and recorded but may be discontinued if the concentration is less than 0.10 µg/L in 12 consecutive samples (MMER, Schedule 5, subsection 4(3)). Guidance on effluent characterization can be found in Chapter 5. Other parameters relevant to EEM, such as arsenic, copper, cyanide, lead, nickel, zinc, total suspended solids and radium 226, are also analyzed as part of Schedule 4.

1.3.1.2 Sublethal Toxicity Testing

Sublethal toxicity testing is conducted on effluent from the mine's final discharge point that has potentially the most adverse environmental impact (MMER, Schedule 5, subsection 5(2)). This testing monitors effluent quality by measuring survival, growth and/or reproduction endpoints in marine or freshwater organisms in a controlled laboratory environment. In the case of effluent deposited into marine or estuarine waters, sublethal toxicity testing is conducted on a fish species, an invertebrate species and an algal species. In the case of effluent deposited into freshwater, sublethal toxicity testing is conducted on a fish species, an invertebrate species, a plant species and an algal species (MMER, Schedule 5, subsection 5(1)). Guidance to determine the appropriate final discharge point to sample can be found in Chapter 2. Guidance on sublethal toxicity testing can be found in Chapter 6.

1.3.1.3 Water Quality Monitoring

Samples for water quality monitoring are collected from the exposure area surrounding the point of entry of the effluent into water from each final discharge point and the related reference areas, as well as from the sampling areas selected for the biological monitoring study (MMER, Schedule 5, subsection 7(1)). Water temperature and dissolved oxygen concentrations are recorded for all samples. As for effluent characterization, the concentrations of aluminum, cadmium, iron, molybdenum, selenium, ammonia and nitrate are measured and recorded in water quality monitoring. Mercury in the water quality monitoring samples is also analyzed and recorded if required for effluent characterization (MMER, Schedule 5, subsection 4(3)). In the case of effluent deposited into freshwater, the pH, hardness, electrical conductivity and alkalinity are recorded. In the case of effluent deposited into estuarine waters, the pH, hardness, electrical conductivity, alkalinity and salinity are recorded. In the case of effluent deposited into marine waters, the salinity is recorded. The concentrations of the following deleterious substances set out in Schedule 4 are also recorded: arsenic, copper, cyanide (if used as a process reagent), lead, nickel, zinc, total suspended solids and radium 226 (unless the conditions specified in subsection 13(2) of the Regulations are met) (MMER, Schedule 5, paragraph 7(1)(d)). Guidance on water quality monitoring can be found in Chapter 5.

1.3.2 Biological Monitoring Studies

Biological monitoring studies are conducted in 36- or 72-month phases. The requirements for each study are dependent on the results of studies from the previous phase(s). Biological monitoring studies to assess effects are described in section 1.3.2.3 and biological monitoring studies to investigate effects are described in section 1.3.2.4.

To assess effects, biological monitoring studies are conducted for the following three components (MMER, Schedule 5, section 9):

- **fish population** to assess effects on fish health;
- **benthic invertebrate community** to assess fish habitat or fish food; and
- mercury concentration in **fish tissue** to assess the human usability of the fisheries resources, in terms of fish consumption.

To investigate effects, biological monitoring studies are conducted for the purpose of:

- assessing the magnitude and geographic extent of effects; and
- determining the cause(s) of effects.

1.3.2.1 Defining and Confirming Effects

The studies on the fish population, fish tissue, and benthic invertebrate community are conducted in both exposure and reference areas. The exposure area refers to all fish habitat and waters frequented by fish that are exposed to effluent, and the reference area refers to water frequented by fish that is not exposed to effluent and that has fish habitat that, as far as is practical, is most similar to that of the exposure area (MMER, Schedule 5, section 1).

The MMER defines effects on the fish population, fish tissue, and benthic invertebrate community (MMER, Schedule 5, section 1) and further prescribes the data assessment required for specific indicators (MMER, Schedule 5, section 16). An “effect” on the fish population or benthic invertebrate community is defined as a statistical difference between data collected in an exposure area and in a reference area or sampling areas within an exposure area where there are gradually decreasing effluent concentrations at increasing distances from the effluent discharge. An effect on fish tissue refers to concentrations of total mercury, exceeding 0.5 micrograms per gram ($\mu\text{g/g}$) wet weight in fish tissue taken in an exposure area and that are statistically different from and higher than the concentrations of total mercury in fish tissue taken in a reference area. Chapter 8 provides information on conducting statistical analyses on EEM data.

Data collected on specific-effect endpoints (Tables 1-1 and 1-2) are assessed to determine if statistical differences are present in order to establish if there are any effects on the indicators. To confirm that observed effects are not artifacts (or due to confounding factors) and are mine-related, biological monitoring studies to assess effects are repeated in a subsequent three-year phase. If a similar type of effect (same endpoint in same direction from zero relative to reference levels) on the fish population, fish tissue or benthic invertebrate community is determined in studies from two consecutive phases, the effect is considered confirmed (MMER, Schedule 5, section 19). Confirmation of an effect for fish endpoints need not be limited to the same sex or same species, unless site-specific conditions warrant a different approach.

If effects are confirmed in one or more components (fish population, fish tissue, benthic invertebrate community), the mine must investigate those effects in subsequent phases (section 1.3.2.3). All confirmed effects must be investigated. If the lack of effects is

confirmed in all three components, a mine must proceed to a reduced biological monitoring frequency (MMER, Schedule 5, paragraph 22(2)(b)).

Attributing cause of an effect to a mine's effluent may be difficult in some circumstances. Environment Canada recommends that where the previous study has determined there is an effect, and there is doubt that the effect is caused by the mine effluent, the second study confirming the effect be designed in a way that maximizes the confidence in establishing that the effect is or is not mine-effluent-related. Adjustments to the study design to eliminate confounding factors are described in the other chapters and could include increased sampling effort in both reference and exposure areas; increase or change in sampling areas; or the use of alternative studies, such as mesocosms or caged bivalves.

1.3.2.2 Historical Information

Mines may have conducted biological monitoring studies prior to becoming subject to the MMER. These studies may be used as part of the EEM program if they determine whether the effluent was causing an effect on fish population, fish tissue or the benthic invertebrate community. However, if the mine operation or environmental conditions changed or any event which may have modified biological effects occurred after the historical monitoring was conducted, then any historical data should be used with caution when interpreting currently observed effects. The results of the historical biological monitoring studies must be submitted to the Authorization Officer³ along with a report that contains scientific data to support the results, not later than 12 months after the day on which the mine becomes subject to the Regulations (MMER, Schedule 5, paragraph 14(b)). Refer to sections 1.4.2 and 1.4.5 for requirements on timelines for submission of study design and interpretative reports for mines using historical information. Further details on historical information are provided in Chapter 13.

1.3.2.3 Biological Monitoring Studies to Assess Effects

To assess effects, biological monitoring studies are conducted for the three components: fish population, fish tissue (mercury concentration) and benthic invertebrate community.

1.3.2.3.1 Fish Population Survey

A fish population survey (Chapter 3) measures indicators of fish population health in exposure and reference areas, or along an exposure gradient, to determine if mine effluent has an effect on fish. A fish survey is required if the concentration of effluent in the exposure area is greater than 1% at a distance of 250 metres from the final discharge point (Schedule 5, paragraph 9(b)).

³ The Authorization Officer for each province is described in Schedule 1 of the MMER. Contact information for Authorization Officers is available on the EEM website: <http://www.ec.gc.ca/esee-eem/>.

The MMER defines the fish population survey effect indicators as growth, reproduction, condition and survival (MMER, Schedule 5, subparagraph 16(a)(i)). When conducting a standard adult fish survey, the collection of adult males and females of two sentinel species is recommended. Data collected on the specific effect endpoints listed in Table 1-1 are evaluated to determine if statistical differences in the effect indicators are present.

Table 1-1: Effect indicators and endpoints for the fish population survey

Effect Indicators	Effect Endpoints
Growth (energy use)	Size-at-age (body weight relative to age)
Reproduction (energy use)	Relative gonad size (gonad weight to body weight)
Condition (energy storage)	Condition (body weight to length) Relative liver size (liver weight to body weight)
Survival	Age

Although the standard fish survey is recommended above, other survey designs, modified methods such as a non-lethal fish survey (Chapter 3) or alternative methods (Chapter 9) may be considered under conditions where the standard survey is not effective or practical.

1.3.2.3.2 Benthic Invertebrate Community Survey

Mines must conduct a benthic invertebrate community survey (Chapter 4) to determine if their effluent has an effect on fish habitat. Benthic invertebrates are collected to determine if there are changes in the effect indicators between exposure and reference areas or along an effluent concentration gradient. Data collected on the specific effect endpoints listed in Table 1-2 are evaluated to determine if statistical differences in the effect indicators are present (Schedule 5, subparagraph 16(a)(iii)). See Chapter 4 for definitions and other details on benthic invertebrate community endpoints.

Table 1-2: Effect indicators and endpoints for the benthic invertebrate community survey

Effect Indicators	Effect Endpoints
Total benthic invertebrate density	Number of animals per unit area
Evenness index	Simpson's evenness
Taxa richness	Number of taxa
Similarity index	Bray-Curtis index

If the designs in Chapter 4 are not effective or practical, an alternative survey may be appropriate (Chapter 9).

1.3.2.3.3 Fish Tissue Survey

A fish tissue survey (Chapter 3, section 3.11) is conducted to assess if mercury from mining effluent may affect the use of the fisheries resources. A survey respecting the fish tissue is required if, during effluent characterization, the concentration of total mercury in the effluent is equal to or greater than 0.10 µg/L (MMER, Schedule 5, paragraph 9(c)).

1.3.2.4 Biological Monitoring Studies to Investigate Effects

To investigate effects, mines assess the magnitude and geographic extent of all confirmed effects and investigate their causes.

1.3.2.4.1 Magnitude and Geographic Extent

When the results of the two previous biological monitoring studies indicate a similar type of effect (same endpoint, same direction from zero) on the fish population, fish tissue or the benthic invertebrate community, an assessment of the magnitude and geographic extent of the effect is required (MMER, Schedule 5, paragraph 19(1)(d)). Magnitude and geographic extent must be assessed for all confirmed effects. The assessment of the magnitude and geographic extent may require additional monitoring efforts to extend the sampling area further downstream, or the necessary information may already exist as part of previous study results. Guidance on magnitude and geographic extent studies can be found in Chapters 2, 4 and 7.

1.3.2.4.2 Investigation of Cause

If the results of the previous biological monitoring study indicate the magnitude and geographic extent of an effect on the fish population, fish tissue or benthic invertebrate community, an investigation of cause (IOC) study is required (MMER, Schedule 5, subsection 19(2)). The goal of an IOC study is to determine the cause of each confirmed effect. Guidance on IOC studies can be found in Chapter 12.

1.4 Steps in Conducting and Reporting Environmental Effects Monitoring Studies

Conducting and reporting EEM studies, as per the MMER, involves the following key steps:

- Conduct and submit results for effluent characterization, sublethal toxicity testing and water quality monitoring
- Submit study design
- Conduct biological monitoring study
- Conduct data assessment
- Submit interpretative report

1.4.1 Conduct and Submit Results for Effluent Characterization, Sublethal Toxicity Testing and Water Quality Monitoring

Effluent characterization is conducted four times per calendar year and not less than one month apart, with the first characterization carried out not later than six months after the day on which the mine becomes subject to the MMER (Schedule 5, subsection 4(2)). Effluent characterization is conducted on an aliquot of effluent collected for the analysis of deleterious substances under Schedule 4. Refer to Chapter 5 for more information on effluent characterization.

Sublethal toxicity tests are conducted two times each calendar year for three years, and once each year after the third year. Sublethal toxicity tests are conducted on an aliquot of effluent collected for the analysis of deleterious substances under Schedule 4. The first testing is to occur not later than six months after the mine becomes subject to the Regulations (MMER, Schedule 5, subsection 6(1)). Sublethal toxicity testing may be conducted once each calendar year, if the results of six sublethal toxicity tests conducted (after December 31, 1997), on a fish species, an invertebrate species and either an aquatic plant species or an algal species, are submitted to the Authorization Officer not later than six months after the mine becomes subject to the Regulations (MMER, Schedule 5, subsection 6(2)). Refer to Chapter 6 for more information on sublethal toxicity testing.

Water quality monitoring is conducted, starting not later than six months after the day on which the mine becomes subject to the Regulations, four times per calendar year, on samples collected not less than one month apart, while the mine is depositing effluent. Water quality monitoring is also conducted on samples collected at the same time that the biological monitoring studies are conducted (MMER, Schedule 5, subsection 7(2)). Refer to Chapter 5 for more information on water quality monitoring.

An annual report on the effluent and water quality monitoring studies conducted during a calendar year is submitted to an Authorization Officer not later than March 31 of the following year (MMER, Schedule 5, section 8). Most of the annual effluent and water quality monitoring reporting requirements may be met by submitting the data results electronically to Environment Canada using the “Regulatory Information Submission System” (RISS) provided on the following website: <https://www.riss-sitdr.ec.gc.ca/riss/Global/Index.aspx>. For the reporting requirements that are not supported by the RISS, a hard copy submission is required to be submitted to Environment Canada also not later than March 31 of the following year. These requirements include the methodologies used to conduct effluent characterization, sublethal toxicity testing and water quality monitoring, as well as the quality assurance and quality control measures implemented and data related to the implementation of those measures.

1.4.2 Submit Study Design

The study design describes how the biological monitoring study will be conducted to meet the regulatory requirements (MMER, Schedule 5, sections 10 and 19). This guidance document is intended as a starting point for study designs and allows for flexibility in the design of studies to accommodate site-specific needs. Various examples of potential study designs are presented in Chapter 4 (see also Chapters 2, 3, 9 and 12 for information related to study designs). When multiple mines discharge to the same drainage basin, joint EEM studies are encouraged, where practical.

The first study design is submitted not later than 12 months after the day on which the mine becomes subject to the Regulations (MMER, Schedule 5, paragraph 14(a)) or not later than 24 months after the day on which the mine becomes subject to the MMER for mines submitting historical information (MMER, Schedule 5, paragraph 14(b)). The study design for the first, second or subsequent biological monitoring study is submitted to the Authorization Officer at least six months before the biological monitoring study is conducted (MMER, Schedule 5, subsections 15(1) and 19(1)). For mines that have applied to become recognized closed mines, the final study design is submitted not later than 6 months after providing the notice informing of the intention to become a recognized closed mine (MMER, Schedule 5, subsection 23(1)).

A mine could be conducting different types of studies, such as a standard fish survey and a magnitude and geographic extent study for the benthic invertebrate community, at the same time. The study design would then describe how these two studies would be conducted.

The information to be provided in the study design depends on the type of biological monitoring study to be conducted.

1.4.2.1 Study Design for Biological Monitoring Studies to Assess Effects

In cases where effects have not been assessed or confirmed, where the most recent interpretative report indicates the cause of the effect or where the 2 most recent interpretative reports indicate no effects, the designs for biological monitoring studies shall include (MMER, Schedule 5, section 11; guidance in Chapter 2):

- a site characterization that describes effluent mixing in the exposure area and a measure of the effluent concentration at 250 metres from the final discharge point;
- descriptions of the exposure and reference area habitat;
- the type of production process and the environmental protection practices at the mine;
- a summary of any federal, provincial or other laws applicable at the mine regarding effluent and environmental monitoring; and
- a description of any anthropogenic, natural and other factors not related to the effluent that may reasonably be expected to contribute to any observed effect.

Also included is the scientific rationale for selecting the fish species, sampling areas, sample size, sampling periods, and field and laboratory methodologies, as well as the

methodology for determining whether the effluent has an effect on the fish population, fish tissue or benthic invertebrate community. Descriptions of the quality assurance and quality control measures that will be implemented to ensure validity of the data collected must also be included as must the summaries of results from previous biological monitoring studies.

Where available, historical data may provide useful information for site characterization and assist in developing EEM study designs, using lessons learned in previous monitoring. If historical information was submitted, the first study design must include a summary of the results of biological monitoring studies completed before the mine became subject to the Regulations.

1.4.2.2 Study Design for Biological Monitoring to Investigate Effects

If the results of the two previous studies indicate a similar type of effect (same endpoint in same direction from zero relative to reference levels) on the fish population, fish tissue or the benthic invertebrate community, the study design shall include, in addition to the information detailed in section 1.4.2.1, a description of one or more additional sampling areas within the exposure area that shall be used to assess the magnitude and geographic extent of the effect (MMER, Schedule 5, paragraph 19(1)(d)).

If the results of the previous biological monitoring study indicate the magnitude and geographic extent of an effect, the study design shall include a detailed description of the field and laboratory studies that will be used to determine the cause of the effect (MMER, Schedule 5, subsection 19(2)).

1.4.3 Conduct Biological Monitoring Study

The biological monitoring study is conducted according to the submitted study design. If circumstances arise that make it impossible to follow the study design, the owner or operator of the mine must inform the Authorization Officer without delay of the circumstances requiring deviation from the study design and of how the study will be conducted (MMER, Schedule 5, subsections 15(2) and 24(2)). It is recommended that the mine's environmental personnel or consultants also notify the Environment Canada regional EEM coordinator of any deviation from the study design.

1.4.4 Conduct Data Assessment

After completing the fieldwork, data assessment and interpretation are conducted to determine if mine effluent is causing an effect or effects (MMER, Schedule 5, section 16). Data assessment and interpretation also determine future monitoring requirements. The specific analyses conducted to determine if there are effects on fish population, fish tissue or the benthic invertebrate community are described in Chapter 8. Data assessment for mines that have confirmed effects entails determining the magnitude and geographic

extent of the effect(s) and assessing cause(s) of any confirmed effect(s). Guidance on IOC studies can be found in Chapter 12.

1.4.5 Submit Interpretative Report

The first interpretative report is submitted not later than 30 months after the date on which the mine becomes subject to the Regulations or not later than 42 months after the date on which it becomes subject to the Regulations, if the mine has submitted a report utilizing historical biological monitoring information (MMER, Schedule 5, section 18).

Subsequent interpretative reports are submitted 36 or 72 months after the day on which the most recent interpretative report was required to be submitted, depending on the results of the previous interpretative report.

The supporting data from biological monitoring studies are submitted to Environment Canada in the electronic format provided on the EEM website: <http://www.ec.gc.ca/esee-eem/> (see Chapter 10 for further information).

The MMER outline the information to be contained in interpretative reports for biological monitoring studies (MMER, Schedule 5, sections 17, 21 and 25). Chapter 10 describes interpretative reports in more detail. Brief descriptions of the different types of interpretative reports are provided below.

1.4.5.1 Interpretative Report for Biological Monitoring Studies to Assess Effects

Interpretative reports for biological monitoring studies to assess effects include, among other items, results of monitoring studies, raw data, results of data assessments, and identification of any effects.

1.4.5.2 Interpretative Report for Biological Monitoring Studies to Investigate Effects

If the magnitude and geographic extent of a confirmed effect on fish population, fish tissue or benthic invertebrate community is not known, then the interpretative report shall include, among other items, the results of a magnitude and geographic extent study. If the magnitude and geographic extent of the confirmed effect is known but the cause of the effect(s) is not known, the interpretative report shall include a description of the cause of the effect. The IOC interpretative report contains the study results and statement identifying the cause of the effect on fish population, fish tissue and/or benthic invertebrate community. If the cause of the effect(s) was not determined, an explanation of why and a description of any steps that must be taken in the next study to determine the cause shall be included in the interpretative report.

1.5 Recognized Closed Mines

An owner or operator of a mine that has ceased operation, and who intends to have that mine become a recognized closed mine, shall provide written notice of that intention to the Authorization Officer and shall maintain the mine's rate of production at less than 10% of its design-rated capacity for a continuous period of three years starting on the day that the written notice is received by the Authorization Officer. A final biological monitoring study must be conducted during the three-year period (MMER, section 32). The final study design shall be submitted to the Authorization Officer not later than six months after the closure notice is provided (MMER, Schedule 5, section 23). The mine shall base the final monitoring phase on the results of the previous biological monitoring study. The final interpretative report shall be submitted to the Authorization Officer not later than 36 months after the day on which the notice to close the mine was provided (MMER, Schedule 5, section 26). Effluent characterization, sublethal toxicity testing and water quality monitoring requirements continue until the mine becomes a recognized closed mine.

1.6 Identifying a Path through the Metal Mining Environmental Effects Monitoring Program

The metal mining EEM program involves monitoring to assess effects, investigate confirmed effects (magnitude and geographic extent and IOC), and reassess effects. When an effect has been confirmed (i.e., similar type of effect in two consecutive studies), the mine is required to assess the magnitude and geographic extent of the effect (MMER, Schedule 5, paragraph 19(1)(d)), and then to investigate the cause of the effect (MMER, Schedule 5, subsection 19(2)).

1.6.1 Critical Effect Sizes

A critical effect size (CES) is a threshold above which an effect may be indicative of a higher risk to the environment. The Metal Mining EEM Review Team recommended that CESs be established for each of the metal mining EEM effect endpoints following the second national assessment of the EEM data from metal mines (Metal Mining EEM Review Team, 2007).

CESs for the fish population and benthic invertebrate community endpoints were initially developed for the pulp and paper EEM program after EEM data showed that most mills observed an effect in at least one of the effect indicators. Once validated, these CESs were adopted for use in the metal mining EEM program (Table 1-3).

Table 1-3: Critical effect sizes for metal mining environmental effects monitoring program

Fish Effect Endpoints	CES¹	Benthic Effect Endpoints	CES¹
Weight-at-age ²	± 25%	Density	± 2SD
Relative fish gonad size	± 25%	Simpson's Evenness	± 2SD
Relative liver size	± 25%	Taxa Richness	± 2SD
Condition	± 10%	Bray-Curtis Index	+ 2SD
Age ²	± 25%		

¹ Differences in fish population effect endpoints are expressed as percentage (%) of reference mean, while differences in benthic effect endpoints are expressed as multiples of within-reference-area standard deviations (SDs).

² Problems associated with determining the age of some species of fish should be discussed and reviewed before effects on weight-at-age and age are used to choose a path through the EEM program. Refer to Chapter 3 for recommendations on age determination.

1.6.2 Magnitude of Confirmed Effects

The magnitude of each effect observed on fish population, fish tissue or benthic invertebrate community can be further evaluated to determine if the magnitude of a confirmed effect is above or below the CES. Table 1-4 outlines how effects from two consecutive studies are to be grouped to determine if confirmed effects are below or above the CES.

Table 1-4: Evaluation of magnitude of confirmed effects in two consecutive phases

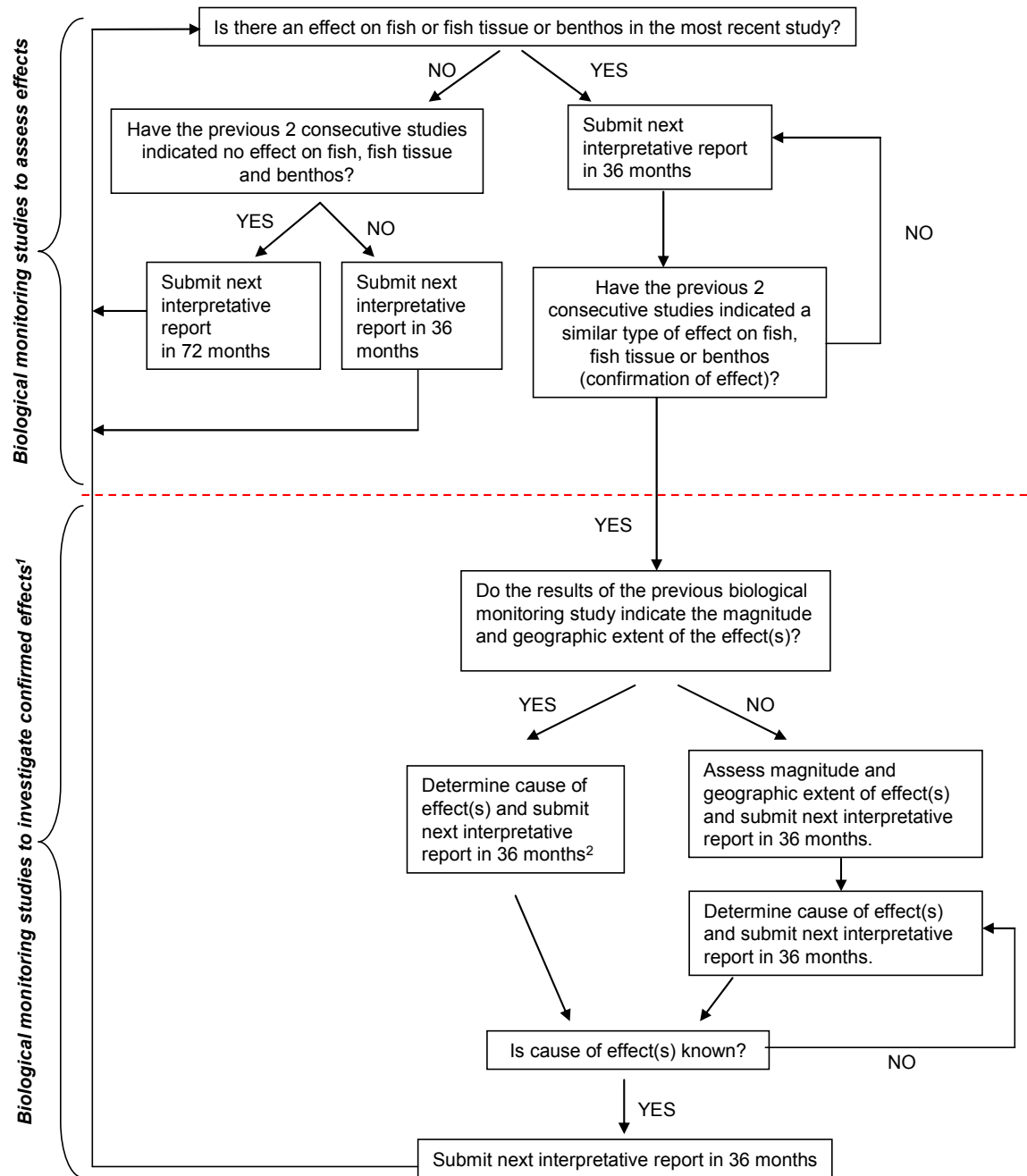
Confirmed effects above or equal to CES	Confirmed effects below CES
Similar effect(s) above or equal to CES observed in 2 consecutive phases	Similar effect(s) below CES observed in 2 consecutive phases
Similar effect(s) in 2 consecutive phases, with the effect(s) above or equal to CES in one phase and below CES in the other phase	Similar effect(s) in 2 consecutive phases, with the effect(s) above or equal to CES in one phase and below CES in the subsequent phase, if there is information available that may explain a change in the observed effects (e.g., improvement of effluent treatment)

1.6.3 Decision Process for the Metal Mining Environmental Effects Monitoring Program

Figure 1-1 is a decision tree to assist mines in identifying an appropriate path through the EEM program, based on their respective situation. CESs are applied to EEM results to assist mines in identifying the level of effort for investigations of confirmed effects. The structure of the decision tree is based on the MMER regulatory requirements, recent scientific knowledge and the experience and knowledge gained through implementing the EEM program.

Site-specific knowledge as well as effluent and water quality data need to be considered before identifying a mine's path through the EEM program. Confirmed effects in supporting endpoints are used as part of the site-specific evaluations to support decisions regarding a path forward (see chapters 3 and 4 for information on supporting endpoints).

Mines are required to continue conducting effluent and water quality monitoring and reporting the results using the timeline prescribed in the MMER and as outlined in section 1.4.1 of this chapter. This requirement is independent of the timeline for conducting biological monitoring studies and submitting interpretative reports.



¹ Level of effort for magnitude and geographic extent and investigation of cause studies is based on magnitude of confirmed effects relative to CES (greater than or equal to CES or below CES).

² If the information on magnitude and geographic extent was not previously reported, it is recommended that it be reported in the investigation of cause study design.

Figure 1-1: Decision tree for the metal mining EEM program

1.6.3.1 Assessing Effects

The interpretative report of the second and any subsequent biological monitoring study is submitted no later than 36 months after the day on which the interpretative report of the previous biological monitoring study was required to be submitted, under the following scenarios:

No effects observed

- The results of a single study indicate no effects on fish population, fish tissue **and** the benthic invertebrate community (MMER, Schedule 5, subsection 22(1)).

Effects observed

- The results of a single study indicate an effect on fish population, fish tissue **or** benthic invertebrate community (MMER, Schedule 5, subsection 22(1)).
- The results of a single study indicate an effect on fish population, fish tissue **and** benthic invertebrate community (MMER, Schedule 5, paragraph 22(2)(a)).

The interpretative report is submitted not later than 72 months after the day on which the interpretative report of the previous study was required to be submitted, under the following scenario:

No effects confirmed

- The results of the previous two consecutive biological monitoring studies indicate no effect on fish population, fish tissue **and** the benthic invertebrate community (MMER, Schedule 5, paragraph 22(2)(b)).

For the purpose of determining the timing of the submission of interpretative reports, if a study respecting the fish population is not required because of the concentration of effluent in the exposure area, as per Schedule 5, paragraph 9(b), then the effluent is considered to have no effect on the fish population. Similarly, if a study respecting fish tissue is not required because of the concentrations of mercury in the effluent as per Schedule 5, paragraph 9(c), then the effluent is considered to have no effect on fish tissue.

1.6.3.2 Investigating Confirmed Effects

If the results of the previous two consecutive biological monitoring studies indicate a similar type of effect (same endpoint in same direction from zero relative to reference levels) on fish population, fish tissue or the benthic invertebrate community, and if the magnitude or geographic extent of the effect or cause of the effect is not known, then the interpretative report is submitted not later than 36 months after the day on which the interpretative report of the previous study was required to be submitted (MMER, Schedule 5, paragraph 22(2)(c)).

1.6.3.2.1 Level of Effort for Investigating Effects

Mines are required to investigate all confirmed effects. The following paragraphs provide recommendations as to how the confirmed effects can be investigated depending on the magnitude of the effects (below or above CESs).

Confirmed effects of magnitude greater than or equal to CESs

Mines with confirmed effects of a magnitude greater than or equal to CESs (Table 1-4) would conduct a field study to assess magnitude and geographic extent of the effects and submit the next interpretative report in 36 months. Subsequently, the mine would conduct field and/or laboratory studies to determine the cause(s) of the effects and submit the IOC interpretative report in another 36 months. If the magnitude and geographic extent of the effect has already been determined, the mine may move directly to determining the cause(s) of the effects. In this case, the mine could report the magnitude and geographic extent of the effects in the IOC study design.

Confirmed effects of a magnitude below the CESs

If a confirmed effect has a magnitude below the CES, it is not expected that larger effects be observed farther away from the final discharge point. The mine could therefore assess the magnitude and geographic extent of a confirmed effect below the CES by providing a scientifically sound rationale using the results and other existing information from studies, and then move directly to determining the cause(s) of the effects. Under this scenario, if the mine uses existing information to assess the magnitude and geographic extent of the effects, it is recommended that this information be reported in the IOC study design and that the next interpretative report be submitted in 36 months. The cause(s) of the effect could be determined by conducting field and/or laboratory studies or by examination and presentation of solid evidence using existing data, alone or in combination with new data and/or a literature review.

Once the cause(s) of the effects has (have) been determined, the next interpretative report is submitted 36 months after the day on which the most recent interpretative report was required to be submitted. In this case, the study design must describe biological monitoring studies to assess effects (see section 1.4.2.1).

1.6.3.3 Timing of Studies to Assess Effects and Magnitude and Geographic Extent and to Investigate Cause

There are different stages in assessing and investigating effects (Figure 1-1). In many cases, the process of assessing and investigating effects may not move together (concurrently) for the different components (fish population, fish tissue and benthic invertebrate community). Once an effect has been confirmed, mines are required to assess the magnitude and geographic extent of the effect and determine the cause of the effect. The magnitude and geographic extent and IOC studies are required for all confirmed effects.

While conducting a study to assess the magnitude and geographic extent of an effect observed in one component (fish population, fish tissue, and benthic invertebrate community), mines are also required to continue monitoring the component(s) where effects were not previously observed or confirmed. Therefore a mine may conduct a study to confirm an effect in one component, or lack thereof, while conducting a study to determine the magnitude and geographic extent of an effect in another component.

While conducting an IOC study for an effect confirmed in one component (fish, fish tissue, benthic invertebrates), mines do not need to conduct simultaneous studies on component(s) where effects were not confirmed. If effects were confirmed in more than one component or for more than one endpoint within a component, then all cause(s) of all confirmed effects need to be determined during the next phase or, if not possible, promptly over subsequent phases.

1.7 References

Metal Mining Environmental Effects Monitoring Review Team. 2007. Metal Mining Environmental Effects Monitoring Review Team Report. National EEM Office, Environment Canada, Gatineau, QC.

<http://www.ec.gc.ca/Publications/default.asp?lang=En&xml=933E74AE-6C80-4A54-9CAB-62E5A7440114>

Table of Contents

2. Study Design, Site Characterization and General Quality Assurance / Quality Control	2-1
2.1 Overview	2-1
2.2 Study Design and Site Characterization	2-1
2.2.1 Site Characterization	2-2
2.2.1.1 Plume Delineation	2-4
2.2.1.2 Measuring of Constituents in the Effluent	2-6
2.2.1.3 Habitat Mapping and Classification	2-7
2.2.1.4 Aquatic Resource Inventory	2-8
2.2.1.5 Classification Scheme for Reference Area Selection	2-9
2.2.1.6 Framework for Rivers	2-13
2.2.1.7 Framework for Lakes	2-14
2.2.1.8 Use of Sublethal Toxicity for Site Characterization	2-16
2.2.1.9 Characteristics of Mining Environments	2-16
2.2.2 Exposure and Reference Areas	2-18
2.2.2.1 Selection of Final Discharge Point for Monitoring	2-18
2.2.2.2 Selection of Exposure and Reference Areas	2-18
2.2.2.2.1 Exposure Area	2-19
2.2.2.2.2 Reference Area	2-19
2.2.3 Reporting of Field Station Positions	2-22
2.2.4 Modifying or Confounding Factors	2-22
2.2.5 Tributaries and Other Point- and Nonpoint-Source Discharges	2-23
2.2.6 Natural Variation in Environmental or Habitat Conditions	2-24
2.2.7 Historical Damage	2-24
2.3 General Quality Assurance / Quality Control and Standard Operating Procedures	2-24
2.3.1 Quality Assurance and Quality Control	2-24
2.3.2 Standard Operating Procedures	2-26
2.4 References	2-27

Table

Table 2-1: Site characterization information for preparing an EEM study design.....	2-4

2. Study Design, Site Characterization and General Quality Assurance / Quality Control

2.1 Overview

This chapter includes information on study design, site characterization, and general quality assurance / quality control (QA/QC) information for the metal mining environmental effects monitoring (EEM) program. The requirements for the study design and site characterization are listed in the *Metal Mining Effluent Regulations* (MMER) (Schedule 5, sections (s.) 10–14) and Chapter 1. This includes information such as timelines for EEM studies (first studies, studies aim to confirm absence or presence of effect, magnitude and geographic extent, investigation of cause and final biological monitoring studies prior to a mine closing), content of study-design reports and submission dates. Each chapter of this document contains additional information on recommended methodologies for the study design for fish, fish tissue, benthic invertebrates and alternative method studies. In addition, each chapter provides more detailed information on QA/QC.

2.2 Study Design and Site Characterization

The objective of a study design is to describe how the biological monitoring studies (a fish survey, fish tissue analysis and benthic invertebrate community survey) are to be conducted.

Study designs should describe the following (MMER, Schedule 5, s. 10–14):

- a summary of previous biological monitoring and effluent and water quality monitoring;
- information related to site characterization, including the results of plume delineation studies;
- the objectives of the field monitoring program, including overall approach and rationale for biological monitoring, which may be based on previous monitoring results;
- statistical design criteria, hypotheses, statistical methods and data needs;
- a description of how the biological monitoring studies are to be conducted to determine if there are effects, taking confounding influences into consideration;
- field sampling plans, including what will be measured, where and when it will be measured, location of exposure and reference sites, and rationale for selection of final discharge point;
- QA/QC measures that will be taken to ensure validity of data; and
- schedules for field monitoring and submission of the interpretative report.

2.2.1 Site Characterization

Site characterization information is submitted as part of the EEM study design (MMER Schedule 5, s. 10 (a)). The requirements for site characterization are described in Schedule 5, s. 11 of the MMER. Table 2-1 summarizes site characterization information that should be included in the study design. For the first EEM study design, site characterization is included in detail. For subsequent EEM studies the site characterization can be submitted in summary format, but new information (e.g., production rates) should be updated in detail. In most cases, mines will have most site characterization information available from previous assessments and historical studies. If information critical to the design of the EEM study is not available, additional field data may be required to provide adequate background for the first EEM study design, particularly with respect to hydrology and aquatic resources.

Site characterization information is used to identify suitable sampling areas that have similar habitats in the exposure and reference areas, and to obtain information on other discharges and confounding factors that may affect the interpretation of data obtained from those areas. Information on some of the unique environmental characteristics of mine sites that should be taken into consideration during the site characterization can be found in Section 2.2.1.9.

For mines with insufficient historical information to locate reference and exposure areas, exploratory sampling may be useful. Exploratory sampling can also be used to identify habitat characteristics for effective selection of sampling stations.

An experienced field crew should be able to approximate the effluent field based on field measurements of water quality tracers (e.g., specific conductance) or preliminary dye study results, and can often identify likely depositional areas based on observed receiving water flow and circulation patterns. Thus, it is usually possible to choose some appropriate water and sediment sampling stations in the field and to complete exploratory sampling of the receiving environment concurrent with plume and depositional zone studies and critical resource/habitat inventories in a single campaign.

Much of the site characterization information can be effectively reported in map form. Maps should be of sufficient scale (e.g., 1:5000) to show the features of the study area in adequate detail. The actual scale should be reported on any map used. The geographic extent of the study area to be mapped should be determined on a site-specific basis, and should include the discharge point as well as the exposure and reference areas.

The requirements of the site characterization section of the EEM study design are outlined in the MMER. Additional information relevant to the site characterization that should be reported may include the following (including Table 2-1):

- a) An identification of the major chemical reagents used in the overall production process since January 1, 1996. Mines are encouraged to report current quantities of

reagents used. This should list reagents of the following types:

- activators
- flocculants
- pH modifiers
- depressants
- frothers
- collectors

- b) An inventory of all discharge points where effluent is deposited into the exposure area. This inventory should identify all known sources of effluent to the aquatic environment, including those regulated under the MMER, as well as any others (e.g., nonpoint sources) that may have the potential to cause an effect on the aquatic environment.
- c) Information on the local climate, particularly seasonal precipitation patterns.

Table 2-1: Site characterization information for preparing an EEM study design

Information Type	Recommended information to be reported (where possible, some of the information can effectively be reported in map form)
General characteristics	• bedrock and surficial geology • topography • soil and vegetation • site accessibility • climatology
Hydrology	• watershed(s) description • water flow (rivers) or dispersion (lakes, estuaries, marine) characteristics • general description of how effluent(s) mix(es) with receiving water • bathymetry mapping (including slope in marine environments) • gradient (rivers) • tides (marine) - mean monthly tide height data • stratification patterns (thermal and chemical) • natural barriers to fish movement • effluent plume delineation
Anthropogenic influences	• docks, wharves, ferry terminals, marinas, boat launches, and public recreational zones • bridges, crossings and fordings • water intakes, effluent discharges, storm water discharges, sewer overflows • waste disposal sites • contaminant source inventory, including point- and nonpoint sources • dams, culverts, waterfalls and other barriers to fish movement • surrounding land use • location of aquaculture facilities
Aquatic resource characteristics	• location of exposure and reference areas used in historical studies • fish and shellfish species present (resident and migratory) • relative abundance of fish and shellfish species • use of the exposure and reference areas by fish and shellfish (spawning grounds, nursery areas, etc.) • rare, threatened or endangered fish species (if present) • non-commercial fisheries (recreational and subsistence) • commercial fisheries • zones of macrophyte growth • ecologically relevant benthic invertebrate habitat(s) and their relative proportions, including: • delineation of depositional and erosional zones • substrate classification
Environmental protection systems and practices	• water management • effluent treatment • residence time • management of tailings • polishing ponds • waste rock (including use of waste rock for mine back fill and building material)

2.2.1.1 Plume Delineation

A description of the manner in which the effluent mixes within the exposure area,

including an estimate of the concentration of effluent in water at 250 m from each final discharge point (MMER, Schedule 5, s. 11 (a)), is to be described in the site characterization. For subsequent biological monitoring studies the site characterization information may be summarized along with, where applicable, a detailed description of any changes to that information since the submission of the most recent study design (MMER, Schedule 5, s. 19 (1) (a)). This description should include an indication of relative flow of the effluent and receiver, as well as seasonal variations in flow. This will give an indication of dilution rate. The description should also give an indication of the density of the effluent, and where within the water column the effluent is likely to be, prior to complete mixing. This estimate may be based on direct measurements in the field or modelling, but it is recommended that modelling be validated with field measurements.

A fish population study is conducted if the concentration of effluent is greater than 1% in the area located within 250 m of a final discharge point (MMER, Schedule 5, s. 9 (b)). If such a study does not need to be conducted due to the effluent concentration being less than 1%, it is recommended that more rigorous plume delineation methods be used to document the effluent concentrations in the exposure area.

It is recommended that the description of the manner in which effluent mixes within the exposure area include the following:

- identification of where in the exposure area the effluent is located, prior to mixing with the receiving water;
- estimation of where in the exposure area the effluent and receiving water begin mixing, and where mixing is complete;
- estimation of the effluent dilution ratio at points downstream of effluent discharge; and
- identification of significant sources of dilution, other than the primary receiver (i.e., tributaries or other streams); and
- how the above vary with the tides and seasons.

For extensive guidance on plume delineation, please consult the *Revised Technical Guidance on How to Conduct Effluent Plume Delineation Studies*, available from Environment Canada (2003) at www.ec.gc.ca/eseec-em/D450E00E-61E4-4219-B27F-88B4117D19DC/PlumeDelineationEn.pdf. This document was prepared for the pulp and paper EEM program but can also be applied to the metal mining EEM program. Additional plume delineation information pertaining to metal mining is discussed below.

2.2.1.2 Measuring of Constituents in the Effluent

Conductivity Surveys

If the conductivity of effluent is consistent during the period of study, a conductivity survey can be used to locate the effluent plume within the exposure area. Survey results can be assessed semi-quantitatively, or conductivity measurements can be converted into relative effluent concentrations lying between 1 (effluent) and zero (background) by applying the following formula:

$$C_r = (C_a - C_b)/(C_e - C_b)$$

Where: C_e = effluent conductivity, $\mu\text{S}/\text{cm}$
 C_b = background conductivity, $\mu\text{S}/\text{cm}$
 C_a = measured conductivity to be converted, $\mu\text{S}/\text{cm}$
 C_r = relative concentration

The relative concentration is an expression of the dilution ratio. Temperature readings need to be taken along with the measurements obtained from the conductivity meter, since conductivity rises approximately 2% for every 1°C rise in temperature. Further information on theoretical considerations for effluent plume delineation using conductivity can be found in relevant reference material (e.g., Fischer et al. 1979; Freeze and Cherry 1979).

Although conductivity surveys can provide valid and cost-effective estimates of the location of effluent within the receiving environment, the natural variability of conductivity in surface waters can interfere with locating the edge of the effluent plume. Such natural variability in conductivity may be observed with depth as well as with surface measurements. The presence of multiple tributaries or receiving water bodies can further exacerbate this difficulty.

Tracing by Effluent Metals

The location of the effluent plume within the exposure area may also be estimated by choosing a reference parameter present in the effluent and tracking its fate in time and space by measuring its concentrations in water samples taken at specific locations. The selection of such a tracer needs to be based on its stability and consistency of concentration, as well as on representativeness and ease of measurement. Since the parameter selected has to be a conservative substance, metals such as copper or nickel may serve as a “natural” tracer. Sulphate is often a good tracer of base metal mine effluents, particularly in massive sulphide deposits.

It should be cautioned that the effluent parameter selected for tracing may sometimes be present at the same order of magnitude in receiving waters. This would rule out its application for determining effluent dilution. Other parameters may be specific to the effluent alone, but present in such low concentrations that it makes their detection

difficult. Several parameters may be present at significantly greater levels in effluent, which would make them ideal tracers for dilution measurement. However, as a result of complications such as the costs of analysis, instability of the substance, difficulty in measurement, or the lack of an adequate in situ measuring apparatus, these parameters may not always prove to be appropriate “natural” tracers. The potential for effluent metals to be used as tracers for plume delineation should therefore be evaluated on a case-specific basis.

2.2.1.3 Habitat Mapping and Classification

Some elements of habitat mapping and classification, as well as aquatic resource inventory, are included as part of site characterization. More detailed habitat mapping may be helpful in identifying habitat types present in the exposure and reference areas. This section provides guidance on habitat mapping and classification.

The recommended method to create a habitat map is to perform a habitat classification. The recommended framework for classifying aquatic features is the classification system developed by the U.S. Fish and Wildlife Service, *Classification of Wetlands and Deepwater Habitats of the United States* (Cowardin et al. 1979; Busch and Sly 1992). This system allows for classification of a wide range of continental, aquatic and semi-aquatic habitats. Cowardin et al. (1979) also provide guidance on habitat description for coastal and estuarine situations.

Classification systems for marine shorelines to deep coastal areas include Frith et al. (1993), Booth et al. (1996), Robinson and Levings (1995), Hay et al. (1996) and Robinson et al. (1996). Specifically, estuarine classification has been reviewed by Matthews (1993), Scott and Jones (1995), Finlayson and van der Valk (1995) and Levings and Thom (1994). In the United States, the most widely used system is that of Cowardin et al. (1979) and Cowardin and Golet (1995), with expansions proposed by other authors.

Listed below are examples of environment-specific conditions for various habitats:

Rivers: It is recommended that river habitat descriptions include information on elevation gradient; the location of dams, falls and other barriers to fish migration; mean annual discharge and ranges; and general substrate characteristics of each river (preferably in the form of a gradient profile chart). Upstream and downstream inputs (e.g., storm water, sewer overflow, effluent from other industrial sites) should be mapped and described.

Lakes: Important habitat features of lakes include bathymetry, the locations of major inlets and outlets, and general oxygen-temperature conditions (e.g., thermal stratification, occurrences of oxygen depletion in deep water).

Open coastlines: Suggested additional mapping parameters for open coastlines (marine, Great Lakes) include depth contours, nearshore substrate characteristics, shoreline configuration, and the locations of inflowing rivers and other discharges and activities.

Estuaries: Estuaries are best described in terms of their general salinity gradients, flows, bathymetries and general substrate features. A description of tidal cycles is recommended for all marine and estuary locations. Most of the above features can be described from navigational maps, topographic maps, government publications on tides and river discharge records, and through interviews with local government officials and knowledgeable individuals.

Natural wetlands: A wetland is defined as land that is saturated with water long enough to promote wetland or aquatic processes as indicated by poorly drained soils, hydrophytic vegetation and various kinds of biological activities that are adapted to a wet environment (Metal Mining EEM Review Team 2007). Wetlands include bogs, fens, marshes, swamps and shallow waters (usually two metres deep or less) (Metal Mining EEM Review Team 2007). During the Metal Mining Program Review (2007) the review team recommended that natural wetlands for EEM studies should be avoided. Where a mine final effluent flows into a natural wetland area, EEM studies should be conducted downstream of the wetland when studies upstream are not possible. This recommendation is consistent with the Federal Policy on Wetland Conservation. This policy is found at the website: http://www.atl.ec.gc.ca/wildlife/wetlands_policy.html

It is recommended that bottom substrates be described. Further guidance on aquatic habitat assessment can also be found in the Department of Fisheries and Oceans and the British Columbia Ministry of the Environment and Parks (1987), Orth (1989), Ontario Ministry of Natural Resources (1989), Plafkin et al. (1989), and the Department of Fisheries and Oceans (1990).

Depositional zones in the exposure area should be identified and illustrated on the habitat map. Any information on sediment characterization (chemistry, toxicity) should be reported. Depositional zones occur where water velocity decreases, resulting in particles settling out; the finest particles settle out in the slowest current speeds. Historical contaminant or benthic invertebrate community data may be helpful in identifying sampling stations within a depositional exposure area. To compare resident benthic invertebrate communities, similar (but uncontaminated) sediment depositional zones should be located in the reference area. In situations where historical contamination was from a source other than the mine, two reference areas could be used; one with and one without the historically contaminated sediment.

2.2.1.4 Aquatic Resource Inventory

An aquatic resources inventory includes the identification of fish and shellfish (resident and transient) that are presently being fished commercially and non-commercially (both

sport [including stocked fish] and subsistence fishery). The inventory should make particular note of fish species that may be present in sufficient numbers to be considered as a sentinel species, and of utilization (e.g., spawning, nursery) of the exposure area by fish species. In addition, any species recognized by federal, provincial or territorial authorities as rare, threatened or endangered should also be included. The Committee on the Status of Endangered Wildlife in Canada website (www.cosewic.gc.ca); district fisheries biologists in federal, provincial or territorial regulatory or museum agencies; local conservation officials; and members of the local community (fishermen, Aboriginal people and public interest groups) are all sources for this type of information.

The potential success of field programs increases with familiarity of the study area. It is recommended that fieldwork be undertaken to verify historical information if this information is not detailed or recent.

Stocked fish are not appropriate for EEM-type monitoring, as these fish are predominately sport fish and are not appropriate indicator species since their growth and reproduction may be altered depending on how and when they were stocked and raised. As well, stocked fish generally have no apparent reproductive success; therefore, this effect indicator could not be evaluated.

2.2.1.5 Classification Scheme for Reference Area Selection

Because reference areas will vary among different landscapes, approaches have been developed to classify land through which rivers run or in which lakes reside in order to predict aquatic biotic assemblages (Corkum 1989, 1992; Hughes 1995, Maxwell et al. 1995; Omernik 1995). A classification system is a way of simplifying sampling procedures and management strategies by organizing a variable landscape (Conquest et al. 1994). The assumption is that the classification scheme is hierarchical. The advantage of a hierarchical classification scheme is that it “offers a way to discriminate among features of the landscape at several scales of resolution” (Conquest et al. 1994). The classification scheme is based (with modifications) on one developed by the U.S. Department of Agriculture’s Forest Service (Maxwell et al. 1995). The hierarchical classification scheme is presented as a guide in the a priori selection of sampling areas.

Habitat-Specific Allocation of Reference and Exposure Areas

The following specific points should be considered during the selection of reference and exposure areas and/or stations:

For Rivers:

- The size of the drainage basin selected is based on stream order. For example, if a mine site is located on a second-order stream, the drainage basin area is delineated at the point the stream becomes third-order (i.e., at the junction of two second-order streams).

- If there are no upstream inputs or confounding factors, reference area(s) can be within the drainage basin and upstream of the mine.
- If confounding factors, such as nonpoint- or point-source inputs, occur upstream of the effluent, the reference area(s) can be selected in nearby drainage basins with comparable habitat features (Figure 4-4).
- If physical disturbance of the river valley is associated with the mine, effluent effects may be confounded by the disturbance. Accordingly, reference areas should be selected to match the physical disturbance, if possible.
- The following features should be similar between reference and exposure areas:
 - ecoregion, drainage basin area, stream order, bankfull width, channel gradient, channel pattern, habitat types, water depth, water velocity, substratum composition, riparian vegetation, shoreline structure and land use, etc.

For Lakes:

- In lakes with a single-mine effluent and without nonpoint sources of pollution, the sphere of influence originating from the effluent should be determined. This is particularly important for lakes in which effluent flow is not unidirectional.
- If effluent plume delineation and former studies indicate that mine effects are likely to be local and restricted, select reference areas within the lake in which the mine discharge occurs. These reference areas should occur in separate but comparable bays or basins of the lake.
- If effluent plume delineation indicates that the identified effluent is dispersed throughout the lake, select reference area(s) in the nearest comparable lake within the same or adjacent drainage basin.
- If nonpoint- or other point-source inputs occur elsewhere on a lake, select reference area(s) in the nearest comparable lakes within the same or adjacent drainage basin.
- If the mine effluent is associated with physical disturbance in the area, effluent effects may be confounded by the disturbance. Accordingly, physically matched reference areas should be selected, if possible.
- The following features should be similar between reference and exposure areas:
 - ecoregion, geological origin, drainage basin area, morphometry, slope from shoreline, habitat types, substratum composition, riparian vegetation, shoreline structure and land use, etc.

For Marine Environments:

- The reference area should be within the same water body and hydrographic current or tidal regime as the exposure area. In other words, the closer the reference area is to the exposure area, the better. Benthic invertebrate communities in marine ecosystems are considerably higher in species richness, have more complex trophic relationships, faunal size ranges and reproductive strategies, than benthic invertebrate communities in freshwater

ecosystems. Because of this complexity, and the multitude of interactions between species in marine benthic invertebrates, small shifts in physical or chemical conditions can dramatically alter the overall benthic faunal community. Add to this the effect of increasing variation in chance larval settlement with increasing geographic distance (geographic “drift” in community structure) and physical barriers in complex coastlines, and it is very rare to find similar invertebrate communities from one bay or fjord to the next, and very difficult to predict specific benthic community structure based on sediment factors (for recent review on marine invertebrate sediment interactions, see Snelgrove and Butman 1994)). In order to have some confidence that the “natural” benthic community is similar enough from one coastal area to the next, there should be sufficient water exchange between them. This is more likely in open coastal areas than in isolated bays and fjords.

- Reference areas that are not in the same water body or hydrographic regime may only be suitable for comparisons of summary characters such as shifts in abundance or species richness. If the habitat conditions are similar enough to the exposure area, it may also be possible to compare larger-scale biotic factors such as the presence of characteristic, long-lived depth/substrate specific taxa described by Thorson (1957) as “parallel communities”.
- Reference and exposure areas should have very similar habitat type, shoreline structure (steep, mountainous, delta, marsh, etc.), bottom topography (sills, sandbars, exposure to open oceanic influences, etc.), substrate type (particle size, sorting, natural chemistry), depth properties, current regimes, physical water properties, nutrient regimes, confounding inputs and drainage characteristics.
- Some special considerations are important for determining the suitability of reference areas for marine and estuarine mines. Physical factors in the estuarine/marine environment that tend to be more complex than in freshwater are salinity (including seasonal freshwater influence), tides (and tidal currents) and sediment sulphides. Other important physical factors include ice-scour or buildup, freezing, water column stagnation due to large summer freshwater runoff, re-suspension due to surface freezing in winter, dams or log-booms, extraordinary siltation or clogging from logging, and periodic flooding.
- In addition to the above important characteristics, the following specific points should be similar between reference and exposure areas:
 - ⇒ intertidal areas: shoreline slope, wave exposure, light and tidal exposures, shoreline vegetation, and encrusting fauna (although the latter may be part of the benthic taxa being monitored for a response to mine effluent).
 - ⇒ sub-tidal areas: seasonal water column stability and bottom oxygen depletion (stagnation).

Ecoregions

The first step in reference site selection is to use terrestrial attributes (ecoregions) with similar features. Ecoregions are defined as “part of an ecoprovince characterized by distinctive ecological responses to climate as expressed by vegetation, soils, water, and fauna” (Wiken 1986; Wickware and Rubec 1989). Ecoregion maps for Canada are available at http://sis.agr.gc.ca/cansis/nsdb/ecostrat/printed_maps.html

Drainage Basin and Geographic Scale

Catchments or drainage basins have clear hydrographic boundaries. A drainage basin is defined as the area that has a common outlet for its surface runoff. Although inter-basin transfer occurs among biota, the geoclimatic history of large basins (1:2 000 000) are known to create barriers to dispersal through hydrographic divides and climate (Maxwell et al. 1995). It is essential to establish the geographic scale appropriate to the study design. In large-scale, synoptic surveys in which relationships are sought between landscape features and aquatic biota, the mapping scale for drainage basins is 1:250 000 (Corkum 1989, 1992, 1996; Reynoldson and Rosenberg 1996). These basins are subdivided into progressively smaller sub-basins.

Land/water interactions with respect to sediment and nutrient transport off the land and from upstream sources is integral in developing predictive models that link environmental variables and associated biota. Drainage basins may occur within ecoregions or may cross different ecoregions. Aquatic fauna are more similar to one another in drainages that occupy the same ecoregion than in drainages that occupy different ecoregions (Corkum 1992; Hughes et al. 1994).

Land Use and Vegetative Buffer Strips

Although ecoregions are defined in terms of climate and natural vegetation, natural vegetation is disturbed with human development. Land-use type is a simple measure of disturbance within the drainage basin. If there is a change in land use (e.g., land clearing for agricultural uses or logging, or fire), the biological assemblages in receiving waters will respond to those changes (Corkum 1992, 1996). Accordingly, site selection should be in drainages with comparable land use.

The degree (width and type) of a vegetated buffer strip adjacent to rivers and lakes should be recorded at all sampling areas. In reference areas where human disturbance cannot be avoided, the effect of a vegetated buffer strip moderates temperature fluctuations through shading (Budd et al. 1987), removes or reduces sediment from runoff (Young et al. 1980), and regulates nutrients and metals entering the water body (Peterjohn and Correll 1984).

2.2.1.6 Framework for Rivers

The river sampling design provides a framework for characterizing habitats at multiple scales (Meador et al. 1993). The framework for rivers is based on how they are organized in hierarchical space and how they change through time (Frissell et al. 1986). The riverine system has several hierarchical or nested levels: drainage or catchment basin, valley segment, stream reach and channel unit (Conquest et al. 1994).

Valley Segment and Stream Order

Valley segments are distinctive sections of drainage basins that possess geomorphic properties and hydrological transport characteristics that distinguish them from other segments (Cupp 1989). Montgomery and Buffington (1993) identified three valley segment types: colluvial (channelized and unchannelized), alluvial and bedrock. Valley segments can be filled with colluvium (sediment and organic matter from landslides) or alluvium (sediment transported by flow). The third valley segment has little soil and is dominated by bedrock.

Valley segments are distinguished by six criteria (Conquest et al. 1994):

1. Stream order (position in drainage network)
2. Valley slope gradient
3. Ratio of valley bottom width to active channel width
4. Channel gradient
5. Stream-corridor geomorphic surface deposits
6. Channel pattern.

Channel segments are assigned stream orders (Strahler 1957) for a particular map scale or aerial photograph (e.g., 1:250 000) (Newbury and Gaboury 1993).

Stream Reach

Stream reaches consist of homogeneous associations of topographic features and channel geomorphic units (Bisson and Montgomery 1996). Stream reaches can be used to predict local stream response to perturbations (Montgomery and Buffington 1993). Stream reaches are useful in assessing habitat quality, aquatic productivity, fish distributions and stream health (Maxwell et al. 1995). Stream reach classification is determined using map scales of 1:12 000 to 1:24 000. Criteria used to classify stream reaches include:

- channel pattern
- channel entrenchment
- channel width
- hydraulic radius
- basin area

- channel material
- stream gradient
- bed form
- riparian vegetation

Simpler approaches have been adopted to identify stream reaches. For example, a straight channel has an undulating bed with alternating riffles and pools spaced at repeating intervals of 5-7 channel widths (Leopold et al. 1964; Leopold 1994). Newbury (1984) defined a stream reach to be equivalent to six times the channel width.

Channel Unit

Channel units are subdivisions of stream reaches that describe uniform microhabitats (depth and flow) and are used to identify factors that limit both invertebrate and fish populations within a stream reach. Hawkins et al. (1993) proposed a three-tiered system of channel units in which the first level distinguishes riffles from pools. The second level identifies turbulent and non-turbulent riffles and distinguishes between pools formed by scour or dams. Dammed pools retain more sediment and organic debris and have more cover than scour pools. The third subdivision identifies microhabitats based on hydraulic processes and structure. Channel units are typically 10 m or less and typically cannot be mapped at a scale appropriate for land management.

Criteria for subdivision of riffles include:

- gradient or water surface profile
- percentage of super-critical flow
- bed roughness
- mean velocity
- step development

Criteria for subdivision of pools include:

- location (main channel or off-channel)
- longitudinal and cross-sectional depth profiles
- substrate characteristics
- pool-forming constraints

2.2.1.7 Framework for Lakes

The geological origin, hydrology and morphometry (obtained from maps and aerial photographs) of lakes are important in identifying sediment-water interactions and productivity of lakes (Wetzel 1975). Although thermal stratification can be predicted from morphological features, field verification is necessary. The mapping scale for lakes is typically 1:24 000 or 1:63 000 (Maxwell et al. 1995).

Origins, Location and Hydrological Linkages

Reference and exposure lakes should have the same origin, location and hydrological linkages. Lake geology ultimately affects the physical, chemical and biological characteristics of water bodies. For example, Hutchinson (1957) identifies 11 types of geomorphic processes (tectonic, volcanic, landslides, glacial activity, solution, fluvial, wind, shorelines, organic accumulation, anthropogenic and natural dams, meteorite impact). Surface geology and location (altitude, latitude and longitude) affect lake chemistry and thermal regimes (Winter 1977). These variables, which can be obtained from maps, are used to predict the biological composition and productivity of lakes (Dolman 1990; Winter and Woo 1990). Hydrological linkage refers to the “connection of a lake to surface or ground water” and can forecast information about lake biota (Maxwell et al. 1995). Maxwell et al. (1995) describe three types of hydrological linkages: riverine linkage (outlets and/or inlets or unconnected), groundwater linkage (gaining, losing, neutral or no recharge), and water storage regime (perennial or intermittent).

Morphometry

Lake morphometry has been used historically to predict fish yields (Ryder 1965; Kerr and Ryder 1988) and to determine species diversity (Eadie and Keast 1984; Marshall and Ryan 1987). With the exception of depth (and volume), other features can be obtained from maps. Hypsographic (cumulative depth-area or cumulative depth-volume) graphs are useful for comparing basin shapes of lakes and predicting surface area or volume for water-level control of reservoirs. Common morphological features of lakes include surface area, volume, mean and maximum depth, shoreline development, and hydraulic residence time.

Trophic State

Many lake classification systems are based on a measure of productivity (oligotrophy, mesotrophy, eutrophy). A fourth lake type (dystrophy) is used to describe lakes that receive large amounts of organic matter from external sources; these lakes are heavily stained and are known as brown-water lakes. Productivity of dystrophic lakes is low and so some limnologists group dystrophic lakes as a subclass of oligotrophic lakes. The following variables have been used to describe the trophic status of lakes:

- dissolved oxygen
- thermal mixing (lake stratification)
- total phosphorus
- soluble reactive phosphorus
- total nitrogen
- nitrite + nitrate

- ammonium
- chlorophyll *a*
- transparency
- organic matter

Zone

Lakes are subdivided into an open-water pelagic zone, a shoreline or littoral zone inhabited by autotrophic vegetation, and a deeper benthic region free of vegetation (the profundal zone). The reference and exposure areas should always be located in the same zone.

2.2.1.8 Use of Sublethal Toxicity for Site Characterization

Since historical sublethal toxicity data for some or all of the required tests (see Chapter 6) have been generated for effluents discharged from a number of mine sites across Canada, the operator may want to use this information during site characterization in the following ways:

- 1) To aid in determining sampling areas for fish or the benthic invertebrate community. If the operator has no historical field data on fish, fish food sources or fish habitat from their effluent exposure zone, historical effluent sublethal toxicity data (if available in sufficient quality) could be used to estimate the potential zone of influence to help in establishing sample collection locations for the fish and benthic invertebrate community surveys in the first biological monitoring (i.e., estimation of extent of response in the high effluent exposure area). Details on how to estimate the geographic extent of a sublethal toxicity response are provided in Chapter 6.
- 2) As an aid in comparing effluent discharge sources. If there are numerous effluent discharge locations at a mine site, one of the recommended sublethal toxicity tests could be used to quantify the degree of sublethal toxicity contributed from the different effluent discharge sources (see chapter 6).

2.2.1.9 Characteristics of Mining Environments

Many mines and mining activities share some similarity in environmental features. They are briefly outlined below:

Headwater location of mine sites: Many mine sites are located near the headwaters of rivers or streams. In some instances the effluent can constitute a significant portion of the flow downstream of the discharge point. This will have an impact on how to characterize the mine's exposure area. Headwater streams, because of their size, gradient or intermittent flow, are often not suitable fish habitat. Therefore, the mine effluent is frequently discharged to receiving waters with small or no fish populations, although the

diluted effluent will, in most cases, eventually reach fish habitat. Nonetheless, some fish species often use accessible headwater areas at some stage of their life cycle, such as for spawning, and this information should be considered when designing an EEM study. Mines may have to move progressively downstream until they reach an area with suitable number and species of fish; however mines should assess fish populations in the immediate receiving environment first.

Effluent quality and volume: Effluent quality and quantity are influenced by several factors including the nature of the ore and host rock, processing methods, effluent treatment methods, climate, and site hydrology. Discharge rates will vary both in volume and duration based on site-specific factors. In Canada, some mining operations discharge seasonally. Reduced natural degradation of substances such as cyanide and ammonia under cold conditions sometimes makes it difficult to achieve discharge limits; therefore, wastewater is discharged during spring and summer. Wastewater may also be discharged during early spring to release large quantities of snow meltwater that has accumulated during the winter months.

The presence of fish in the initial receiving environment can also be a factor in the discharge of effluent volume. For example, the need to protect overwintering fish in pools when natural stream flow is minimal may cause a mine to reduce its effluent discharge to the exposure area. Other mines minimize discharges in late summer when stream flow is low and fish may be spawning. Conversely, spring discharge rates may be high, allowing water inventories to be relieved while stream flows are high. However, not all mines have sufficient reservoir capacity to optimize effluent releases.

Mine ore bodies are variable, resulting in variations in effluent discharges. Each ore is different, not only between mines but also within each mine itself. One ore body may contain more or less of certain minerals. In addition, while mines may have a short life span, the mill at the mine site may process ore from several mines. Virtually every mine has its own distinct suite of parameters of concern in effluents that influence site-specific factors related to bioavailability and hardness. However, similarities in operations among mines in a region may allow for efficiencies to be realized by a regional EEM study design.

Mines incorporate a variety of effluent treatment options (e.g., lime addition, settling ponds, water treatment plants) in their processes, and the type and effectiveness of each treatment option will influence the effluent discharged to the environment. The retention time in tailings ponds can affect the composition of the effluent. For example, cyanide breakdown and settlement of particulate matter are time-dependent. A sufficiently long residence time can potentially modify the concentrations in the effluent.

Local geology: Mine location is determined by the regional geology and the exact locations of ore deposits. Local mineralization around ore deposits influences the natural background levels of metals in streams. The net result is often a naturally higher

background concentration of metals in streams located near mine sites, which should be considered during reference area selection such that comparability to the exposure area is optimized.

Bioavailability of metals: The bioavailability of metals is an important aspect to consider when assessing the effects of mine effluent in the EEM program, as ongoing research continues to identify modifying factors that can change their bioavailability. For example, the effects of soluble metals on the biota can be mitigated during effluent treatment (e.g., adding lime to precipitate metals). In addition, the effect of mine effluent and associated sediments on the aquatic environment can vary during the mine's life cycle as the bioavailability of certain parameters changes.

2.2.2 Exposure and Reference Areas

An area is qualitatively defined for sampling purposes and relates to the appropriate geographical scale encompassing one or more fundamental sampling locations ("stations"). A station is a fixed sampling location that can be recognized, re-sampled and defined quantitatively (e.g., latitude/longitude). Within EEM, the overall study area is subdivided into reference and exposure areas (for control-impact designs) or within an exposure area where there are gradually decreasing effluent concentrations (for gradient designs). The MMER definition of exposure area is "all fish habitat and waters frequented by fish that are exposed to effluent," and the definition of reference area is "water frequented by fish that is not exposed to effluent and that has fish habitat that, as far as practical, is most similar to that of the exposure area." (MMER, Schedule 5, s. 1).

2.2.2.1 Selection of Final Discharge Point for Monitoring

In cases where the mine has more than one final discharge point, it is recommended that sampling be done in an exposure area where the effluent has the greatest potential to have an adverse effect on the receiving environment. The mass loadings of the deleterious substances, the manner in which the effluent mixes in the exposure area, and the sensitivity of the receiving environment should all be considered when selecting which final discharge point should be used for biological monitoring.

2.2.2.2 Selection of Exposure and Reference Areas

The selection of the sampling areas is one of the most critical components of the study design and should be considered carefully to maximize the quality of the information gained from the study. The design of biological surveys is site-specific, and various examples of potential study designs are presented in Chapter 4. However, this guidance is not intended to limit the mine's flexibility to propose other potential study designs that may be suitable to the site.

2.2.2.2.1 Exposure Area

Exposure area sampling should be done in an area proximate to the effluent discharge where effects may be found. Sampling areas should ideally support both appropriate habitat for the benthic invertebrate community and populations of the selected fish species. The study design should also consider the use of the exposure area by fish species (e.g., spawning, nursery). Identification of the exposure area and its habitat features should precede the selection of reference areas, because reference areas will, as far as practicable, match the physical and chemical habitat features of the exposure area (other than the features expected to change due to the effluent).

The exposure area may extend through a number of receiving environments (e.g., different stream orders, lakes or marshes, estuarine to marine, or intertidal to sub-tidal) and contain a variety of habitat types. In most cases, the boundary of the exposure area is defined by the zone of effluent mixing. Within an exposure area, there may be a high effluent and a low effluent exposure area often referred to as near-field and far-field areas, respectively. Additional sampling areas within the exposure area may be used during phases to assess magnitude and geographic extent of effects, or during periodic monitoring to provide an enhanced study or address site-specific needs. High effluent exposure (near-field) areas are outside the initial discharge zone (as described below) and have higher exposure to effluent than far-field areas. The initial discharge zone is the area where the effluent exceeds the velocity of the receiving water and the effluent is buoyant. The initial discharge zone is often characterized by visual turbulence and typically does not extend more than 5-50 m from the outfall. At least one of the high effluent exposure (near-field) stations should be as near as possible to the effluent discharge point but located outside the initial discharge zone. For magnitude and geographic extent studies, the exposure area extends along the effluent gradient so that additional lower effluent exposure areas (far-field areas) are included. The exposure area extends geographically until a return to reference area conditions (regulatory definitions of exposure and reference areas are provided above). Lower effluent exposure areas are recommended to be positioned close to the boundary of the zone of effluent mixing. Multiple sampling stations in each defined area should be used to determine spatial variation. In a gradient design, there is no reference area per se, but the response variables are evaluated along the effluent gradient.

In practical use, there will probably always be one or more low effluent exposure (far-field) areas for media other than fish (e.g., water, sediment and benthos). Recommended positioning of low effluent (far-field) exposure areas should be such that each area differs in regard to degree of effluent exposure. If possible, all exposure areas should be located so as to minimize or avoid exposure to non-mine discharges.

2.2.2.2.2 Reference Area

Reference areas do not need to represent pristine (pre-European settlement) conditions, but, rather, can comprise areas in which anthropogenic impacts, unrelated to the mine effluent, are similar to exposure areas (Simon 1991; Omernik 1995). Where feasible, the

reference area should be located in the same water body as the effluent discharge, upstream of or beyond any influence from the discharge. The reference area should be suitable physically and biologically, and outside the influence of the mine or other confounding factors. When a mine is located at a headwater, and/or where no suitable reference area on the same water body is available (e.g., dams and reservoirs may be upstream), the reference area should be located in an adjacent water body with similar characteristics or a non-impacted tributary to the receiving water body. Another possibility is sampling a number of exposure areas at increasing distances from the discharge point, representing an exposure gradient (gradient design). More than one reference area may be used, where appropriate. During magnitude and geographic extent studies, it may be necessary to sample more than one reference area if multiple exposure areas with different habitat types are sampled. As well, a more regional approach could be considered, particularly for benthic invertebrate community surveys, such as looking at several non-impacted streams (or lakes) in the area (i.e., reference condition approach).

Where historical monitoring data exist, the mine should consider using the same sampling areas from previous studies, provided they are appropriate for use in the EEM program. This will help ensure that monitoring data collected as part of the EEM program may be compared with historical data.

Baseline data (pre-effluent discharge) and multiple reference areas may assist in data interpretation. It is possible to use historical data as a baseline comparison to determine effects, but it should be treated as data additional to the mine's own. The mine's design should therefore include both a reference area and an exposure area (or follow a gradient design). This ensures that reference conditions have not changed and that changes observed are not incorrectly attributed to the mine, because there can be changes in parameters related to changes in environmental conditions (e.g., due to flooding or variability in annual temperatures). A reference area should be used to allow characterization of those changes that are mine-related compared to those that are not. Practitioners can also take advantage of the environmental assessment phase for new projects to provide information complimentary to the EEM program (Kilgour et al. 2007).

Where possible, sampling areas for different components (fish, benthic invertebrate community, water) should be co-located. The characteristics of the selected fish species, (e.g., mobility, habitat usage) and the different sampling gear may not always make this practical. The reference areas for benthic invertebrate sampling in some cases can be directly upstream of the exposure area, which may not be the case with fish (due to mobility). In addition, mines are encouraged to conduct benthic invertebrate community and fish monitoring studies concurrently, if justifiable biologically (e.g., if the ideal time for sampling fish reproductive parameters coincides with a suitable time for benthic community sampling; see Chapter 3 for additional information on timing for fish reproduction). Where there is more than one mine in close proximity and effluents are discharged to the same drainage basin, joint EEM studies are encouraged. Where studies are proposed jointly, sampling areas may be shared.

Data obtained from reference areas when compared to exposure areas can detect impairment of aquatic life (Yoder 1991), diagnose stressors (Hughes et al. 1994), provide data on temporal and spatial trends (Yoder 1989) and provide data for water resource summaries for government agencies (OEPA 1990). The identification of “least impacted” areas will differ across the country. Reference areas in extremely disturbed areas may be impossible to locate. Here, studies should be designed so that reference areas with minimal degradation are located in comparable drainage basins within the same ecoregion (Hughes et al. 1994).

Usually, coastal mines do not have strictly “upstream” areas for reference sampling, because of variable directionality of current due to tidal effects. Estuarine mines may have upstream areas that are too different physically and biologically to be suitable for reference sampling. The reference area is therefore usually at least periodically downstream from the effluent discharge. Accordingly, it is important to understand the current flow patterns in the area to determine whether a potential reference area is “outside” mine effluent exposure.

Selection of a reference area in the initial phase should not necessarily dictate its use in future phases.

In selecting sampling areas, the mine should take the following factors into consideration:

- the location of sampling areas in previous surveys;
- the location of confounding influences;
- the size of area needed to accommodate the number of samples to be collected;
- habitat type;
- site access; and
- other issues that could affect the mobility of fish.

In general, both sampling areas should be as follows:

- As similar as possible except for exposure to effluent. Although the two areas are unlikely to be identical, it is assumed that the differences in natural characteristics (e.g., depth, substrate, flow and water quality) will, other than mine-related factors, be small relative to the potential effect associated with the presence of effluent. If this is not the case, it should become apparent, and study design changes should be made in subsequent phases.
- Situated as closely as possible to each other (but sufficiently distant to be confident that fish from the reference area are not exposed to effluent).
- Accessible, and offer safe sampling during the most appropriate season (i.e., when measurements on fish growth, reproduction, condition and survival can be taken).
- Described in as much detail as possible, including the latitude and longitude as well as a written description of the area (physical, chemical and biological habitat, including measurements of temperature, depth and flow).

At a minimum, for a control/impact study, sampling should be conducted at no less than 1 reference area and 1 exposure area during the first EEM study, and subsequent EEM studies (studies aim to confirm presence or absence of effect and magnitude and geographic extent). The use of multiple reference areas offers the greatest statistical power to detect a meaningful difference between a reference area and an exposure area (Foran and Ferenc 1999). It can also give an indication of variability among reference areas (Munkittrick et al. 2000). Differences found in the exposure area that are outside the range of values seen at a number of reference areas are more likely to be ecologically relevant (Munkittrick et al. 2000). Sampling multiple reference areas is preferred over increasing sample size (e.g., number of fish) at a single area (Environment Canada 1997).

When possible there are advantages to selecting similar sampling areas for benthic and fish sampling so that the data can be used to help interpret responses. However, optimal benthic sampling areas may be inappropriate for the fish survey because of the characteristics of the fish species selected, the mobility of the fish, different habitat selection, and the type of sampling gear required. The sampling areas may be the same in many circumstances, but this should not be a sufficient criterion in itself for selecting the fish sampling area.

2.2.3 Reporting of Field Station Positions

The interpretative report shall contain the latitude and longitude of sampling areas in degrees, minutes and seconds; a description of the sampling areas sufficient to identify them is required (MMER Schedule 5, s. 17(b)). The latitude and longitude coordinates can be obtained using a variety of methods. Global positioning systems are a common tool for locating the position of field stations and are recommended for this purpose. In some instances, the coordinates with stream-wise distances (e.g., river kilometres) may be useful. The recommended positioning accuracy should be determined on a site-specific basis. In some cases, where there are multiple outfalls, industrial sites may decide to collaborate on their studies.

Additional stations may be included to better represent spatial patterns in a large zone of effluent mixing, such as at a location with transects (right, mid, left), and reference/high effluent exposure area (near-field) /low effluent exposure areas (far-field) sampling areas.

2.2.4 Modifying or Confounding Factors

Modifying or confounding factors can alter the interpretation of the results of biological monitoring. If the sampling areas are fairly similar, effects of modifying factors can be considered negligible. However, when there are significant differences among sampling areas, the survey design becomes confounded. In this case, it may be difficult to differentiate the effects of mine effluent from the effects of the modifying factor(s) on the response variable. For example, if the habitat type (e.g., pool) downstream of the mine

was different from the habitat found upstream (e.g., riffle), the effects of habitat type on variables would confound any effects related to mine effluent; both mine effluent and habitat type may be good predictors of any differences in variables observed upstream and downstream of the effluent discharge.

Incorporating multiple reference locations into the study design can aid in designing against spatial confounding factors, and practitioners are encouraged to do so. Design considerations for the detection of anthropogenic disturbances have been presented in the literature (see Green 1993 and references therein; Underwood 1994; Underwood 1997), and practitioners are encouraged to incorporate these considerations into their study designs.

Some examples of potential confounding factors include:

- tributaries and other point- and nonpoint-source discharges (e.g., other industrial discharges, agricultural runoff, aquaculture facilities, sewage treatment plants);
- natural environmental/habitat variables; and
- historical damage.

Attention to potential confounding factors identified during site characterization should be considered during the study design. In this way confounding factors can be minimized, or accounted for in the study design so that their influence can be assessed during data interpretation.

2.2.5 Tributaries and Other Point- and Nonpoint-Source Discharges

Tributaries provide dilution water to a main channel, lake, estuary or open ocean. This dilution water may or may not have similar chemical properties to the water body under study. Tributaries also require time and distance to mix with the water body under study. Therefore, if there is a tributary between the reference and exposure area, the additional flow from the tributary can potentially confound the interpretation of data.

Other point- and nonpoint-source discharges may make it difficult to distinguish between effects due to the mine effluent and other discharges, particularly if they are found within close proximity to the mine effluent discharge. When other discharges are present immediately above the mine, multiple reference areas should be used. One reference area should be set between the other discharge and the effluent discharge. In this way it may be possible to account for the influence of the other discharge. As well, the reference area should have similar background levels of metals. If no differences between the two reference areas are found, they can be pooled to compare against the exposure area.

If there are other point-source effluent discharges not related to mines in the study area, the study design should attempt to minimize the potential effects of the confounding factors. When it is not possible to resolve the confounding factors by modifying the study design, alternative sampling designs and methods should be evaluated.

2.2.6 Natural Variation in Environmental or Habitat Conditions

Natural biological communities can differ both temporally and spatially. Particularly if study areas are extensive, it is possible that natural biological communities and characteristics will be different from one location to another. It may be difficult to distinguish the influences of mine effluent, if any, relative to natural variation.

Examples of common, naturally occurring and sometimes confounding factors include:

- habitat type (riffle, run, pool);
- substrate type (organic content, particle size);
- water depth;
- water flow rate/discharge;
- tidal action / currents / wave exposure;
- salinity;
- dissolved oxygen/temperature;
- emergent/submergent vegetation cover;
- water chemistry (conductivity, hardness, pH, etc.); and
- biological properties.

Once present in the study design, confounding effects cannot be eliminated. Only by giving careful attention to potential modifying factors identified during the pre-design phase or the previous phases of the survey can the influence of modifying factors be removed or controlled in subsequent phases. Where it is not possible to eliminate confounding factors, increasing the number of sampling areas or including additional chemical and/or biological parameters may allow the investigator to assess their influence on data interpretation.

When it is not possible to resolve the confounding factors by modifying the study design, alternative sampling designs and methods (Chapter 9) should be considered.

2.2.7 Historical Damage

When the area in which a mine releases its effluent has been subject to damage from previous activities, it may be difficult to determine differences due to current effluent-release practices. The use of an alternative method may have to be used in these situations.

2.3 General Quality Assurance / Quality Control and Standard Operating Procedures

2.3.1 Quality Assurance and Quality Control

Detailed QA/QC is described in each chapter. QA/QC is a documented system

incorporating adequate review, audit and internal quality control. The objective of a QA/QC program is to ensure that all field sampling and laboratory analyses produce technically sound and scientifically defensible results.

QA is a planned system of operations and procedures, the purpose of which is to provide assurance to the client that defined standards of quality are being met. Analytical QA defines the way in which tasks are to be performed in order to ensure that data meet predefined data quality goals. These tasks include not only the analysis itself but all aspects of sample handling and data management.

QA encompasses a wide range of internal and external management and technical practices designed to ensure data of known quality commensurate with the intended use of the data. External QA activities include participation in relevant inter-laboratory comparisons and audits by outside agencies. Outside audits may be based on performance in analysis of standard reference materials, or on general review of practices as indicated by documentation of sampling, analytical and QA/QC procedures, test results, and supporting data. QC is an internal aspect of QA. It includes the techniques used to measure and assess data quality and the remedial actions to be taken when data quality objectives (DQOs) are not realized. Within the context of a particular study, assurance of adequate data quality is only possible when DQOs have been defined. Users of the data should play a lead role in defining DQOs for a study and in ascertaining whether laboratory quality-control limits are consistent with these objectives.

Data quality measures should be defined in the same terms as DQOs, so that the two can be compared in project evaluation. DQOs are normally derived from intended data uses (e.g., hypotheses to be tested, summary statistics involved, and total uncertainty that can be tolerated). Total uncertainty includes imprecision (sampling, analytical, environmental) plus any analytical bias that may occur (Taylor 1987). Objectives can be established for each component and for total uncertainty, and should be incorporated into the QA project plans. The various components of imprecision can be estimated using field replicate data and laboratory replicate data.

QA functions, the personnel responsible for each QA function, and corrective actions when performance limits are exceeded should be identified in the quality management plan.

QC activities define the boundaries of acceptable performance for the measurement system, and include the routine checks (data quality measures) that indicate whether the system is performing to specification. Data reporting generally stops and corrective actions are initiated when the system goes out of control. Range and average-control charting methods have been described elsewhere (OMOE 1984; ASTM 1985, 1986; Dux 1986).

An outline of recommended QA/QC requirements for specific components of the fish survey (Chapter 3), benthic invertebrate survey (Chapter 4), effluent and water quality analysis (Chapter 5), and sediment quality analysis (Chapter 7) are presented in each respective chapter. This information focuses on QC in the field, laboratory, data analysis, and

reporting.

2.3.2 Standard Operating Procedures

Standard operating procedures (SOPs) are fundamental to any QA/QC program. All field and lab procedures should be conducted according to SOPs to ensure quality control. SOPs should describe the following in detail:

- the field program's requirement for sampling methods and procedures, sample handling, labelling, equipment, preserving, record keeping, and shipping; and
- the analytical methods and procedures, sample handling, labelling, equipment, test system implementation, record keeping and so forth of all laboratory analyses.

Each SOP should be a written, detailed method accessible to each analyst. SOPs should be based on procedures developed by a standard-setting organization such as Environment Canada, the U.S. Environmental Protection Agency, the American Society for Testing and Materials, or the American Public Health Association. Where methods are not well validated, it is recommended that the SOP be thoroughly referenced to the relevant literature and contain all the elements outlined in CALA (1991). In-house validation of data should be appended to the SOP and should contain the QA/QC procedures, including the types and frequencies of QC samples to be analyzed, expected levels of precision, accuracy and recovery, and the method detection limits.

While chemical analysis procedures tend to be reasonably well documented, sampling procedures in general, and sampling design in particular, are often overlooked. Sampling error is usually a large component and often the dominant component of uncertainty in environmental measurements. SOPs that include field operations will help to reduce this uncertainty or at least ensure that it is quantified. All field staff should be familiar with the SOPs for any field survey work.

Emphasis should be placed on measures to prevent inadvertent contamination of samples and to ensure sample integrity. In addition, SOPs should specify the proper preparation of all sampling gear and supplies, and the proper calibration of all instrumentation (such as meters).

2.4 References

[ASTM] American Society for Testing and Materials. 1985. Standard practice for establishing conditions for laboratory sensory evaluation of foods and beverages. Philadelphia (PA): American Society for Testing and Materials. ASTM E480-84.

[ASTM] American Society for Testing and Materials. 1986. Physical requirement guidelines for sensory evaluation laboratories. Philadelphia (PA): American Society for Testing and Materials. ASTM STP 913.

Bisson PA, Montgomery DR. 1996. Valley segments, stream reaches, and channel units. In Hauer FR, Lamberti GA, editors. *Methods in stream ecology*. San Diego (CA): Academic Press. p. 23-52.

Booth J, Hay D, Truscott J. 1996. Standard methods for sampling resources and habitats in coastal subtidal regions of British Columbia: Part I - Review of mapping with preliminary recommendations. Canadian Technical Report of Fisheries and Aquatic Science 2118.

Budd WW, Cohen PL, Saunders PR, Steiner FR. 1987. Stream corridor management in the Pacific Northwest: 1. Determination of stream-corridor widths. *Environ Manag* 11:587-597.

Busch W-DN, Sly PG, editors. 1992. *The development of an aquatic habitat classification system for lakes*. Boca Raton (FL): CRC Press.

[CALA] Canadian Association for Laboratory Accreditation: formerly CAEAL (Canadian Association of Environmental Analytical Laboratories). 1991. Code of practice and QA manual for laboratory analysis of sewage treatment effluent in support of the MISA Program; Draft report prepared for CAEAL and the Ontario Ministry of the Environment by Zenon Environmental Laboratories.

Conquest LL, Ralph SC, Naiman RJ. 1994. Implementation of large-scale stream monitoring efforts: sampling design and data analysis issues. In Loeb SL, Spacie A, editors. *Biological monitoring of aquatic systems*. Boca Raton (FL): Lewis Publ. p. 69-90.

Corkum LD. 1989. Patterns of benthic invertebrate assemblages in rivers of northwestern North America. *Freshwat Biol* 21:191-205.

Corkum LD. 1992. Spatial distributional patterns of macroinvertebrates along rivers within and among biomes. *Hydrobiologia* 239: 101-114.

Corkum LD. 1996. Responses of chlorophyll *a*, organic matter and macroinvertebrates to

nutrient additions in rivers flowing through agricultural and forested land. Arch Hydrobiol 136:391-411.

Cowardin LM, Carter V, Golet FC, LaRoe ET. 1979. Classification of wetlands and deepwater habitats of the United States. U.S. Fish and Wildlife Service. FWS/OBS-79/31.

Cowardin LM, Golet CF. 1995. US Fish and Wildlife Service 1979 wetland classification: a review. Vegetatio 118:139-152.

Cupp CE. 1989. Valley segment type classification for forested lands of Washington. Timber, Fish & Wildlife AM-89-001.

Department of Fisheries and Oceans. 1990. Coastal/estuarine habitat description and assessment manual. Part II. Habitat description procedures. Coquitlam (BC): Prepared by G.L. Williams and Associates Ltd.

Department of Fisheries and Oceans and British Columbia Ministry of the Environment and Parks. 1987. Fish Habitat Inventory and Information Program: Stream Survey Field Guide.

Dolman WB. 1990. Classification of Texas reservoirs in relation to limnology and fish community associations. Trans Am Fish Soc 119:511-520.

Dux JP. 1986. Handbook of quality assurance for analytical chemistry laboratory. New York (NY): Van Nostrand Reinhold Co.

Eadie JM, Keast A. 1984. Resource heterogeneity and fish species diversity in lakes. Can J Zool 62:1689-1695.

Environment Canada. 2003. Revised technical guidance on how to conduct effluent plume delineation studies. Available from: www.ec.gc.ca/esee-eem/D450E00E-61E4-4219-B27F-88B4117D19DC/PlumeDelineationEn.pdf

Environment Canada. 1997. Fish Survey Expert Working Group report. EEM/1997/6.

Finlayson CM, van der Valk AG. 1995. Wetland classification and inventory. A summary. Vegetatio 118(1-2):185-192.

Fischer HB, List EJ, Koh RCY, Imberger J, Brooks NH. 1979. Mixing in inland and coastal waters. San Diego (CA): Academic Press Inc.

Foran J, Ferenc S. 1999. Multiple stressors in ecological risk and impact assessment. Pensacola (FL): SETAC Press.

Freeze RA, Cherry JA. 1979. Groundwater. Englewood Cliffs (NJ): Prentice-Hall. .

Frissell CA, Liss WL, Warren CE, Hurley MD. 1986. A hierarchical framework for stream habitat classification: Viewing streams in a watershed context. *Environ Manag* 10:199-214.

Frith HR, Seraring G, Wainwright P, Harper H, Emmett B. 1993. Review of habitat classification systems and an assessment of their suitability to coastal B.C. Unpub. report to Environment Canada from L.G.L. Ltd., Sidney (BC).

Green RH. 1993. Application of repeated measures designs in environmental impact and monitoring studies. *Australian J Ecol* 18:81-98.

Hay DE, Waters RD, Boxwell, editors. 1996. Proceedings Marine Ecosystem Monitoring Network Workshop, Nanaimo, B.C. March 28-30. 1995. Canadian Technical Report of Fisheries and Aquatic Science 2108.

Hawkins CP, Kershner JL, Bisson PA, Bryant MD, Decker LM, Gregory SV, McCullough DA, Overton CK, Reeves GH, et al. 1993. A hierarchical approach to classifying stream habitat features. *Fisheries* 18:3-12.

Hughes RM. 1995. Defining acceptable biological status by comparing with reference conditions.. In Davis WS, Simon TP, editors. *Biological assessment and criteria: tools for water resource planning and decision making*. Boca Raton (FL): Lewis Publishers. p. 31-47.

Hughes RM, Heiskary SA, Matthews WJ, Yoder CO. 1994. Use of ecoregions in biological monitoring. In Loeb SL, Spacie A, editors. *Biological monitoring of aquatic systems*. Boca Raton (FL): Lewis Publ. p. 125-149.

Hutchinson GE. 1957. *A treatise on limnology. Vol I. Geology, physics, and chemistry*. New York (NY): John Wiley & Sons Inc.

Kerr SR, Ryder RA. 1988. The applicability of fish yield indices in freshwater and marine ecosystems. *Limnol Oceanogr* 33:973-981.

Kilgour BW, Dubé MG, Hedley K, Portt CB, Munkittrick KR. 2007. Aquatic environmental effects monitoring guidance for environmental assessment practitioners. *Environ Monit Assess* 130:423-436.

Leopold LB. 1994. *A view of the river*. Cambridge (MA): Harvard University Press.

Leopold LB, Wolman GM, Miller JP. 1964. *Fluvial processes in geomorphology*. San Francisco (CA): W.H. Freeman & Co.

Levings CD, Thom RM. 1994. Habitat changes in the Georgia Basin: Implications for resource management and restoration. In Wilson RCH, Beamish RJ, Aitkens F, Bell J,

editors. Review of the marine environment and biota of Strait of Georgia, Puget Sound and Juan de Fuca Strait: Proceedings of the BC/Washington Symposium on the Marine Environment, Jan 13,14 1994. Canadian Technical Report of Fisheries and Aquatic Science 1948. p. 330-351.

Marshall TR, Ryan PA. 1987. Abundance and community attributes of fishes relative to environmental gradients. *Can J Fish Aquat Sci* 44:196-215.

Matthews GWT. 1993. The Ramsar Convention: its history and development. Glan (CH): Ramsar Convention Bureau.

Maxwell JR, Edwards CJ, Jensen ME, Paustian SJ, Parrott H, Hill DM. 1995. A hierarchical framework of aquatic ecological units in North America (Nearctic Zone). St. Paul (MN): U.S. Department of Agriculture, Forest Service, North Central Forest Experimental Station. Gen. Tech. Rep. NC-176.

Meador MR, Hupp CR, Cuffney TF, Gurtz ME. 1993. Methods for characterizing stream habitat as part of the national water-quality assessment program. Raleigh (NC): U.S. Geological Survey. Open-File Report 93-408.

Metal Mining EEM Review Team. 2007. Report of the Metal Mining EEM Review Team. Available from: <http://www.ec.gc.ca/ese-eem/default.asp?lang=En&n=2DAFFC56-1>.

Montgomery DR, Buffington JM. 1993. Channel classification, prediction of channel response, and assessment of channel condition. Olympia (WA): Department of Natural Resources. Washington State Timber/Fish/Wildlife Agreement. Report TFW-SH10-93-002.

Munkittrick KR, McMaster M, Van Der Kraak G, Portt C, Gibbons W, Farwell A, Gray M. 2000. Development of methods for effects-based cumulative effects assessment using fish populations: Moose River Project. Pensacola (FL): SETAC Press.

Newbury RW. 1984. Hydrological determinants of aquatic insect habitats. In Resh VH, Rosenberg DM, editors. The ecology of aquatic insects. New York (NY): Praeger. p. 323-357.

Newbury RW, Gaboury MN. 1993. Stream analysis and fish habitat design. A field manual. Newbury Hydraulics Ltd., The Manitoba Habitat Heritage Corporation, Manitoba Fisheries Branch.

[OEPA] Ohio Environmental Protection Agency. 1990. Ohio water resource inventory. Columbus (OH): Ohio Environmental Protection Agency.

Omernik JM. 1995. Ecoregions: a spatial framework for environmental management. In

Davis WS, Simon TP, editors. Biological assessment and criteria. Tools for water resource planning and decision making. Boca Raton (FL): Lewis Publishers. p. 49-62.

Ontario Ministry of Natural Resources. 1989. Manual of Instructions. Aquatic Habitat Inventory Surveys. Toronto (ON): Ontario Ministry of Natural Resources.

[OMOE] Ontario Ministry of the Environment. 1984. Principles of control charting. King DE, Ronan RC, editors. Laboratory Services Branch, Data Quality Report Series. Rexdale (ON): Ontario Ministry of the Environment.

Orth DJ. 1989. Aquatic habitat measurements. In Neilson LA, Johnson DL, editors. Fisheries Techniques. Bethesda (MD): Am. Fish. Soc. p. 61-84.

Peterjohn WT, Correll DL. 1984. Nutrient dynamics in an agricultural watershed: observation on the role of a riparian forest. *Ecology* 65:1466-1475.

Plafkin JL, Barbour MT, Porter KD, Gross SK, Hughs RM. 1989. Rapid bioassessment protocols for use in streams and rivers: benthic macroinvertebrates and fish. EPA/444/4-89-001.

Reynoldson TD, Rosenberg DM. 1996. Sampling strategies and practical considerations in building reference data bases for the prediction of invertebrate community structure. In Bailey RC, Norris RH, Reynoldson TB, editors. Study design and data analysis in benthic macroinvertebrate assessments of freshwater ecosystems using a reference site approach. Technical Information Workshop North American Benthological Society, 44th Annual Meeting, Kalispell, Montana. p. 1-31.

Robinson CLK, Levings CD. 1995. An overview of habitat classification systems, ecological models and geographic information systems applied to shallow foreshore marine habitats. Canadian Management Report of Fisheries and Aquatic Science 2322.

Robinson CLK, Hay DE, Booth J, Truscott J. 1996. Standard methods for sampling resources and habitats in coastal subtidal regions of British Columbia: Part 2 - Review of sampling with preliminary recommendations. Canadian Technical Report of Fisheries and Aquatic Science 2119.

Ryder RA. 1965. A method for estimating the potential fish production of north temperate lakes. *Trans Am Fish Soc* 94:214-218.

Scott DA, Jones TA. 1995. Classification and inventory of wetlands. *Vegetatio* 118:1-16.

Simon TP. 1991. Development of biotic integrity expectations for the ecoregions of Indiana. I. Central Corn Belt Plain. U.S. Chicago (IL): Environmental Protection Agency, Region V, Environmental Sciences Division, Monitoring and Quality Assurance Branch. Ambient Monitoring Section. EPA 905/9-91/025.

Snelgrove RVR, Butman CA. 1994. Animal sediment relationships revisited: cause versus effect. *Oceanogr Mar Biol Ann Rev* 32:111-178.

Strahler AN. 1957. Quantitative analysis of watershed geomorphology. *Am Geophys Union Trans* 38:913-920.

Taylor JK. 1987. Quality assurance of chemical measurements. Chelsea (MI): Lewis Publishers Inc.

Thorson G. 1957. Bottom communities(sublittoral or shallow shelf). In Hedgpeth JW, editor. *Treatise on marine ecology and paleoecology*. Vol 1..Memoirs of the Geological Society of America 67. p. 461-534.

Underwood AJ. 1994. On beyond BACI: sampling designs that might reliably detect environmental disturbances. *Ecol Applications* 4(1):3-15.

Underwood AJ. 1997. *Experiments in ecology: their logical design and interpretation using analysis of variance*. Cambridge University Press. United Kingdom.

Wetzel RG. 1975. *Limnology*. Philadelphia (PA): W.B. Saunders Co.

Wickware GM, Rubec CDA. 1989. Ecoregions of Ontario. *Ecological Land Classification Series* 26. Environment Canada.

Wiken E. 1986. Terrestrial ecotones of Canada. *Ecological Land Classification Series* 19. Ottawa (ON): Environment Canada.

Winter WT. 1977. Classification of hydrological settings of lakes in the north-central United States. *Water Resources Research* 134:753-767.

Winter WT, Woo MK. 1990. Hydrology of lakes and wetlands.. In Wolman MG, Riggs HC, editors. *Surface water hydrology*. Boulder (CO): The Geological Society of America. p. 159-187.

Yoder CO. 1989. The development and use of biological criteria for Ohio surface waters.. In *Water quality standards for the 21st century*. Washington (DC): U.S. Environmental Protection Agency, Office of Water. p. 139-146.

Yoder CO. 1991. Answering some concerns about biological criteria based on experiences in Ohio. In *Water quality standards for the 21st century*. Washington (DC): U.S. Environmental Protection Agency, Office of Water. p. 95-104.

Young RA, Huntrods T, Anderson W. 1980. Effectiveness of vegetated buffer strips in controlling pollution from feedlot runoff. *J Environ Qual* 9:483-497.

Table of Contents

3.	Effects on Fish and Fisheries Resources.....	3-1
3.1	Overview.....	3-1
3.2	Study Design Considerations.....	3-1
3.2.1	Selection of Reference and Exposure Areas.....	3-2
3.2.1.1	Sampling of Exposure and Reference Areas	3-3
3.2.1.2	Adequacy of the Reference Area	3-3
3.2.2	Confounding Factors.....	3-4
3.2.3	Marine Discharges	3-4
3.2.4	Historical Data	3-5
3.3	Selection of Sentinel Fish Species	3-5
3.3.1	Community Survey	3-7
3.3.2	Immature Fish	3-7
3.3.3	Small-bodied Fish Species.....	3-8
3.3.4	Live-bearers	3-9
3.4	Effect Indicators.....	3-10
3.4.1	Lethal Sampling.....	3-10
3.4.1.1	Survival.....	3-12
3.4.1.2	Energy Use (Growth and Reproduction)	3-13
3.4.1.3	Energy Storage (Condition)	3-14
3.4.1.4	Abnormalities.....	3-15
3.4.2	Non-lethal Sampling.....	3-16
3.4.2.1	Survival (Size Distributions)	3-17
3.4.2.2	Energy Use.....	3-17
3.4.2.3	Condition.....	3-19
3.4.3	Wild Molluscs	3-19
3.5	Timing of Sampling	3-20
3.6	Verification of Fish Exposure.....	3-24
3.7	Power Analysis	3-26
3.7.1	Power and Significance Level	3-26
3.7.2	Effect-Size.....	3-26
3.8	Fish Sampling Methods	3-27
3.8.1	Bycatch	3-28
3.8.2	Remote Sensing	3-28
3.8.3	Alternative Methods	3-28
3.9	Fish Survey Quality Assurance and Quality Control.....	3-29
3.9.1	Field Practices to Improve Data Analysis and Interpretation	3-29
3.9.2	Quality Control in the Field	3-29
3.9.3	Determination of Sampling Effort	3-30

3.9.3.1	Examples of Calculations of Sampling Effort	3-31
3.9.4	Consultation with Regional EEM Coordinators and Implementation in the Field.....	3-33
3.9.5	Data Entry	3-34
3.9.6	Quality Control in the Laboratory	3-34
3.10	Data Analysis.....	3-35
3.10.1	Statistical Analysis.....	3-36
3.11	Methods for Analysing Fish Usability.....	3-36
3.11.1	Selection of the Fish Species	3-38
3.11.2	Tissue Sample Collection and Preparation	3-38
3.11.3	Supporting Analyses: Lipid and Moisture Percentages.....	3-38
3.11.4	Guidance for Fish Tissue Analysis for Mercury Using Non-Lethal Methods.....	3-39
3.11.4.1	Recommended Methodology	3-40
3.12	References.....	3-42

List of Tables

Table 3-1: Required fish survey measurements, expected precision and summary statistics.....	3-11
Table 3-2: Suggested aging structures for Canadian fish species.....	3-13
Table 3-3: Fish survey effect indicators and endpoints for various study designs	3-15
Table 3-4: Generalizations and suggested optimal sampling times for fish species in EEM	3-21
Table 3-5: Fish species commonly used in EEM, aspects to consider during study design, and recommended sampling times.....	3-22
Table 3-6: Suggested reporting format for the parameters (A) and the resulting regressions (B) required for the fish survey analysis.....	3-36

3. Effects on Fish and Fisheries Resources

3.1 Overview

Fish monitoring for the environmental effects monitoring (EEM) program may consist of a fish population survey and tissue analyses to determine if the mine effluent is having an effect on fish and fisheries resources. Detailed requirements and timelines are found in Chapter 1 and in the *Metal Mining Effluent Regulations* (MMER) (SOR/2002-222).

For the purposes of EEM, fish includes shellfish, crustaceans and marine animals, as per section 2 of the *Fisheries Act*, but excludes parts of these organisms (MMER Schedule 5, section (s.) 1).

The fish survey provides an assessment of whether there are differences in the growth, reproduction, condition and survival of the fish population between exposed and reference areas or within an exposure area along a gradient of effluent concentrations. Note that a mine is required to conduct a study of the fish population if the concentration of effluent in the exposure area is greater than 1% in the area located within 250 metres (m) of a final discharge point (MMER, Schedule 5, s. 9(b)).

In addition to the fish survey, biological monitoring studies may also include a study respecting fish tissue if, during effluent characterization conducted under paragraph 4(1)(d), a concentration of total mercury (inorganic and organic mercury) in the effluent is identified that is equal to or greater than 0.10 µg/L (MMER, Schedule 5, s. 9(c)). “Effect on fish tissue” is defined as measurements of concentrations of total mercury that exceed 0.5 µg/g wet weight in fish tissue taken in an exposure area and that are statistically different from and higher than the measurements of concentrations of total mercury in fish tissue taken in a reference area (MMER, Schedule 5, s. 1).

3.2 Study Design Considerations

General information regarding study designs is discussed in Chapter 2. The study design requirements and the definitions of effect for the fish population survey and fish tissue survey are discussed in Chapter 1.

To evaluate the effect of effluent on fish, the following questions should be answered:

- Is there an effect?
- Is the effect mine-related?
- Is the magnitude and extent of the effect known?
- Is the mine-related cause of the effect known?

Each mine’s EEM representatives or consultants should consult with the regional EEM authority to review the results of the previous phase’s site selection, species selection,

fishing effort, etc., and to discuss the selection of the most appropriate options for the current phase. The results of previous phases, historical data, and local knowledge should be used to assess:

- the suitability and capture success of selected sentinel species
- the adequacy of the reference area
- the appropriateness of sampling methods and required equipment.

Mines may want to make changes between phases, including increasing sampling effort; changing sampling methods, equipment or fish species; selecting different exposure or reference areas; or using alternative monitoring techniques. Changes in the study design may need to be made for various reasons, including the following:

- The results indicate that power was insufficient in the previous phase due to collection of a low number of fish or high variability.
- The species characteristics were not measurable, not suitable, or there are concerns about the status of fish populations.
- It is uncertain if the fish were exposed to effluent.
- Reference sites were inappropriate.

Concerns raised about EEM studies (and field studies in general) can be separated into concerns about the adequacy of the reference sites, the potential impacts of confounding factors (e.g., potential influences of genetics on the variability of species characteristics), the ecological relevance of effect indicators used, the influences of natural variability, and concerns over statistical design issues. This guidance will attempt to provide input to deal with these issues.

3.2.1 Selection of Reference and Exposure Areas

The two main study designs are control-impact and gradient designs. The choice of reference area is the number one problem with control-impact field studies (Munkittrick et al. 2009). Ideally, a reference site would be located upstream, in similar habitat, and free of confounding influences, with a natural barrier that limits movement between sites. This situation is seldom available. The main issues cited regarding reference areas include whether the reference site is (a) comparable in terms of habitat; (b) free from the issue of concern (i.e., exposure) and from confounding influences (further discussed in Chapter 2); (c) open to movement of fish from the exposure area (fish in an upstream reference area could have been exposed previously or fish in the exposure area could be transient, reducing exposure to potential effects); and (d) whether the exposed fish were exposed to the effluent or stressor of interest.

No reference site is perfect. The ideal situation involves having data from before construction or initiation of the stressor of interest (e.g., before/after control-impact [BACI] design; outlined further in Chapter 4). Study sites that have barriers that prevent

fish from moving between sites (e.g., dams, waterfalls, beaver dams) may be a good alternative, providing that the barrier does not alter the habitat. In open-water areas, choosing a species that has limited mobility improves the confidence that fish are not moving, but increases the potential influences of local differences in habitat. One difficult situation to interpret arises when there are no statistical differences in fish measurements between the areas, and there are no barriers restricting movement. In these cases, an indicator of exposure to the effluent is recommended, which can be chemical or physiological (e.g., liver enzyme induction, stable isotope signatures [Galloway et al. 2003; Dubé et al. 2006]).

If there are significant differences in fish characteristics between reference and exposure areas, there can be high confidence that fish are not moving between sites. Although differences are seen, variability in fish parameters (e.g., growth, weight, condition) is a function of a number of factors, not all of which will be related to the discharge of effluent. The selection of appropriate fish species for monitoring, survey timing and sampling gear will also facilitate the interpretation of any differences detected. Nevertheless, other natural and anthropogenic factors may influence effects on the fish and fish tissue and confound interpretation of the data. The requirement to confirm effects was developed to increase the confidence, over two phases, that effects are mine-related.

3.2.1.1 Sampling of Exposure and Reference Areas

The exposure area should be selected to ensure that the fish collected have been exposed to the effluent. It should be sampled first to determine which fish species are present, and their relative abundance within the area. The reference area can then be selected to provide fish of the same species that are available at the exposed site. Timing of sampling should be as close as possible between sites, to minimize temporal variability. The choice of time period for sampling will depend on factors such as time of year, stage of reproductive development, and potential habitat differences between sites (water temperature differences, etc.), but it is recommended that, if possible, all sampling be done within the same week. If a longer time period is required, reference samples should be collected before and after the collection of exposure samples, to allow comparison.

If fish are found in the reference area, but not in the exposure area where they are expected to occur (i.e., fish were historically found in the sampling area), the absence of fish in the exposure area should be reported as an effect. More information on reference site selection can be found in Chapter 2.

3.2.1.2 Adequacy of the Reference Area

It is now common in research programs to use a large number of reference sites. As an example, over the first 3 cycles of monitoring for the pulp and paper EEM program, there has been a trend toward using more reference sites. In Cycle 1, 3% of studies used

multiple reference sites; in Cycle 2, 9%; and in Cycle 3, 25%. Including additional reference sites increases the ability to evaluate issues related to natural variability, ecological relevance and confounding factors, and improves the ability to evaluate the adequacy of the chosen reference site. Studies that use a gradient approach and multiple reference sites are statistically stronger than studies that depend only on a single reference site.

Other new approaches include reference condition approaches (Bailey et al. 1998), and using negative reference sites (using the exposed site as your reference). Regardless, the existence of consistent changes over two phases increases the level of confidence that changes are real. Follow-up studies must evaluate the adequacy of the reference site, especially if consistent results are not found.

3.2.2 Confounding Factors

In Cycle 2 of the pulp and paper EEM program, almost 90% of studies that detected effects also concluded that factors other than pulp mill effluent were responsible for such observations. Potential confounding factors exist at most sites and include other outfalls, habitat changes, historical uses and contamination, tributaries and non-point-source inputs. In highly confounded situations, alternative methods should be considered, but it should be emphasized that it is possible to obtain interpretable field results at most sites with adjustments to the study design. Given the complexity of certain situations, it is recommended that as much data as possible be gathered in order to demonstrate that other discharges or contaminant sources are primarily responsible for observed changes or an absence of observed changes. If changes are seen and determined to be influenced by confounding factors, the objective of subsequent study designs should be to eliminate the confounding factors or determine their significance.

3.2.3 Marine Discharges

Metal mines discharging into marine or estuarine receiving waters may face a number of problems and confounding factors that should be considered when developing an EEM fish survey study design. These problems may include the following:

- Some marine and estuarine areas are difficult to sample (e.g., tides, currents, high flushing rates or unsuitable habitat) and alternative approaches should be considered.
- There can be gradients for current, temperature and salinity, which can affect physical processes and the uptake of contaminants as well as have consequences for physiological changes within organisms.
- Selection of reference areas can be more difficult in marine situations.
- Different life stages of fish may utilize different habitats at different times of the year.
- Species availability can be low in marine environments. In many situations,

small-bodied resident fish species are available and should be investigated. These species may be multiple spawners or live-bearers, or species for which there is little background information. However, this should not restrict or inhibit attempts to use these species, especially if they are abundant. The assumption inherent in an EEM program is that a fish community should be intact, with the normal abundant species present. The second priority, and underlying assumption, is that a fish population which shows a growth rate, reproductive development, and an age distribution indistinguishable from a reference area, is unaffected.

Potential solutions to these difficulties include using alternative species or caged bivalves, or mesocosms for confounded receiving environments. New facilities that will have collected baseline information prior to initiating effluent discharge will be in a better position to assess the effect of their effluent on the receiving environment compared to confounding factors.

3.2.4 Historical Data

Mines have the opportunity to submit historical data if there is previous biological monitoring information that could determine if there are effects on fish, fish tissue or the benthic invertebrate community. Historical data can be used to assist in the development of the first EEM study. See Chapter 13 for additional information on historical data.

3.3 Selection of Sentinel Fish Species

The recommended method for carrying out the fish survey is to monitor adults of two species of relatively sedentary finfish that have been exposed to effluent over a long period of time. Sexually mature finfish are preferred, but where they are not available, it is possible to design a program using shellfish or juvenile fish, although it will not be possible to analyze all the same effect endpoints. If available, at least one of the species selected should be a benthivore. The most important factors when selecting fish species for the EEM program are exposure, abundance, relevance to the study area (Munkittrick et al. 2000; McMaster et al. 2002), and sensitivity to effluent. In selecting the two species, the species used in previous EEM studies at the site (if applicable) should be considered, and preference should be given to:

- resident (non-migratory) fish species identified in a site characterization
- sexually mature female and male fish species that are abundant in both the exposure and reference areas
- fish species for which fishing or sampling permits can be obtained
- fish species that have the highest exposure to effluent

At any given site, there may be limited choices of potential species for monitoring. It will often be necessary to obtain the advice of an experienced fisheries biologist with knowledge of fish species present in the study area. More than 60 species have been used

as sentinels in EEM pulp and paper and metal mining programs to date, and mines and their consultants are encouraged to contact regional, federal and provincial government agencies for fisheries information and additional guidance.

Some receiving environments do not support adequate numbers of fish for sampling. In situations where it has been determined that fisheries resources may be impacted by a destructive fish survey, non-lethal sampling techniques may be used. In environments that do not support adequate numbers of fish to meet the recommended sample sizes or where there are not two suitable fish species for monitoring, the following options, in order of preference, may be considered:

- one sexually mature fish species and one sexually immature fish species
- two sexually immature fish species
- one sexually mature fish species
- one sexually immature fish species.

The mine should consider changing its study design (e.g., species, methods of collection) if the results from the previous phase suggest that the species is long-lived (> 30 years); that it was not possible to measure all survey parameters on the fish (e.g., age, liver and gonad weight); that an insufficient number of individuals were collected; and that the degree of variability was such that the numbers of fish required by power analysis for subsequent designs are unreasonable, and it is not possible to reduce this by selective sampling methods. If the fish species available at a site are present in the high effluent exposure (near-field) area only during certain times of the year or life-history stages, the life stage and sampling time should be selected to maximize exposure to effluent.

Some of the challenges with species selection may relate to attempts to design a single program for multiple purposes. Concerns about contamination of fishery resources for human consumption would direct the study design to collect a species that is long-lived (so that contaminants accumulate longer), is piscivorous (so that biomagnification is greater), matures late (to increase concentration), preferably focuses on male fish or species that do not spawn every year (so that elimination of contaminants through egg deposition is lessened), and are of importance for local consumption. To improve the sensitivity of detecting environmental impacts, it is preferable that species are benthic (because generally they will move less), are not commercially or recreationally important (because it obscures the determining cause), mature early, contribute much energy to reproduction (so that energy demands are high), and are short-lived (so that impacts are recent)—with a focus on female fish (environmental impacts are often more serious on female egg producers).

A number of other factors need to be considered when selecting a sentinel species (see Munkittrick and McMaster 2000; Munkittrick et al. 2000), including ensuring that the species are active participants in the local aquatic food web. Other life-history characteristics, such as spawning time and migration, need to be evaluated

site-specifically, because the interaction between discharge site, spawning habitats, seasonal changes in flow and dilution can all influence results and potentially impact the sensitivity of the monitoring program.

A key consideration when selecting a species is the mobility and residence time of that species, as this determines effluent exposure. Species that are resident in the system for most or all of their life cycle and exhibit territorial behaviour or limited mobility relative to the size of the study area are preferred, because the observed responses of these species reflect their localized environment. Species that are migratory or spend only a small proportion of their life cycle in the system under investigation (e.g., anadromous salmonids, some marine fishes) are not suitable, because exposure to effluent is minimal or transient and difficult to determine. This is also true for species that are highly mobile and are likely to be moving in and out of the effluent exposure area. In some cases, it may be possible to select more mobile species (e.g., Mountain Whitefish) (Swanson 1993), due to physical constraints that limit movement (e.g., dams, natural barriers, changes in habitat). In general, the greater the likelihood that a fish species is exposed to effluent, the greater its value as a monitoring species.

3.3.1 Community Survey

If a mine is new or has no historical survey information available, a fish community survey should be done to aid in the selection of appropriate fish species. Fish community surveys evaluate whether there are differences between areas in the diversity and abundance of fish species present.

A change in fish community has occurred when species that are expected to be abundant from the collections conducted at reference areas are not present in the effluent discharge area. If the exposure areas do not support one or more of the abundant species found at the reference area, it will be necessary to document the geographical extent of this absence. When the fish community composition has changed because of the presence of an effluent, there will also likely be measurable changes in the fish populations that remain. Results from the EEM program should document this, and may help in determining whether other fish species are at risk of disappearing from the exposure area.

Fish communities often include a number of species that are not abundant for a variety of reasons that may be unrelated to the presence of mine effluent. Non-lethal techniques (e.g., electro-fishing) are preferred for the community survey where possible, and field sampling should be designed to limit mortality of the existing species.

3.3.2 Immature Fish

The recommended method for carrying out a fish survey is to monitor adults (sexually mature fish) of two species of relatively sedentary finfish that have been exposed to effluent over a long period. However, there have been situations where no adult fish can

be collected in a receiving environment. For example, some areas may not be inhabited by adult finfish, but are nursery areas for their juveniles. If sexually mature fish do not reside in an effluent exposure area, the suitability of juvenile fish may be considered. When sexually immature species are used, there is no direct measurement of reproductive development. However, the relative abundance of young of the year (YOY) can be used as a measurement of reproductive success.

Relevant measures for juvenile fish would be similar to those of mature fish, but without gonad measurements: growth (length, weight, or weight-at-age, if possible); condition (length-at-body-weight relationships); liver-weight-to-body-weight ratio; abundance (YOY survival, percent composition of age classes); deformities associated with exposure to effluents, such as vertebral fusions and compressions, spinal curvatures including lordosis and scoliosis, and fin erosion; and growth in juveniles exposed to effluent compared to juveniles in the reference area. Methods for the collection of juvenile fish are well established and many juvenile fishes can be aged (e.g., Secor et al. 1995).

3.3.3 Small-bodied Fish Species

The trend toward the increasing use of small-bodied forage-fish species (Munkittrick et al. 2002) has continued, rising from their use in 10% of surveys in the pulp and paper Cycle 1, to 26% in Cycle 2 and 34% in Cycle 3. A small-bodied fish can be considered a fish species that has a maximum size of 150 mm or less. Their use has several advantages and disadvantages. On a practical level, small-bodied fish species are usually more abundant, easy to capture, and more sedentary than larger-bodied fish species. Small-bodied fish have also been shown to be more sensitive to environmental changes, such as pH (Shuter 1990). Their home-range size has been positively correlated with body size (Minns 1995), and many small-bodied species integrate local conditions very well.

On the other hand, small-bodied fish require more sensitive analytical balances and more careful measurements. They are more sensitive to microhabitat differences because they integrate the local habitat so well. They are also more sensitive to differences in timing of sampling (see section 3.5).

In addition, small-bodied fish often have a shorter life span, so if they are chosen as one or both of the fish species, an additional 20 sexually immature fish (0+ and 1+) should be collected to aid in size-at-age (growth) analysis. Also, because a small-bodied fish species may only have a life expectancy of 3 to 4 years, the 0+ and 1+ will constitute a significant portion of the population (e.g., the 0+ and 1+ Slimy Sculpin [*Cottus cognatus*] are up to 50–70% of the population). This measurement of the proportion of a sample composed of YOY fish does add another surrogate measurement for reproductive performance (Gray et al. 2002).

There are other considerations as well. The life history, biology, and reproductive characteristics of some small-bodied species are unknown, making it difficult to determine the best sample areas, times and methods. Some are multiple spawners, which means reproductive effort in these species is difficult to estimate from a single sample because the reproductive tissue can be turned over almost completely between clutches (i.e., most of the mass of ova in the ovary will be spawned, and then a new clutch of mature ova will be developed). The ovary will generally contain two or more class sizes of ova and the spawning season may last from several weeks to more than a month. The number of clutches produced during the spawning season becomes the important reproductive variable and is difficult to estimate for an individual female in the field, even with frequent sampling. It will be difficult to evaluate the significance of changes in egg production in multiple spawners if they show normal reproduction in the first clutches.

Species identification of small-bodied fish should be verified, especially for cyprinids, which can appear very similar without careful examination. Useful references for this purpose include Scott (1967), Scott and Crossman (1973), Roberts (1988), Nelson and Paetz (1992), Jenkins and Burkhead (1993) and Coad et al. (1995). The smaller organ size of these fish requires a more sensitive balance. Dissecting microscopes may be necessary for removing the organs properly and avoiding extraneous tissue or moisture, which could affect results. Dissection on recently collected, fresh fish is recommended. Differentiation among tissues and separation of the liver and gonads from intestinal tissue is easiest when the tissue is fresh. Dissection of frozen specimens of small fish can be difficult and lead to errors in organ measurements. Preservation in a formalin solution may give adequate results, but care must be taken to treat exposure and reference fish the same (e.g., duration of storage) in order to minimize preservation distortion.

Measurement of fecundity and egg weight requires special consideration. Many small species have few, large eggs. Gonadal estimates will be easier closer to spawning. The timing of sampling will also be affected by residency, and the two factors have to be optimized. The entire gonad should be preserved and fecundity counts conducted with the aid of a dissecting microscope.

3.3.4 Live-bearers

Live-bearers are not common in Canadian freshwater receiving environments, but if used, require special attention regarding measurement of reproductive variables. Live-bearing species have been used successfully for detecting responses in exposures to Swedish pulp mills (Larsson et al. 2000, 2002; Larsson and Forlin 2002). To estimate fecundity, the gonad must be preserved and the number of live and dead embryos counted. Proper sampling requires some preliminary data on spawning time and gonadal development so that sampling procedures can be optimized.

3.4 Effect Indicators

The effect indicators for the various types of study designs for the fish survey are listed in Table 3-3. For a much more detailed discussion on these topics, consult Munkittrick et al. (2009), where the authors re-emphasize the original purpose of the EEM program and discuss why the current EEM effect indicators are used in place of other levels of monitoring. Additional issues raised and addressed by Munkittrick et al. (2009) include the influence of natural variability (i.e., the tendency for parameter values to change from year to year, or potentially from site to site), genetic adaptation, and four important statistical design issues (site selection, pseudo-replication, power analysis, and concern over the number of comparisons made).

The EEM program focuses on parameters measurable in groups of individuals, for several reasons:

- The approach offers a compromise between the sensitivity and reversibility of biochemical approaches, and the relevance of community-level parameters.
- Monitoring at the community level will miss reversible, important effects at the population level.
- Changes to fish growth, reproduction, condition or survival puts fish at risk, and therefore, focusing on these population level parameters addresses the overall objective of the *Fisheries Act*, which is to protect fisheries resources.
- Knowing this level of risk is important to the management of ecosystems.

3.4.1 Lethal Sampling

In answer to the question “have fish been modified by the effluent?” effects on growth, reproduction, condition and survival of the fish population are examined. The program recommendation for the fish survey is that key indicators be measured in both sexes of adults of two species of fish. The precision for the measurements is listed in Table 3-1. The intention is to obtain estimates of age or size distribution, how well fish are using available energy for growth and reproduction, and the storage of energy as reserves. The required numbers of samples can be calculated from a statistical equation using the standard deviation (SD) of gonad sizes for the species and site (from previous samples), and a critical effect size (CES) of 25% (see section 3.7.1). The minimum sample size recommended for a lethal fish survey when there are insufficient data to calculate sample size by power analysis is 20 sexually mature males and 20 sexually mature females of 2 fish species, in each sampling area. The rationale for using 20 fish of each sex for lethal sampling is that there is little change in the 95% confidence limits with increasing sample size beyond 20 fish. For example, Munkittrick (1992) found that there was little improvement in White Sucker variance estimates with a sample size above 16.

When there is background information available, it should be used to calculate adequate sample-size requirements prior to conducting the fish survey. It is important that sample size and variability be examined early in the study design phase so that the study can be

redesigned if the variability estimates are sufficiently high for the survey not to achieve adequate power. Fish surveys benefit most by decreasing variability. When variability is so high that sample sizes are not justifiable or cost-effective, the first consideration should be to redesign the study to a) reduce variability, b) select alternative species that may be less variable, or c) consider an alternative method.

It is strongly recommended that sampled fish be processed and sexed immediately in the field on sample days to ensure the collection of fish with an equal sex ratio. Subsequent sexing of the fish in the lab using frozen samples may show a skewed sex ratio if it is assumed that fish sampled in the field displayed a 1:1 sex ratio.

It is important to identify immature fish (fish not developing to spawn) so that they can be excluded from the statistical analysis. There are three situations where gonadal development of fish is not uniform: a) situations with multiple spawning species where spawning is not synchronized; b) multiple spawning species where the number of spawns per year is influenced by fish size or age; and c) in northern populations, where fish may not acquire sufficient energy reserves to spawn each year. In all cases, fish should be analyzed within a group: comparisons should be conducted between fish developing to spawn and fish that are not. As well, the proportion of fish in each category can be analyzed. In situations where the existence of two or more groups is known before sampling, it may be possible to separate fish into categories during sampling based on condition or fish size.

The EEM program operates in an iterative fashion, so it is not necessary to develop a full assessment of the fish populations in a single sample, and the measurements are meant to act as surrogates to assist in the development of an assessment over more than one phase. Any effects in the fish survey must be confirmed in a subsequent phase, and be assessed against the CESs before studies progress (CESs are discussed in Chapter 1). While the measurements listed below are the required measurements, it may be necessary to provide alternative measurements due to site-specific or species-specific issues.

Table 3-1: Required fish survey measurements, expected precision and summary statistics

Measurement Requirement (MMER Schedule 5, s. 16 (a) and (b))	Expected Precision***	Reporting of Summary Statistics (MMER Schedule 5 s. 16) and other general reporting
Length (fork or total or standard)*	+/- 1 mm	Mean, median, SD, standard error, minimum and maximum values for sampling areas
Total body weight (fresh)	+/- 1.0%	Mean, median, SD, standard error, minimum and maximum values for sampling areas
Age	+/- 1 year (10% to be independently confirmed)	Mean, median, SD, standard error, minimum and maximum values for sampling areas
Gonad weight (if fish are	+/- 0.1 g for	Mean, median, SD, standard error,

Measurement Requirement (MMER Schedule 5, s. 16 (a) and (b))	Expected Precision***	Reporting of Summary Statistics (MMER Schedule 5 s. 16) and other general reporting
sexually mature)	large-bodied fish species and 0.001 g for small-bodied fish species	minimum and maximum values for sampling areas
Egg size (if fish are sexually mature)	+/- 0.001 g	Weight, (recommended minimum sub-sample sizes of 100 eggs), mean, median, standard error, minimum and maximum values for sampling areas
Fecundity** (if fish are sexually mature)	+/- 1.0%	Total number of eggs per female, mean, median, standard error, minimum and maximum for sampling areas
Weight of liver or hepatopancreas	+/- 0.1 g for large-bodied fish species and 0.001 g for small-bodied fish species	Mean, median, SD, standard error, minimum and maximum values for sampling areas
Abnormalities	N/A	Presence of any lesions, tumours, parasites, or other abnormalities
Sex	N/A	

* If caudal fin is forked, use fork length (from the anterior-most part to the fork of the tail). Otherwise, use total length, and report type of length measurement conducted for each species. In cases where fin erosion is prevalent, standard length should be used.

** Fecundity can be calculated by dividing total ovary weight by weight of individual eggs. Individual egg weight can be estimated by counting the number of eggs in a sub-sample. The sub-sample should contain at least 100 eggs.

*** For small-size fish weights, use at least a 3-decimal scale.

3.4.1.1 Survival

Mean age is meant to give an assessment of the relative ages of the reference and exposed populations. If size-selective gear such as gillnets are used, and there is a significant difference in mean ages of fish sampled at both sites with identical gear, the difference indicates a need to further investigate the population and the reason for the difference in subsequent phases. More detailed information can be obtained through age distributions (or size distributions if aging is not possible), if adequate sample sizes are available or if aging is not possible. Furthermore, since many fish species have short life spans (< 4 years), it may be necessary to obtain immature fish and juveniles in order to conduct an appropriate assessment of this effect indicator. It is also very difficult to obtain a 25% difference in age when species are short-lived, and it may be possible to substitute a

difference in average size (length) of 25% as a surrogate for age when species are short-lived.

A list of appropriate aging structures for a variety of potential sentinel species is provided in Table 3-2. In addition, there are many references that can be referred to for aging methods (e.g., Mackay et al. 1990). Methods of aging should be consistent at each sampling area and among phases, and appropriate quality assurance / quality control (QA/QC) procedures should be followed (e.g., independent confirmation). It is recommended that all aging structures be archived for future reference. If fish cannot be aged reliably or if it is not cost- or time-effective, the age can be determined by using size-frequency distributions. This may be especially useful when sampling small-bodied fish species or when conducting non-lethal sampling. It may also be possible to confirm the size-frequency distributions by aging representative sub-samples from each size class. For more information on size-frequency distributions, consult Nielsen and Johnson (1983).

3.4.1.2 Energy Use (Growth and Reproduction)

Growth and reproduction measures give an assessment of the ability of fish to use the food available to them. Growth is the change in size (weight or length) with time or age. In the case of growth, it may be helpful to collect information on other age classes, such as whether there are changes in growth of early life stages. This will assist in determining the magnitude of the effect. Subsequent phases should focus on confirming responses detected and examining the relevance of the changes to other size classes and species.

Reproduction is expressed as reproductive effort, fecundity, egg weight or gonad weight relative to body size. Reproduction may be the most sensitive measurement in resident fish. Changes in reproductive investment can be evident within a year, because the reproductive tissue is generally turned over annually. Fecundity and gonad weight are easy to measure if an appropriate sampling time is chosen. Confirmed changes in gonad size could lead to additional work related to magnitude, such as determining whether the change occurs at other times of the year (for multiple spawners) or whether the changes are present in other species in the same area.

Table 3-2: Suggested aging structures for Canadian fish species

Structure	Family (common name/species)	Comments
Dorsal spine	Squalidae (Dogfish Shark)	
Dorsal spines or scales	Percidae (Yellow Perch)	Spines more precise for older fish
Otoliths	Anguillidae (freshwater eel), Atherinidae (silverside), Batrachoididae (toadfish), Carangidae (jacks), Clupeidae (herring), Haemulidae (grunt), Gasterosteidae	

Structure	Family (common name/species)	Comments
	(stickleback), Percopsidae (Trout-perch), Cottidae (sculpin)	
	Gadidae (codfish, Burbot)	Preferred; pectoral fin rays are difficult to age
Otoliths, fin ray	Scombridae (mackerel)	
Otoliths, first four marginal pectoral fin rays, scales	Coregoninae (whitefish)	
Otoliths, pectoral fin ray	Acipenseridae (sturgeon)	
Otoliths, pectoral fin rays, dorsal spines or scales	Percidae (Walleye, Sauger)	Scales preferred for fast-growing populations or < 40 cm; otoliths or spines for fish > 40 cm (or > 8 years of age), especially slow-growing populations
Otoliths, pectoral fin rays, or scales	Catostomidae (all sucker species), Coregoninae (cisco), Cyprinidae (minnow), Salmonidae (trout, char), Sciaenidae (drum)	Need fin rays for very old suckers, only otoliths will work for Golden Shiner, otoliths for every drum
Otoliths, scale	Bothidae (lefteye flounder), Pleuronectidae (righteye flounder)	
Pectoral fin rays, scales	Esocidae (Northern Pike, Muskellunge)	Scales are appropriate but fin rays have a higher confidence; cleithra are appropriate sometimes
Pectoral spine	Ictaluridae (catfish)	
Scales	Centrarchidae (sunfish, bass), Cichlidae (cichlid), Cyprinodontidae (killifish), Hiodontidae (Goldeye and mooneye), Mugilidae (mullet), Percichthyidae (temperate bass), Serranidae (sea bass), Sparidae (porgie)	Need fin rays for very old specimens
Vertebrae, fin ray	Lophiidae (goosefish)	
Vertebral centrum	Rajidae (skate)	

3.4.1.3 Energy Storage (Condition)

Measures of energy reserves provide valuable information on the availability and quality of food to the fish. The EEM program uses condition (body-length-to-body-weight relationships) and liver size as indicators of energy reserves. As with other indicators, the consistency in response between indicators is important. Liver size can increase for several reasons, including storage of lipids and glycogen and enhanced detoxification activity.

3.4.1.4 Abnormalities

During the fish survey, a visual examination of fish is also conducted in order to identify the presence of any internal or external abnormalities, such as of body form, body surface, fins, eyes, lesions, tumours, neoplasms, scars or other abnormalities such as eroded, frayed or hemorrhagic fins, internal lesions, abnormal growths, parasites, and any other unusual observations. An area on the data sheet should also be included for other significant observations. Photographs can be a useful tool to document any obvious abnormalities.

It is recommended that a rough illustration of the selected fish species be incorporated into the data collection sheet for the recording of abnormalities in the external appearance. This information can then be used by others at a later date if significant differences exist between reference and exposure areas.

More information on fish anatomy can be found in general fish biology textbooks. Instructions on tumour descriptions are available in *Gross Signs of Tumors in Great Lakes Fish: A Manual for Field Biologists* (www.glfc.org/tumor/tumor1.htm).

Table 3-3: Fish survey effect indicators and endpoints for various study designs

Effect Indicators	Lethal Effect and Supporting Endpoints	Non-lethal Effect and Supporting Endpoints	Sentinel Mollusc Effect and Supporting Endpoints
Survival	*Age *Age-frequency distribution Length-frequency distribution	*Length-frequency distribution Age-frequency distribution (if possible)	*Length-frequency analysis
Growth	*Size-at-age (body weight at age) Length-at-age	*Length of YOY (age 0) at end of growth period *Weight of YOY (age 0) at end of growth period Size of the 1+ fish Size at age (if possible)	Whole animal wet weight Shell length and width Soft tissue fresh weight
Reproduction	*Gonad weight at body weight Gonad weight at length Fecundity (number of eggs/female at body weight, length, and/or age)	*Relative abundance of YOY (% composition of YOY) YOY survival	*Gonad weight at body weight (gonadosomatic index [GSI]) (bivalves only)
Condition	*Body weight at length *Liver size at body weight Liver weight at length Egg weight at body weight and/or age (mature females only)	*Body weight at length	*Weight (whole animal dry weight, dry shell or soft tissue weight) related to shell length Soft tissue weight related to shell weight Soft tissue weight related to

Effect Indicators	Lethal Effect and Supporting Endpoints	Non-lethal Effect and Supporting Endpoints	Sentinel Mollusc Effect and Supporting Endpoints
			shell volume

*Fish survey effect endpoints used for determining effects as designated by statistically significant differences between exposure and reference streams. Other supporting endpoints can be used to support analyses.

3.4.2 Non-lethal Sampling

Non-lethal sampling should only be used in situations where it is warranted, i.e., where there is a concern about the potential impacts of sampling on small fish populations. Lethal sampling of adults is preferred where possible, although information on non-lethal samples can be valuable when large numbers of fish are collected during the sampling procedures. The indicators used for non-lethal sampling are contained in Table 3-3, and additional information on statistical analysis for the non-lethal sampling is contained in Chapter 8.

If the only option for a facility is to do a non-lethal sampling of fish in order to evaluate the effects of effluents on the fish population at a facility, a minimum of 100 fish older than YOY is recommended from each study site. The YOY acquired during the collection for the 100 non-YOY fish should also be retained and sampled (measured). YOY can usually be separated from older age-classes by size distributions; however, this may not be possible for species with extended spawning periods. The proportion of fish that are YOY should be estimated from the first 100 fish collected. If YOY abundance is extremely high (> 80-90%), sampling should then continue until 100 non-YOY are captured to calculate size-distributions of older fish. The collection of the additional non-YOY fish allows for a higher discrimination of the older fish classes to be achieved. The fish older than YOY that are collected should represent the whole range of fish sizes and be representative of the population (mature and immature). The recommended sample sizes in each area will give a good idea of the population distribution when plotting parameters such as the length or weight frequency. As well, when examining differences between the relative abundance of young versus mature fish, fairly good resolution is achieved (Gray et al. 2002).

When possible, sampling should be conducted when YOY are a catchable size in the gear being used. The same sampling gear should be used in both the exposure and reference areas; if it is not possible to use the same gear, or multiple gears must be used, the size distributions within a site should be compared between gears. If there are differences in the sizes of fish collected with different gear, comparisons between sites should be restricted within gear type. The sampling techniques and relative effort should be the same in all sampling areas. Pooling of data from different fish-sampling techniques should be avoided, and all methods used should be reported. If more than one gear type is used, the records of fish caught by each method should be reported, and any pooling of data clearly described. Fish should be measured for length (± 1 mm), weight (± 0.01 g) (Gray and Munkittrick 2005), assessed for the presence of abnormalities, and external sex determination should be made, if possible. All fish should then be released. If possible, a

small number of larger fish should be sacrificed to verify ages of older individuals. If only adults are used, the priority should be to sample prior to or at the start of the spawning season (see guidance on preferred sampling times in Table 3-5). However, if YOY are to be collected, the timing should move to the late fall, when it will be easier to measure YOY for most species. Fall sampling of YOY will be much more difficult if the fish are not single, synchronous spring spawners, as the size distributions of YOY fish will be broad.

A large number of areas can typically be sampled when conducting a non-lethal survey and the facility is encouraged to sample multiple exposure and reference areas. Programs that sample adults and YOY will allow for maximal assessment of effect indicators.

Species selection for non-lethal sampling can be difficult and is often based on availability. When choices are available, a synchronous spring spawner will offer the most advantages in terms of differentiating YOY from older year-classes. Discrimination of year-classes can also be affected by the longevity of the species. An annual species such as silverside will have a single year-class, eliminating the need to differentiate year-classes. A short-lived species (2-3 years) with fast growth and easily distinguished year-classes also offers advantages. However, these species are not always available.

When multiple species are available to choose from, it is recommended to collect initial samples and examine the ability to discriminate YOY and age-classes between species.

3.4.2.1 Survival (Size Distributions)

There are challenges to using age information on many short-lived species of fish. If a fish only lives 2-3 years, it will not be possible to measure a 25% difference in mean age. If non-lethal aging structures have not been validated for the sentinel species being used, size-distribution should be examined as a surrogate for differences in age.

Size distributions should be compared between exposure and reference areas with the Kolmogorov-Smirnov test, although this test is not very sensitive. Size comparisons should also examine distributions for YOY alone, for both sizes combined. If a site difference is present, subsequent phases should focus on understanding the difference and possible causes. When possible, verifying the ages of larger fish and YOY can be useful.

3.4.2.2 Energy Use

It should be possible at most sites to get estimates of growth and reproduction using non-lethal methods. Growth can be evaluated by the size of YOY at the end of the growing season and by the size of the 1+ fish. A comparison of the size of YOY fish between sites gives a good indicator of growth, as it is a direct indicator, in comparison to size-at-age, which is indirect. YOY are used because all of their growth is attributable to environmental conditions since the spawning time, and growth is not complicated by diverging energy into reproductive development. Differences between sites in spawning

times will be integrated into this analysis. It is also possible to get a growth estimate by a shift in size distributions over time (e.g., repeating measurements 2 months apart at the same sites), or differences in average size (this would require a second sampling trip to determine). If the fish species chosen is externally sexually dimorphic, it is possible to examine whether there are gender-specific differences in growth rate.

Reproduction can be assessed using relative age-class strength or by the relative abundance of YOY individuals (Gray et al. 2002) or by YOY survival, which requires two sampling periods. A length-frequency distribution may be plotted as a surrogate of an age-frequency distribution. Size-frequency analysis can be used to examine size distributions and distributions of condition factors (using length and weight data), and can be used to infer age distributions and size-at-age data (if ages can be inferred) (Gray et al. 2002). It is recommended that, if possible, aging structures be collected from a sub-sample of each size-class, for situations where age may need to be verified (as in section 3.4.1.1, the utility of the age information is reduced in situations where the species is short-lived). In Slimy Sculpin, rapid growth of YOY fish in the spring can cause some overlap with the 1+ age-class, making resolution difficult (Gray et al. 2002). The ability to discriminate the YOY will depend on the duration of the spawning season, and the amount of time elapsed between the spawning time and sampling time. It may be easiest (for spring and early-summer spawning species) to examine length-frequency distributions using late summer and early fall data, when the YOY should be easiest to distinguish. To test for differences in relative abundance of YOY between the exposure and reference areas, a Kolmogorov-Smirnov test can be performed on length-frequency distributions with and without the YOY included. If inclusion of the YOY changes the interpretation of the significance of the difference (i.e., it is different with them included, and not different without them), there is then a difference in the relative abundance of YOY. Alternatively, replicate areas can be sampled to allow for the use of more statistical approaches, or the proportions of YOY can be tested using a Chi-squared test.

It may not be possible to distinguish YOY in species that spawn multiple times, in northern areas where YOY may emerge later in the year, or in situations where there are habitat-preference differences that are age-dependent in a species. In those cases it will not be possible to easily infer potential reproductive impacts. Some professional judgement will be required. If the species lives multiple years and immature fish can be distinguished non-lethally (condition near spawning time can be used in many situations for this), the proportion of immature fish can be used as a substitute. In cases where this is not possible, interpretation will need to be made based on size distributions alone, and care must be exercised to be conscious of the potential impacts of adult mortality on interpretation.

It is important to remember that a difference in water temperature between sites will affect spawning time. End-of-summer differences in size distributions could as easily result from differences in spawning time due to temperature as other potential causes. If there are temperature differences between sites that are suspected to be a major cause in the differences in size distribution observed, then subsequent studies should determine whether

these site differences are a consequence of the facility or an inadequacy in choosing reference sites.

If it is possible to make multiple sampling trips, it may also be possible to measure changes in condition before and after spawning as an indicator of reproductive investment. For some small-bodied species, spawning females are very easy to distinguish by condition factor. Differences in condition factor of females between sites before spawning, or an indication of the change before and after spawning in females, could be used to infer reproductive investment, if females can be distinguished after spawning.

3.4.2.3 Condition

Condition factor (k) can be evaluated by the relationship ($k = 100\,000 * (wt/l^3)$) of the fish examined (where wt = weight [in grams] and l = length [in mm]). The appropriate analysis for final interpretation is an analysis of covariance (ANCOVA) of weight versus length, by site.

3.4.3 Wild Molluscs

Where there are no appropriate finfish present, collection of wild molluscs, such as oysters or mussels, may be considered. Shellfish are included under the definition of fish in the *Fisheries Act* and they have been used by some pulp and paper mills in the EEM program. However, there are some drawbacks, including difficulties in aging individuals and in estimating reproductive investment in some species. Crabs and lobsters are not suitable species because they cannot be reliably aged at the present time (Environment Canada 1997). Currently, guidance is available on the relative gonad index (mantle somatic index) for bivalves (see mesocosm guidance in Chapter 9).

Molluscs are a diverse taxonomic group that include bivalves and gastropods, and are widely distributed throughout Canada. Molluscs possess many qualities that a species for monitoring should exhibit:

- they are relatively sedentary, although some species (i.e., unionids) may migrate short distances (metres) within their habitat;
- they are widely distributed across Canada and are identified with limited taxonomic expertise;
- most unionid bivalves are large enough to provide sufficient tissue for analyses;
- several bivalve species have been shown to readily accumulate many chemicals from a variety of pathways (water, sediment, food) and show sublethal effects associated with exposure; and
- bivalve growth is relatively easy to measure and has been shown to be as sensitive or more sensitive than mortality in other standard assays on species such as *Daphnia*, Fathead Minnow and Rainbow Trout (see Salazar and Salazar 2001).

In general, reproductive periods for molluscs and patterns of abundance are related to climate and the abundance of food supply. For most freshwater lotic or lentic habitat types, sampling is best conducted during the fall when the majority of taxa will be present and/or are large enough to be easily collected. In marine environments, sampling should be conducted in late summer or fall, as populations with spring recruits have stabilized by this time.

3.5 Timing of Sampling

A variety of factors need to be considered when deciding on a time to sample, including potential migratory behaviour of the sentinel species, water conditions (e.g., flow, turbidity, wave action), accessibility, and the cycle of gonadal development for the sentinel species. Where historical data exist, it would be useful to examine the data and, if appropriate, conduct the survey during similar periods so that the surveys can be compared.

The timing of sampling should be synchronized with the development of sufficient gonadal tissue so that effects on the reproductive function can be assessed. However, such information is unavailable for many species of fish. Species for which there exists extensive background information on their biology and life history characteristics should be preferred as sentinel species in order to ensure that sampling can be synchronized with sufficient gonadal tissue development.

Recent research has been conducted to evaluate the optimal timing for interpreting gonadal development, using seasonal collections from a variety of species. Five types of fish categorized by spawning characteristics have been identified, and Table 3-4 provides the recommended sampling time based on the following background collection studies: background collections followed Canadian freshwater species that were synchronous spawners (such as Slimy Sculpin; Gray et al. 2005; Brasfield 2007), multiple spawners with few spawnings per year (such as Blacknose Dace [*Rhinichthys atratulus*]; Galloway and Munkittrick 2006; Hicks and Munkittrick, unpublished data), multiple spawners with many spawns per season (such as Redbelly Dace [*Chrosomus eos*]; Carroll 2007), and asynchronous spawners (every few days, such as Mummichog [*Fundulus heteroclitus*; McMullin et al. 2009]). There is a fifth type of freshwater species that has asynchronous development, where individuals may take a year off from spawning because of cold temperatures or low food availability. This variability has a major impact on power and sample size requirements.

Examination of these data confirms that there are specific times when power is higher for detecting differences, and when gonadal development is adequate for detecting impacts. The generalizations in Table 3-4 may not apply to all species or all regions; the regional EEM contact should be consulted for any available updates to regional guidance.

Synchronous spawners show a difference in timing of gonadal development between

males and females. For synchronous spring spawners, adequate data can usually be obtained as late as possible in the fall, or prior to spawning in the spring. If previous data are available for a site, the reproductive strategy can usually be estimated from the magnitude of the correlation coefficient (R^2) between gonad weight and body weight, if the previous collections were done at a time when the gonads were well developed.

Table 3-4: Generalizations and suggested optimal sampling times for fish species in EEM

Reproduction Type	Sample Time	R^2 for Gonad Weight vs. Body Weight Relationship for Reference-site Females
Synchronous spawners	Late fall (if spring spawner) Early summer to mid-summer (if fall spawner)	> 0.85
Multiple spawners, few spawns	4-6 weeks before first spawn (usually April to early May)	$0.4 < R^2 < 0.8$
Multiple spawners, many spawns	As close to start of spawning as possible	< 0.4
Asynchronous spawning	After spawning has started or near start of spawning period	Not significant
Asynchronous development (year off)	Separate groups and treat independently	Two groups of fish seen with different slopes within a site

Evaluation in multiple-spawning species is complicated by the duration of the spawning period. In the case of such species, the frequency distribution of age- or length-classes may provide valuable complementary information on the reproductive success.

Multiple spawners with few spawns should be sampled at least 6 weeks prior to the initiation of the spawning season (for information on spawning temperatures, consult references such as Scott 1967; Scott and Crossman 1973; Roberts 1988; Nelson and Paetz 1992; Jenkins and Burkhead 1993; Coad 1995) due to an increased variability in the gonad-weight-to-body-weight relationship as the spawning season approaches, because of a lack of synchronization in timing for the second clutch of eggs (Galloway and Munkittrick 2006). Multiple spawners with many spawns, and asynchronous spawners, should be sampled close to the start of the spawning period because of the rapid development of the gonads in both species.

The consequences of sampling at an inappropriate time have been examined using data from the pulp and paper EEM program (cycles 1 to 3). For large-bodied species, fish were sampled outside of the optimal window in more than 33% of previous studies, but interpretation was not strongly affected when optimal and suboptimal studies were compared. However, small-bodied species were sampled at suboptimal times more than 75% of the time, and data collected outside of the optimal windows failed to detect significant effects on gonad or liver size (Barrett and Munkittrick, unpublished data).

Table 3-5: Fish species commonly used in EEM, aspects to consider during study design, and recommended sampling times (based on Barrett and Munkittrick 2010)

Family	Species	Scientific Name	Considerations	Sample
Salmonidae	Lake Trout	<i>Salvelinus namaycush</i>	F M*	4-6
	Brook Trout	<i>Salvelinus fontinalis</i>	F M A	4-6
	Arctic Charr	<i>Salvelinus alpinus</i>	F M D	4-6*
	Dolly Varden	<i>Salvelinus malma</i>	F M?	4-6
	Bull Trout	<i>Salvelinus confluentus</i>	F M?	4-6
	Cutthroat Trout	<i>Salmo clarki</i>	W M* A	LF
	Rainbow Trout	<i>Oncorhynchus mykiss</i>	S (F) M A	LF*
	Arctic Grayling	<i>Thymallus arcticus</i>	S M A	LF
	Mountain Whitefish	<i>Prosopium williamsoni</i>	F M?	4-6
	Round Whitefish	<i>Prosopium cylindaceum</i>	F/W M?	4-6
	Lake Whitefish	<i>Coregonus clupeaformis</i>	F M*	4-6
	Cisco	<i>Coregonus artedii</i>	F	4-6
Hiodontidae	Goldeye	<i>Hiodon alosoides</i>	S M P H	SPAW
	Mooneye	<i>Hiodon tergisus</i>	S M?	LF
Esocidae	Northern Pike	<i>Esox lucius</i>	S M	LF
Cyprinidae	Carp	<i>Cyprinus carpio</i>	X M*	4-6
	Fallfish	<i>Semotilus corporalis</i>	S X? G?	LF
	Creek Chub	<i>Semotilus atromaculatus</i>	S X*	4-6
	Peamouth Chub	<i>Mylocheilus caurinus</i>	S	LF
	Lake Chub	<i>Couesius plumbeus</i>	S M	LF
	Longnose Dace	<i>Rhinichthys cataractae</i>	X	4-6
	Blacknose Dace	<i>Rhinichthys atratulus</i>	X	4-6
	Pearl Dace	<i>Margariscus margarita</i>	X?	4-6
	Redbelly Dace	<i>Chrosomus eos</i>	A	SPAW
	Spottail Shiner	<i>Notropis hudsonius</i>	X? M?	4-6
	Mimic Shiner	<i>Notropis volucellus</i>	X? 2 yr?	4-6
	Emerald Shiner	<i>Notropis atherinoides</i>	X? M?	4-6
	Blacknose Shiner	<i>Notropis heterolepis</i>	X	4-6
	Common Shiner	<i>Luxilus cornutus</i>	S?	4-6
	Golden Shiner	<i>Notemigonus crysoleucas</i>	X?	4-6
	Redside Shiner	<i>Richardsonius balteatus</i>	A?	SPAW
	Bluntnose Minnow	<i>Pimephales notatus</i>	S G	4-6
	Fathead Minnow	<i>Pimephales promelas</i>	A G	4-6
Catostomidae	White Sucker	<i>Catostomus commersoni</i>	S M	LF
	Longnose Sucker	<i>Catostomus catostomus</i>	S M	LF
	Largescale Sucker	<i>Catostomus macrocheilus</i>	S M?	LF
	Bridgelip Sucker	<i>Catostomus columbianus</i>	S M?	LF
	Shorthead Redhorse	<i>Moxostoma</i>	S M	LF
	Sucker	<i>macrolepidotum</i>		
	Silver Redhorse Sucker	<i>Moxostoma anisurum</i>	S M?	LF
Ictaluridae	Brown Bullhead	<i>Ameiurus nebulosus</i>	S G	4-6
	Channel Catfish	<i>Ictalurus punctatus</i>	S G D	4-6

Family	Species	Scientific Name	Considerations	Sample
Fundulidae	Mummichog	<i>Fundulus heteroclitus</i>	A R	SPAW
Gadidae	Burbot	<i>Lota lota</i>	W M? H	LF
Atherinidae	Atlantic Silverside	<i>Menidia menidia</i>	S 1 yr.	SPAW
Gasterosteidae	Brook Stickleback	<i>Gasterosteus acconstans</i>	X	SPAW*
	Threespine Stickleback	<i>Gasterosteus aculeatus</i>	X M*	SPAW*
	Ninespine Stickleback	<i>Pungitius pungitius</i>	X M*	SPAW*
Percopsidae	Trout-perch	<i>Percopsis omiscomaycus</i>	X M*	4-6
Centrarchidae	Rock Bass	<i>Ambloplites rupestris</i>	S G	4-6
	Pumpkinseed Sunfish	<i>Lepomis gibbosus</i>	S G	4-6
	Smallmouth Bass	<i>Micropterus salmoides</i>	S G	4-6
Percidae	Walleye	<i>Sanders vitreus</i>	S M	LF
	Yellow Perch	<i>Perca flavescens</i>	S M* H D	LF
	Iowa Darter	<i>Etheostoma exile</i>	S G M*	4-6
	Johnny Darter	<i>Etheostoma nigrum</i>	S G M*	4-6
	Logperch	<i>Percina caproides</i>	S M*	4-6
Cottidae	Mottled Sculpin	<i>Cottus bairdii</i>	S G LM	LF
	Slimy Sculpin	<i>Cottus cognatus</i>	S G R LM	LF
	Torrent Sculpin	<i>Cottus rhotheus</i>	S G LM	LF
	Spoonhead Sculpin	<i>Cottus ricei</i>	S G LM	LF
	Shorthorn Sculpin	<i>Myoxocephalus scorpius</i>	W G	LF

Note: The best designs for sampling fish fall under three categories: a) close to the start of spawning (SPAW); b) late fall (LF) or about 6 months prior to spawning; and c) 4–6 weeks prior to first spawn (late April/early May).

Spawning

S - spring or early summer spawners, usually a single spawning time, usually migrate to spawn

F - fall spawners, usually single spawn

W - winter spawners, usually single spawn

X - multiple spawners, usually summer spawners: assess whether small fish are reproducing less (or more) than larger fish (view GSI vs. length); variability will increase as spawning season approaches; best to sample 4–6 weeks prior to first spawn

G - males are guarders; reproductive investment in males will be low, but they are likely not very mobile

A - asynchronous spawners

? - some doubt about number of spawns: either multiple or asynchronous

Migration

M - mobile; are known to migrate considerable distances (or change habitats significantly) for spawning at some sites

* - may be site-specific

M? - probably migrate

A - anadromous forms may be present, and may be much larger and much more mobile

C - catadromous; migrate out as adults

R - research available to show that they are usually resident

Other

P - plankton feeder; migratory feeder

1 yr - 1-year life cycle

2 yr - 2-year life cycle

H - juvenile and mature fish may utilize very different habitats

LM - females may start to mature very late in the year, fish may spawn during freshet; can be difficult to obtain mature female gonads

D - diet changes within the size of sexual maturity, to piscivorous at larger sizes (when available)

3.6 Verification of Fish Exposure

It is crucial that studies be designed to maximize the possibility of detecting effects if they are present. This can be accomplished by sampling at the proper time of year, with appropriate gear, at appropriate reference areas and during the period of residence in the effluent area. If fish exposure to the mine effluent is uncertain, redesigning the survey (selecting different species, using tracers, changing sampling time or changing exposure or reference areas) or using alternative monitoring methods should be considered for the subsequent phase.

Controversy arises when fish show no differences in characteristics among sites, and there are no indicators of exposure. In this case, it is difficult to determine whether the fish at both sites belong to the same population. In order to verify the exposure of fish to effluent in the exposure areas, and to verify the lack of exposure at reference areas, it may be necessary to select a tracer which accumulates in fish tissue. The selection of a tracer depends on the type of mine involved and the complexity of the receiving environment.

It is possible to infer exposure by examining metal levels in indicator tissues. The indicators and the tissues will vary with the mine type and species being used. In general, gills, liver and kidney have the greatest potential for estimating exposure and bioavailability of metals. Mercury is the only metal element of concern that has been found to accumulate in muscle tissue, so if mercury is a contaminant of concern, dorsal muscle tissue should be analyzed. Blood and bone tissue may reflect exposure to lead, and might be considered if lead is the primary element of concern (Hodson et al. 1984). Bone concentrations are expected to be most indicative of long-term metal exposure, while blood concentrations are indicative of short-term exposure (AETE 1998). For larger species, samples of liver or kidney can be collected. The tissues should be frozen for later preparation and analysis. For small species (< 10 cm), whole body levels can be examined, or levels in the carcass after removing the digestive tract. See section 3.11 for fish usability methods.

Large statistical differences between sites in whole-organism characteristics in a number of parameters give some confidence that the samples are from different populations of

fish. If there are no differences between sites, it may be that fish are moving or that there is no impact. Stable isotopes of carbon and nitrogen can be used to document that there are differences in fish residence times, provided that the stressors in question locally alter stable isotopes (i.e., Farwell 1999; Galloway et al. 2004), or there are local geochemical differences that alter stable isotopic signatures and that can be used to demonstrate local residency (i.e., Gray et al. 2004). However, the stable isotopes are not always sufficiently different between sites to be useful, and their suitability has to be evaluated on a site-specific basis (Dubé et al. 2006).

By selecting a sampling time and fish species that have life history habits that may increase the likelihood of exposure, potential exposure can be maximized. For example, for species having spawning movements that take them away from or temporarily into the effluent exposure area, a survey conducted during the spawning season would be ineffective. Thus, for spring-spawning freshwater species, a fall survey would be appropriate. For fall spawners, a spring or summer survey is appropriate. This may not apply to fish in which ova mature rapidly; for example, as some late-spring-spawning cyprinids should be sampled in early spring, rather than in fall when ova may still be immature, it is pertinent to have some background biological information, if possible.

The timing of sampling and the choice of fish species should be made according to normal operation of the facility to ensure that the effluent is present in the environment. Sampling when effluent has not been discharged for long periods (months) should be avoided. However, the selected sampling gear, flow conditions and effluent conditions may limit the preferred season for the survey.

If no fish are captured (or they are captured in reduced density) and there are no fish resident in the exposure area, it could be interpreted that fish are avoiding the exposure area. The suitability of fish species should be evaluated at the end of each monitoring phase, based on the site-specific nature of the results and the site-specific concerns about residency and exposure.

There are some situations where fish may move freely in and out of the exposure area, and no species spend significant periods of time in the effluent. In these cases, the sampling should be designed to maximize exposure time in the effluent area and possibly during periods of optimal gonadal development.

There are two main issues dealing with residency: whether the fish from reference and exposure areas were mixing; and whether the fish captured in the exposure area were indeed exposed. If fish demonstrate exposure, are collected in the exposure area, and demonstrate differences from reference-area fish, there should be no controversy. Follow-up studies can examine other species to see if they demonstrate effects.

If fish demonstrate exposure, are collected in the exposed area, and show no differences, it is outside the scope of the EEM to determine why exposure-area effects are not seen. If

subsequent monitoring phases confirm the absence of demonstrated effects and the study design was adequate, it would be concluded that the conditions of the area allow for fish that are exposed to effluent not to be affected, using the current design.

3.7 Power Analysis

The purpose of defining an effect-size and power level is to determine if the sampling program is collecting sufficient information for decisions to be made. The statistical power of a comparison is a function of the sample size, the variability and the target difference set between areas. To determine the sample size for detecting a specific difference, some knowledge is needed about the statistical power level that is acceptable for the decision-making process and the variability of the population.

3.7.1 Power and Significance Level

Earlier cycles of the pulp and paper EEM program set the power ($1-\beta$) at 0.80 and α at 0.05. The EEM program now encourages setting α and β equal to one another. If values are set at $\alpha = \beta = 0.10$, the sample sizes required to detect the same effect are approximately the same as in earlier cycles. Where possible, provided sample sizes determined by the power analysis are not unreasonably large, mines are encouraged to reduce $\alpha = \beta = 0.05$ (the traditional level for *alpha*). In many statistical programs, the default β is 0.20, and needs to be adjusted. Again, these recommendations are to help ensure that studies are designed to provide a reasonably high probability of statistically detecting a predetermined effect size if it has occurred, (i.e., the power of the test [$1-\beta$] should be high). Refer to Chapter 8 for the rationale for setting α and β at equal levels.

It is important to understand that variability and power will vary with the parameter being studied. Fish are not equally variable across all of their characteristics. Reproductive variables are usually as changeable, or even more on a relative scale, than parameters such as length, weight and liver weight (Environment Canada 1997). If effect sizes are also expressed on a relative scale (i.e., as percent differences), any study that can detect a $\pm 25\%$ difference in relative gonad size can detect similar or smaller differences in other important parameters.

3.7.2 Effect-Size

It is recommended that the EEM program be designed to detect a difference of 20-30% in gonad size, using a recommended power level of 0.90 ($1-\beta$). The magnitude of the difference that could be detected for other parameters would be fixed based on the sample size for determining an effect on gonad size. The power for detecting differences in other parameters should be reviewed during study design to ensure that reasonable power is achieved for as many variables as possible. The same approach used to identify a target effect-size for relative gonad weight should be applied to other variables. Sensitivity analyses using population models should be used to explore the consequences of the effect-size chosen for any and all variables (Environment Canada 1997).

An extensive literature review has shown that CESs that have been defined in other programs are often consistent with a CES of around 25% or 2 SDs for many biological or ecological monitoring variables. This value appears to be reasonable for use in a wide variety of monitoring programs and with a wide variety of variables (Munkittrick et al. 2009). Barnthouse et al. (1989) argue that a 10% change in variables would be societally and ecologically significant, although they were concerned primarily with laboratory toxicity tests and not field surveys. Their proposed effect-size was deliberately conservative (small) because of concerns about the uncertainty in extrapolating laboratory results to the field.

When preliminary analyses show that power will be insufficient given reasonable sample sizes, the assessments should be redesigned. Studies are designed site-specifically, and priority should be given to reducing variability rather than increasing sample size. As variability will also vary between sampling campaigns, the target effect-size should not be a fixed number, but rather should be a range of changes that you wish to detect, such as 20-30% difference. Sample sizes can be calculated using methods described in Green (1989); sample size calculators can also be downloaded from the Internet, such as the common one that can be found at <http://biostat.mc.vanderbilt.edu/wiki/Main/PowerSampleSize>.

A priori power calculations and CES calculations are described in section 8.6.2.1 of Chapter 8.

3.8 Fish Sampling Methods

Sampling methodology should be chosen site-specifically, and capture gear and effort should be focused on methods shown to be successful. The same sampling methods can be used for population and community surveys. The difference lies with the selectivity of the fishing gear. During a community survey, the gear should be as non-selective and non-destructive as possible. For the population survey, which focuses on one or two species, the gear will be more selective. For example, trap netting may be preferred during a community survey, while a one-size mesh gillnet of the appropriate size could be appropriate for a population survey.

Standardized sampling is a priority. Therefore, in situations where sentinel species are the same as for a previous phase, and the sampling techniques used previously were sufficient to capture the target number of each sentinel species, these same sampling techniques should be retained unless good reasons for change, such as unacceptable bycatch, are documented. The sampling techniques and relative effort should be the same in all sampling areas. Pooling of data from different fish-sampling techniques should be avoided, and all methods used should be reported. If more than one gear type is used, the records of fish caught by each method should be reported, and any pooling of data clearly described.

A number of good guidance documents fully describe fish collection methods (Schneider 2000; Portt et al. 2006). Portt et al. (2006) describe the use and efficacy of 1) gillnets; 2) beach seines; 3) hoop, fyke and trap nets; 4) electro-fishing; 5) underwater observation; 6) Gee or minnow traps; and 7) enclosure (drop, pop and throw) traps. However, methods will usually have to be developed and optimized site-specifically.

3.8.1 Bycatch

It may be possible to obtain samples using the bycatch of commercial, research or other fisheries operations in either marine or freshwater situations. The investigator is responsible for ensuring and documenting that sampling procedures and conditions are met (QA/QC), and that fish are exposed. Capture techniques also have to be standardized between sites.

3.8.2 Remote Sensing

Fish abundance near outfalls can be monitored using video or still cameras mounted on remotely operated vehicles. This technique may be particularly effective in rocky and steep areas where use of fishing gear may be difficult. Camera surveys may also be useful in reconnaissance surveys of bottom conditions before trawls or traps are deployed for fishing. Any proposed methods should be clearly outlined in the study design for review.

3.8.3 Alternative Methods

There may be situations where conducting the fish survey is not suitable. The reasons for this are site-specific, but the most common reasons are the presence of hazardous conditions (e.g., strong currents) or the presence of confounding factors such as other effluent discharges in the exposure area, which will make it difficult or impossible to isolate any effects attributable to the effluent being monitored. Under these circumstances, the mine may select an alternative option to the fish survey and/or the fish usability survey. Recommended alternative monitoring methods for the fish survey are mesocosm studies and caged bivalves. Detailed guidance on how to conduct the alternative monitoring methods and interpret the data can be found in Chapter 9.

3.9 Fish Survey Quality Assurance and Quality Control

3.9.1 Field Practices to Improve Data Analysis and Interpretation

The quality of data collected in the field influences the ease of data analysis and interpretation. The preparation of data recording sheets beforehand will save time in the field, and the use of waterproof paper is encouraged. Field conditions, habitat, gear used and information for catch-per-effort calculations should be recorded. The use of the same balance and measuring device for all measurements, and having the same person taking the measurements, will reduce measurement error. If the person taking the measurements is reporting the data to a person recording the measurements, avoid the use of decimal points and report all measurements as digits and not numbers to avoid transcription errors (e.g., report 14.5 cm as 1-4-5 and use units of mm); some numbers can be easily confused when reported orally, such as “fourteen” vs. “forty.”

It is essential that the sampling gear be consistent between the sampling areas, because most sampling methods select for certain age- or size-classes, and thus inconsistent sampling gear between sampling sites could result in detecting false differences (e.g., in age or size).

3.9.2 Quality Control in the Field

This is the first stage of data collection. QA/QC procedures for the fish survey should be outlined during the development of the study plan and should be followed precisely in order to maintain high-quality data. While a QA/QC plan for field sampling can have many components, some of the main procedures are as follows:

- initiate and maintain communication with local government agencies (e.g., fishing licence, dates of fish collection, location of collection, endangered species, etc.);
- all personnel involved in field sampling should have appropriate education and/or training and be familiar with the written standard operating procedures for the survey;
- all safety measures should be identified, understood and adhered to;
- fish collection methods and equipment should be appropriate for the specific water body and fish species;
- habitat descriptions, including possible modifying factors (water depth and current, dissolved oxygen concentration, temperature, substrate classification, evidence of pollution [discolouration, odour, residues], salinity, conductivity, etc.);
- date and time of collection;
- collection methods need to be consistent throughout the study;
- location of sampling areas and fish collection areas documented (geographic coordinates); photograph the collection location;
- record of the number of fish species and incidental species caught per collection

- stations;
- estimate of catch per unit effort;
- samples from fish (e.g., ovaries, age structures, stomach content) should be placed in appropriate containers;
- suitable preservatives/fixatives (e.g., ovaries—frozen or formalin) should be used;
- all samples should have appropriate labelling;
- all measurements will be taken using appropriate equipment of acceptable accuracy and precision (this should be documented);
- instruments should be calibrated and maintained in good working order (records and methods should be available);
- detailed field notes should be maintained in a bound notebook; and
- chain-of-custody forms and appropriate shipping and storage procedures should be used.

3.9.3 Determination of Sampling Effort

To aid in assessment of expected effort requirements at individual sites, the study plan submitted to Environment Canada should include details on how fish sampling will be performed.

The following are performance-based criteria and guidance to determine a “reasonable level of fishing effort.” Each site is unique. It is uncertain whether fishing success will be achieved at a site just because a certain level of sampling effort has been successful in the past at other sites or even at the same site at other times.

1. **The study design** should document all details on how the adult fish sampling will be performed, to aid in assessment of effort. Details to include in the study design (where applicable) are:
 - how and why the sentinel species were selected;
 - who was consulted on the locations and techniques chosen to collect the proposed sentinel species;
 - contingency plans regarding alternative gear and sentinel species;
 - scheduling of dates for work that will be performed so that EEM contacts can be available for consultation;
 - type, location(s) and dimensions of gear (e.g., gillnet, trapnet, hoopnet, fish trap, trawl; in some cases more than one type of gear may be advisable);
 - mesh type (e.g., nylon, cotton fibre or wire, knotted or knotless) and size;
 - proposed level of gear / fishing effort;
 - sampling time (i.e., time of day);
 - sampling duration (i.e., time interval between gear placement and retrieval); and
 - frequency of checks.

Any preliminary fish survey results or observations made during pre-design activities

should be provided where they have guided selection of sentinel species or procedures. The regional EEM contact will review these data and may request further information to clarify sampling procedures.

2. **Proper operating procedures** should be used. These include use of gear as outlined in the study plan. Gear should be checked at a frequency that ensures the recovery of sentinel species in useful condition and the release of non-target (especially protected and endangered) species. The use of non-lethal and/or selective techniques should be a consideration. A record of the identity and estimated numbers of non-target fish may be a useful addition to contingency plans. The mill and consultant should have a good understanding of the habitat, the characteristics of the species, and the gear being considered.
3. **Consultation with local experts** (e.g., provincial and federal fisheries personnel, Aboriginal groups, individuals and associations involved in local sport and commercial fisheries, the public, and others with knowledge of local fisheries resources) should be conducted to ascertain that the selection of sentinel species, location of nets, timing of collections, etc., are optimal.
4. **Personnel** tasked with the fish collection and sampling procedures should have documented experience.
5. **Licences** for collecting should be obtained from local fisheries agencies.
6. **Records** should be kept that document the operating procedures used (e.g., mesh size, sampling time, location, frequency of checks, etc.). These records may be required in order to properly assess the manner in which the study was conducted.

Although not required, it is recommended that an estimate of catch per unit effort (CPUE) be provided for each sampling area (e.g., number of fish caught per unit of time or area or net). The CPUE information is useful in documenting the effort expended in situations where collection of the minimum number of fish may be difficult.

3.9.3.1 Examples of Calculations of Sampling Effort

Some examples of fishing methods that have been successful in collecting fish in a timely manner are provided below. These are provided as examples to guide consultants in the development and implementation of their study design, and to indicate when it may be advisable to consult with the Environment Canada regional EEM coordinator.

1. Data from two Ontario lakes indicated that 40 individuals of any of 6 warm-water fish species were collected in 1-6 sets for 24 hours duration. The equipment included a 6-x-6-foot trap net. Details of the process are presented in the Ontario Ministry of Natural Resources Fisheries Assessment Unit Newsletter (FAU Update Issue 94-1,

OMNR 1994).

2. During the Assessment of the Abundance of Cold Waters Ontario Fish Communities Program, the fishing effort recommended to collect 40 Lake Trout in 7 lakes varied from approximately 12 to 120 hours. The equipment included a 46-m gillnet gang with 3 panels of 15.2 m. Details of the process are in the Ontario Ministry of Natural Resources Fisheries Assessment Unit Newsletter (FAU Update Issue 94-2, OMNR 1994). Mesh size should be consistent and selected according to the target species.
3. Experience has shown that gillnets made up of 4 panels of 50 m of net, set 24 hours/day for 5 days (equivalent to 24 000 metre-hours of effort) in freshwater systems, should allow for 20 fish of each sex to be collected. This is contingent on correct deployment of the panels and shifting of the panels to cover areas inhabited by the fish. Mesh size should be consistent and selected according to the target species.
4. Alternatively, a good strategy would be to initially set a minimal amount of net to decrease bycatch (< 400 m). If fishing is selective enough, and the amount of bycatch acceptable, up to 2 km of a single mesh size has been necessary in small unproductive rivers.
5. In marine situations, experience has shown that 48 hours of beam trawling, long-lining using a variety of hook sizes or other methods including traps (alone or in combination), should provide the 20 fish of each sex.
6. Consultation with users of electro-fishing technique (large rivers using boat-mounted apparatus) indicates that all fish can be obtained in one day. Procedures enhancing success include operating at dusk or night, passing over the same area at least three times, and using intermittent pulses of current (since a continuous field may actually chase fish away). In small streams, lakes and rivers, more time will often be necessary for sampling because of the difficulties encountered in moving through these environments.
7. In 1999, a consultant was collecting fish for an East Coast paper mill located on a tidal estuary. The target species were Mummichogs and the consultants used a 15-x-1.5-m beach seine with 0.5-cm mesh. One end of the beach seine was extended about 10 m out from shore by a technician wearing chest waders, while a second technician held the other end of the seine along the shore. The technicians towed the seine perpendicular to the shore for a distance of 20-30 m and then the outer end was towed into the shore to close the net. Once both ends of the seine were secured on the shore, the upper and lower lines were carefully retrieved to capture any fish enclosed in the net. The consultants found that fishing was most successful at slack high tide. A total of 108 Mummichogs were captured and retained over 4 days of sampling. Total fishing time was 12.5 hours. Many more Mummichogs were captured and

released due to the need to balance male and female numbers. In addition, 11 other species of fish were captured and released (Final Report, Repap New Brunswick Inc. Kraft Mill, Second Cycle Aquatic EEM Study, Jacques Whitford Environment Limited, April 2002).

Although the above techniques and gear may apply to a variety of species, these examples are not all-inclusive because each site is unique and the examples are provided as suggested effort only. Local expertise can serve as further advice. The previous examples are adequate guidelines toward catching a minimum of 20 fish per sex, species and area.

3.9.4 Consultation with Regional EEM Coordinators and Implementation in the Field

If all of the above criteria are met and a mine/consultant is having problems meeting the minimum data requirements of the adult fish survey, the owner or operator may deviate from the study design but is required to inform the Authorization Officer without delay of those circumstances and of how the study was or will be conducted. All reasonable efforts should be made to collect target sample sizes of two species of fish and demonstrate due diligence on the part of the mine.

Possible outcomes and options of the consultation with the EEM coordinator are as follows:

1. **Continue** – Advice on the following situations will depend on site-specific conditions. Set further consultation dates if required.

- a) Absence of target species at reference area:
 - continue with current gear and technique;
 - continue at alternative reference area identified in contingency plan;
 - continue at existing reference area with alternative target species, gear and/or technique identified in contingency plan.
- b) Absence of target species at exposure area:
 - continue with current gear and technique;
 - continue at alternative exposure area identified in contingency plan;
 - continue at existing exposure area with alternative target species, gear and/or technique identified in contingency plan.
- c) Absence of target species at reference and exposure areas:
 - continue with alternative areas, target species, gear and/or technique identified in contingency plan.

2. **Postpone (not to continue)** – Existing dangerous conditions; sampling conditions (e.g., weather, cold) will not allow collection of fish; alternative gear is not available; no further contingencies are available (e.g., no further alternative species; further investigation is needed):

- design new sampling plan in consultation with regional EEM contact;
- redeploy at a later date with original or alternative areas, target species, gear or technique, but under more favourable conditions;
- set dates for further consultation.

3. **Discontinue** – If the full complement of fish is not obtained, the absence (or paucity) of fish will be considered a result that will be thoroughly explained in the study findings, taking all possible contributing factors into account. If the minimum number of fish is not caught, this could result in inflated variance estimates. The decision on whether to continue will be influenced first by safety considerations. In all scenarios refer to the contingency plan where appropriate, and set date(s) for further discussion. Field technicians should speak directly with the regional EEM contact. Sentinel species choices will apply to both reference and exposure areas. Pooling of data from different seasons is not valid.

3.9.5 Data Entry

Data entry and preparation of analysis is discussed in Chapter 8, and reporting is discussed in Chapter 10.

3.9.6 Quality Control in the Laboratory

Although much of the survey information is collected while in the field, variables such as fecundity, egg weight, and age are usually determined later in the laboratory. With each measurement, the primary concern of the laboratory QA/QC program is to ensure consistency (precision) and accuracy of the data. The following issues should be considered as part of the measurement procedures:

- all personnel involved in sample processing and analyses should have appropriate education and/or training;
- measurements should be conducted using recognized protocols and methods (these should be documented), and all instruments used should be properly calibrated and maintained (records, methods available);
- keep fish measurements recorded for each fish (target species);
- keep a record of external lesions, tumours, parasites, etc;
- fecundity data, including methods and sub-sampling precision (if applicable);
- aging data, including methods and independent confirmation of estimates;
- maintain records that describe the sample, measurement, and responsible personnel; if possible, a minimum number of individuals should conduct a particular measurement to maintain consistency and reduce measurement error (especially for age determination);
- if sub-sampling is necessary (e.g., fecundity, egg weight), information describing the efficiency and accuracy of the sub-sampling technique should be documented; this information should also be used to calculate appropriate correction or scaling factors

- (if needed) to minimize possible differences in methods and efficiency;
- all data should be verified; for example, measurements such as fecundity and egg weight should be replicated to ensure precision and accuracy; a recognized expert should verify estimates of age;
- literature and taxonomic keys used for fish identification should be documented;
- archive samples and voucher specimens; and
- maintain detailed sample processing and laboratory notes in a bound notebook.

3.10 Data Analysis

QA/QC concerns regarding data analysis include data verification and validity, repeatability and robustness of statistical analyses, and rigour and defensibility of analyses. For the most part, validation and verification of data depends on the success of QA/QC procedures during field sampling, sample processing, and laboratory analyses (see above). However, there are other considerations regarding the data verification and analyses:

- conduct screening techniques to identify possible transcription errors, outliers and other potentially questionable data points;
- maintain tabular summaries of the general descriptive statistics (sample size, mean, minimum, maximum, standard error, and SD) of fish measurements (e.g., see Table 3-6);
- provide results of assessing assumptions of normality and homogeneity of variance;
- maintain a record of transformation used;
- provide parameter estimates of variability (analysis of variance [ANOVA] mean square error [MSE], ANCOVA MSE, SD for age-to-maturity);
- provide calculations of sample size requirements for each parameter;
- provide a summary of adherence to data quality objectives, standard operating procedures and identification of any QA/QC problems, which should incorporate considerations related to laboratory and field QA/QC;
- to allow reproduction of analyses and results, provide all raw data in an appendix and archive computer data files for an approved period of time after the analyses are published in a report;
- document in detail the methods used for analyses;
- verify that statistical software packages used produce the same output and results as other packages;
- evaluate the robustness of the analyses, (i.e., the results and conclusions should be similar);
- take note of whether outliers are included or excluded, and whether transformations are used, etc.; the objective is to ensure that results are not a function of some manipulation or assumption prior to or during analyses; and
- maintain detailed notes regarding the analyses of the survey data.

3.10.1 Statistical Analysis

The standard statistical assumptions required for many parametric statistical tests are those of independence, normality, and homogeneity of variances. These three assumptions and additional information on data assessment and interpretation are discussed in Chapter 8.

Table 3-6: Suggested reporting format for the parameters (A) and the resulting regressions (B) required for the fish survey analysis.

A. Parameter Summaries

Species	Sex	Parameter	Reference		Exposed		% Diff	Stat. Sign
			Ref (n)	Reference Mean and SD	Exp (n)	Exposed Mean and SD		

B. Regression Analyses

Species	Sex	Regression	Ref (n)	Reference Adj. Mean	Exp (n)	Exposed Adj. Mean	% Diff.	Stat. Sign.	Sign. Interax.

Note: The percentage difference should be reported as exposed relative to reference site. Statistical significance should be given as p-value.

Legend: Diff = difference, stat sign = statistical significance (p-value), ref = reference, exp = exposed, adj = adjusted, sign interax = significant interaction.

3.11 Methods for Analysing Fish Usability

The objective of the question “has there been a change in fish usability due to effluent?” is to determine whether effluent has altered fish in such a way as to limit their use by humans. Fish usability can be affected by altered appearance, altered flavour or odour, or contaminant levels that exceed consumption guidelines for human health and are statistically different from levels measured in the reference area. This section examines fish usability with respect to contaminant levels of mercury.

Mercury is the only metal for which there is a standard Health Canada tissue consumption guideline for humans, and therefore is a pollutant of national concern. Health Canada recently completed a study on mercury and reaffirmed the standard (maximum limit) of 0.5 µg/g with the exception of fresh/frozen tuna, shark, swordfish, escolar, marlin and orange roughy. Provincial and territorial governments are responsible for implementing fish consumption advisories for sport fisheries with the exception of federal parks. Consumption restrictions for sport fish begin at levels above 0.45 µg/g total mercury.

Biological monitoring studies consist of a study respecting fish tissue, if during effluent characterization conducted under paragraph 4(1)(d) a concentration of total mercury in the effluent is identified that is equal to or greater than 0.10 µg/L (MMER, Schedule 5, s. 9(c)).

An effect on fish tissue means measurements of concentrations of total mercury that exceed 0.5 µg/g wet weight in fish tissue taken in an exposure area and that are statistically different from and higher than the measurements of concentrations of total mercury in fish tissue taken in a reference area (MMER, Schedule 5 s. 1). At some mine sites there may be reference areas that have levels of total mercury in fish tissue higher than the guideline (e.g., northern Quebec – Schetagne et al. 1997; Schetagne and Verdon 1999); therefore to be considered an effect in fish tissue, there must be a statistical difference between the areas and an exceedance of the guideline using a one-tailed statistical test.

As discussed in Chapter 5, the method detection limit for mercury in effluent has been changed to 0.01 µg/L (0.00001 mg/L) so that the concentration of 0.1 µg/L specified in Schedule 5, s. 9(c) of the *Metal Mining Effluent Regulations* can be detected with confidence. Analytical methodologies suitable to achieve this level of detection include cold vapour atomic absorption spectrometry (CVAA), cold vapour atomic fluorescence spectrometry (CVAFS), and inductively coupled plasma mass spectrometry (ICP-MS).

In the *Guide to Eating Ontario Sportfish*, it is suggested that other metals—lead, copper, nickel, zinc, cadmium, magnesium, chromium, arsenic and selenium (Pb, Cu, Ni, Zn, Cd, Mg, Cr, As and Se, respectively)—may be found in fish tissue, but not at levels for consumption restrictions. On a site-specific basis these metals may be identified as a concern if there are human health consumption guidelines from another regulatory agency (e.g., provincial or territorial), that are applicable to the region where the study is being conducted **and** if the metal for which there is a consumption guideline is present in the effluent. Local consumption and commercial fisheries should be considered to determine which edible tissues (liver, kidney, bones, flesh or even entire fish) should be analyzed. It is recommended that other metals in fish tissue be analyzed where there are site-specific concerns. The *Guide to Eating Ontario Sportfish* is available at the following website: <http://www.ene.gov.on.ca/en/water/fishguide/index.php>.

Molluscs can accumulate metals (Cd, Cu, Zn, Pb, Ni, mercury [Hg], As, silver [Ag], and Cr). Field studies suggest that the relationship between mollusc tissue metal concentrations and ambient metal concentrations are influenced by a number of biological, physical and chemical parameters that need to be taken into account. Ultimately, the relationship is metal-specific and depends on the availability of the metal from the dissolved and particulate phase (AETE 1997).

Below is the EEM protocol for fish tissue analysis. Other protocols may be used provided

they meet the minimum EEM standards. For example, the Hydro Quebec protocol for the monitoring of mercury levels in fish (Tremblay et al. 1998) has been widely used in Quebec, as this protocol provides an examination of mercury in different age-classes of fish. The protocol can be found on the EEM website.

3.11.1 Selection of the Fish Species

In selecting the location of the sampling areas for the fish tissue sampling, the same factors considered for the fish survey should be taken into consideration. The species selected for tissue analyses should be, if present, sport, subsistence and/or commercial species, including molluscs and crustaceans, where relevant. The fish species used for the tissue analysis may or may not be the same as the species used in the fish survey. On a site-specific basis, the tissue used for the analysis should be chosen based on the portion of the fish constituting the edible portion locally consumed, including the muscle, liver, eggs, hepatopancreas (crustaceans), bone or any other relevant portion. For molluscs, whole soft body parts should be collected, and it may be necessary to produce a composite sample from more than ten specimens to create an adequate sample weight. For lobster or crab, edible tissue (e.g., muscle, eggs, hepatopancreas) should be collected.

3.11.2 Tissue Sample Collection and Preparation

Tissue analyses should be conducted on 8 samples (to achieve 95% power) of a single species from each of the exposure area and reference area, for a minimum of 16 samples. The 8 samples may be tissue from 8 individual fish, or each sample may be a composite of a number of fish; however, tissue from an individual fish should be used in one sample only. If possible, the samples should be of **one** sex- and age-class. The sex of each fish making up the sample should be reported. If fish are not of the same age-class, the age-classes of the fish should be consistent among sampling areas. Although the largest (oldest) fish of a similar size are preferred, the size specifications set by the responsible authority for fishing regulations in the jurisdiction where the study is undertaken should be respected.

The amount of tissue collected should be appropriate for the analytical method being used. Fish should be used independently in a sample and not mixed between samples. Tissues collected for analysis should be handled in such a way as to avoid contamination from sources such as boat fuel. Each sample should be clearly labelled, sealed in an appropriate contaminant-free container, frozen and forwarded to the analytical laboratory. The individual samples should be homogenized separately and sub-sampled for mercury analysis.

3.11.3 Supporting Analyses: Lipid and Moisture Percentages

Monomethylmercury (MeHg) comprises almost all (95% or greater) of the total mercury

found in muscle tissue of fish regardless of the composition of diet sources and exposure water (Bloom 1992). Because of its strong affinity for sulfhydryl groups of proteins, the relative ease with which it passes through the digestive wall and slower depuration rate relative to inorganic mercury, MeHg is accumulated and retained in biological tissues (Clarkson 1994; Saouter et al. 1993).

Lipid concentration has been used to normalize tissue residues among species or within species between seasons, as well as being a key variable in modelling bioaccumulation. Lipid extraction methods by Randall et al. (1991) and the chloroform-methanol extraction method are recommended. Lipid analysis should only be completed when the contaminant being tested is known to be lipophilic.

Percent lipid and percent moisture determinations should be provided for every sample submitted for total metal analysis. Also, percent lipid values should be reported for the replicates analyzed in the same batch with the submitted sample. The percent lipid precision for the replicate samples should be $\pm 30\%$ for tissues containing more than 2% and $\pm 60\%$ for tissues with less than 2% lipid. The method for the lipid determinations would be reported and the solvents used clearly specified.

3.11.4 Guidance for Fish Tissue Analysis for Mercury Using Non-Lethal Methods

Tissue analysis for mercury has been traditionally conducted by extracting a fillet from fish. Non-lethal harvesting methods can produce accurate and reliable measures of fish muscle mercury concentrations provided appropriate analytical techniques are used (Tyus et al. 1999; Baker 2002; Baker et al. 2004; Peterson et al. 2005). The use of non-lethal methodologies for mercury analysis are particularly attractive at sites where destructive sampling methods would be detrimental to fish populations, for example, at sites where fish density is low. The purpose of this section is to describe appropriate non-lethal methodologies for tissue sampling and analysis.

Currently, it is recommended that tissue analysis be conducted on 8 samples (to achieve 95% power) from the exposure area and 8 samples from the reference area of a single species from one sex and age class during a lethal sampling study. This guidance should also be followed in a non-lethal survey with the exception of determining sex. It will not be possible to determine for most species if non-lethal sampling is used. However, several studies failed to find differences in mercury concentrations between males and females, although they can differ in energy requirements (Lange et al. 1994; Henderson et al. 2003; Ward and Neumann 1999).

Baker et al. (2004) demonstrated that small tissue quantities collected with two different types of non-lethal biopsy tools (dermal punch and a Tru-Cut™ biopsy needle) provided accurate and precise estimates of mercury concentration in fish muscle relative to benchmark values from the traditional, fillet-style methods and did not reduce survival of

recaptured Northern Pike. Tyus et al. (1999) examined survival of Rainbow Trout and Razorback Sucker subjected to tissue collection using dermal punches, fin punches or liver punches and found no significant differences in growth or survival in any of the treated fish.

3.11.4.1 Recommended Methodology

Reliability of the non-lethal technique can depend on the biopsy tool, analytical methodology and tissue sample weight (Baker et al. 2004). The following recommended methodology for extraction of fish muscle tissue using a non-destructive approach is based on the work of Baker (2002) and Baker et al. (2004).

1. **Practice** – If at all possible, attempt to collect tissue from archived material or incidental mortalities before trying this method on a living fish. Becoming familiar with a technique will minimize possible handling and sampling stress.
2. **Capture and anaesthetize fish** – Prepare two holding containers, one with well-oxygenated water and another containing an anaesthetic (e.g., MS222). Capture fish by non-lethal means such as angling, short-set gill nets or electrofishing and place in the holding container. Transfer fish to the container containing the anaesthetic, one at a time, as necessary.
3. **Obtain external fish measurements** – Once anaesthetized, weigh and measure the fish, and obtain an aging structure (such as a pelvic fin ray) if appropriate.
4. **Tissue extraction** – Two tools currently available for harvesting small tissue samples include dermal punches or the Tru-Cut™ biopsy needle.
 - Tru-Cut™. Remove two or three scales from the dorsal region of the fish just below the dorsal fin using a sterilized needle. The outer barrel is then inserted to a depth of about 1 cm into the fish muscle tissue beneath the scale at an oblique angle (to minimize penetration depth). The 2-cm-long notched needle (inner barrel) is then extended into the flesh. The containment cover (i.e., sharp outer barrel) slides over the extended needle to cut the tissue and capture it within the notch. The needle is then withdrawn, the barrel opened and tissue slug removed with stainless steel (which should be acid washed between samples) or disposable plastic tweezers and placed in a small labelled vial. Samples obtained are approximately 25 mg. At least two tissue samples should be harvested and composited per fish to obtain a sufficient quantity to permit analysis. Baker et al. (2004) indicate that this procedure requires about 10 seconds for an experienced person to harvest a single sample.
 - Dermal punch. The dermal punch harvests a larger quantity of tissue and, for this reason it is the recommended harvesting method if only cold vapor atomic absorption spectrophotometry (CVAAS) is available for tissue analysis. This

- method can be used on fish greater than 200 mm in size. A few scales are removed and the dermal punch is placed on the skin. Moderate pressure and twisting action is applied to penetrate the epaxial musculature to harvest a small slug of tissue (approximately 60 mg of tissue). As with the biopsy approach, two samples should be harvested per fish and composited.
5. **Sample preservation** – Samples should be frozen using dry ice or liquid nitrogen to prevent decomposition during storage and transport to an analytical laboratory. Samples should be freeze-dried and weighed prior to analysis.
 6. **Infection prevention** – Tissue extraction methods, particularly the dermal punch, leaves an open wound that may lead to an increased likelihood of infection. Sterile crazy glue, such as Nexaband™, which acts like a waterproof bandage, should be used to close the wounds to decrease the chance of infection.
 7. **Monitor and reintroduce fish** – Once the tissue samples are harvested, return the fish to the holding container until it appears to have recovered and swims normally. The fish is then released back into the receiving water body.
 8. **Chemical Analysis** – Selection of an analytical method must consider the accuracy of chemical measurement for small tissue quantities. CVAAS requires a minimum of 100 mg sample weight. Cold vapour atomic fluorescence spectrophotometry has lower detection limits and is better suited to determining mercury concentrations in small tissue quantities. Combustion atomic absorption spectrometry with gold amalgamation is a simplified and rapid procedure for analyzing small tissue quantities for total mercury (Cizdziel et al.2002).

3.12 References

[AETE] Aquatic Effects Technology Evaluation. 1997. Technical evaluation of molluscs as a biomonitoring tool for the Canadian mining industry. Prepared by Robin Stewart and Diane F. Malley for Aquatic Effects Technology Evaluation (AETE) Program, CANMET, Natural Resources Canada. March 1997.

[AETE] Aquatic Effects Technology Evaluation. 1998. Technical evaluation of fish methods in environmental monitoring for the mining industry in Canada. Prepared by EVS Environment consultants for Aquatic Effects Technology Evaluation (AETE) Program, CANMET, Natural Resources Canada. Draft, July 1998.

Bailey RC, Kennedy MG, Dervish MZ, Taylor RM. 1998. Biological assessment of freshwater ecosystems using a reference condition approach: comparing predicted and actual benthic invertebrate communities in Yukon streams. *Freshwat Biol* 39:765-774.

Baker R. 2002. Fish mercury database summary – 2001, British Columbia. Prepared by the Aqualibrium Environmental Consulting Group (now the Azimuth Consulting Group, Vancouver BC) for BC Hydro.

Baker RF, Blanchfield PJ, Paterson MJ, Flett RJ, Wesson L. 2004. Evaluation of nonlethal methods for the analysis of mercury in fish tissue. *Trans Am Fish Soc* 33:568-576.

Barnthouse LW, Suter II GW, Rosen AE. 1989. Inferring population-level significance from individual-level effects: an extrapolation from fisheries science to ecotoxicology. In Suter II GW, Lewis MA, editors. *Aquatic toxicology and environmental fate*, 11th edition, ASTM STP 1001. Philadelphia (PA): American Society for Testing and Materials. p. 289-300.

Barrett TJ, Munkittrick KR. 2010. Seasonal reproductive patterns and recommended sampling times for sentinel fish species used in environmental effects monitoring programs in Canada. *Environ Rev* 18:115-135.

Bloom NS. 1992. On the chemical form of mercury in edible fish and marine invertebrate tissue. *Can J Fish Aquat Sci* 49:1010-1017.

Brasfield SM. 2007. Investigating and interpreting reduced reproductive performance in fish inhabiting streams adjacent to agricultural operations [doctoral dissertation]. Saint John (NB): University of New Brunswick.

Carroll L. 2007. The reproductive cycle of the redbelly dace (*Phoxinus eos*) [honours thesis]. Saint John (NB): University of New Brunswick, Department of Biology.

Cizdziel JV, Hinnert TA, Heithmar EM. 2002. Determination of total mercury in fish tissues using combustion atomic absorption spectrometry with gold amalgamation. *Water Air Soil Pollut* 135:355-370.

Clarkson TW. 1994. The toxicology of mercury and its compounds. In Watras CJ, Huckabee JW, editors. *Mercury pollution: integration and synthesis*. Boca Raton (FL): Lewis Publishers.

Coad BW, Waszczuk H, Labignan I. 1995. *Encyclopedia of Canadian fishes*. Canada Museum of Nature and Canadian Sportfishing Productions Inc. 928 pp.

Dubé MG, Benoy GA, Wassenaar LI. 2006. Contrasting pathways of assimilation: Stable isotope assessment of fish exposure to pulp mill effluents. *J Environ Qual* 35:1884-1893.

Environment Canada. 1997. Fish Survey Working Group final report. Recommendations from Cycle 1 review. EEM/1997/6.

Farwell A. 1999. Stable isotope study of riverine benthic food webs influenced by anthropogenic developments [doctoral dissertation]. Guelph (ON): University of Guelph, Dept. Environ. Biol.

Galloway BJ, Munkittrick KR. 2006. Influence of seasonal changes in relative liver size, condition, relative gonad size and variability in ovarian development in multiple spawning fish species used in environmental monitoring programmes. *J Fish Biol* 69:1788-1806.

Galloway BJ, Munkittrick KR, Currie S, Gray MA, Curry RA, Wood CS. 2003. Examination of the responses of slimy sculpin (*Cottus cognatus*) and white sucker (*Catostomus commersoni*) collected on the Saint John River (Canada) downstream of pulp mill, paper mill, and sewage discharges. *Environ Toxicol Chem* 22:2898-2907.

Galloway BJ, Munkittrick KR, Curry RA, Wood CS, Dunn S. 2004. Identifying a suitable fish species for monitoring multiple effluents in the Upper Saint John River, Canada. In Borton DL, Hall TJ, Fisher RP, Thomas JF, editors. *Pulp and paper mill effluent environmental fate and effects*. Lancaster (PA): DesTech Publications. p. 169-181.

Gray MA, Munkittrick KR. 2005. An effects-based assessment of slimy sculpin (*Cottus cognatus*) populations in agricultural regions of northwestern New Brunswick. *Water Qual Res J Can* 40:16-27.

Gray MA, Curry RA, Munkittrick KR. 2002. Non-lethal sampling methods for assessing environmental impacts using a small-bodied sentinel fish species. *Water Qual Res J Can* 37:195-211.

Gray MA, Cunjak RA, Munkittrick KR. 2004. Site fidelity of slimy sculpin (*Cottus cognatus*): insights from stable carbon and nitrogen analysis. Can J Fish Aquat Sci 61:1717-1722.

Gray MA, Curry RA, Munkittrick KR. 2005. Impacts of nonpoint inputs from potato farming on populations of slimy sculpin (*Cottus cognatus*). Environ Toxicol Chem 24:2291-2298.

Green RH. 1989. Power analysis and practical strategies for environmental monitoring. Environ Res 50:195-205.

Henderson BA, Collins N, Morgan GE and Vaillancourt A. 2003. Sexual size dimorphism of walleye (*Stizostedion vitreum vitreum*). Can J Fish Aquat Sci 60:1345-1352.

Hodson PV, Blunt BR, Whittle DM. 1984. Monitoring lead exposure of fish. In Cairns VW, Hodson PV, Nriagu JO, editors. Contaminant effects of fisheries. New York (NY): John Wiley and Sons. p. 87-98.

Jenkins RE, Burkhead NM. 1993. Freshwater fishes of Virginia. Bethesda (MD): American Fisheries Society.

Lange TR, Royals HE, Connor LL. 1994. Mercury accumulation in largemouth bass (*Micropterus salmoides*) in a Florida lake. Arch Environ Contam Toxicol 27:466-471.

Larsson DGJ, Forlin L. 2002. Male-biased sex ratios of fish embryos near a pulp mill: Temporary recovery after a short-term shutdown. Environ Health Perspect 110:739-742.

Larsson DGJ, Hallman H, Forlin L. 2000. More male fish embryos near a pulp mill. Environ Toxicol Chem 19:2911-2917.

Larsson DGJ, Mayer I, Hyllner SJ, Forlin L. 2002. Seasonal variations of vitelline envelope proteins, vitellogenin, and sex steroids in male and female eelpout (*Zoarces viviparus*). Gen Comp Endocrinol 125:184-196.

Mackay WC, Ash GR, Norris HJ, editors. 1990. Fish ageing methods for Alberta. Edmonton (AB): R.L. & L. Environmental Services Ltd. in assoc. with Alberta Fish and Wildl. Div. and Univ. of Alberta.

McMaster ME, Frank M, Munkittrick KR, Riffon R, Wood C. 2002. Follow-up studies addressing questions identified during cycle one of the adult fish survey of the pulp and paper EEM program. Wat Qual Res J Can 37(1):133-153.

McMullin VA, Munkittrick KR, Methven DA. 2009. Latitudinal variability in lunar spawning rhythms: absence of a lunar patten in the Northern Mummichog (*Fundulus heteroclitus macrolepidotum Walbaum*). J Fish Biol 75(4):885-900.

Minns CK. 1995. Allometry of home range size in lake and river fishes. Can J Fish Aquat Sci 52:1499-1508.

Munkittrick KR. 1992. A review and evaluation of study design considerations for site-specificity in assessing the health of fish populations. J Aquat Ecosys Health 1:283-292.

Munkittrick KR, McMaster ME. 2000. Assessment of multiple stressors in aquatic ecosystems by directed assessment of cumulative effects using fish populations. In Ferenc SA, Foran JA, editors. Multiple stressors in ecological risk and impact assessment: approach to risk estimation. Pensacola (FL): SETAC Press. p. 27-65.

Munkittrick KR, McMaster M, Van Der Kraak G, Portt C, Gibbons W, Farwell A, Gray M. 2000. Development of methods for effects-based cumulative effects assessment using fish populations: Moose River Project. Pensacola (FL): SETAC Press.

Munkittrick KR, McGeachy SA, McMaster ME, Courtenay SC. 2002. Overview of freshwater fish studies from the pulp and paper environmental effects monitoring program. Water Quality Res J Can. 37:49-77.

Munkittrick KR, Arens CJ, Lowell RB, Kaminski GP. 2009. A review of potential methods for determining critical effect size for designing environmental monitoring programs. Environ Toxicol Chem 28:1361-1371.

Nelson JS, Paetz MJ. 1992. The fishes of Alberta. Calgary (AB): The University of Calgary Press.

Nielsen LA, Johnson DL. 1983. Fisheries techniques. Bethesda (MD): American Fisheries Society.

[OMNR] Ontario Ministry of Natural Resources. 1994a. Nearshore community index netting (NSCIN): indexing the abundance of the warm water fish community. FAU Update 94-1, Fisheries Assessment Unit Newsletter. Sutton West (ON): Lake Simcoe Fisheries Assessment Unit.

[OMNR] Ontario Ministry of Natural Resources. 1994b. Spring littoral index netting (SLIN): indexing the abundance of the coldwater fish community. FAU Update 94-2, Fisheries Assessment Unit Newsletter. Bracebridge (ON): Muskoka Lakes Fisheries Assessment Unit.

Peterson SA, Van Sickle J, Hughes RM, Schacher JA, Echols SF. 2005. A biopsy procedure for determining filet and predicting whole-fish mercury concentration. *Arch Environ Contam Toxicol* 48:99-107.

Portt CB, Coker GA, Ming DL, Randall RG. 2006. A review of fish sampling methods commonly used in Canadian freshwater habitats. Canadian Technical Report of Fisheries and Aquatic Science 2604. Catalogue number Cat. No.Fs97-6/2604E.

Randall RC, Lee II H, Ozretich RJ, Lake JL, Pruell RJ. 1991. Evaluation of selected lipid methods for normalizing pollutant bioaccumulation. *Environ Toxicol Chem* 10:1431-1436.

Roberts WE. 1988. The sculpins of Alberta. *Alberta Naturalist* 18:121-127.

Salazar M, Salazar S. 2001. Standard guide for conducting in-situ field bioassays with caged marine, estuarine and freshwater bivalves. Philadelphia (PA). American Society for Testing and Materials (ASTM). 2001 Annual Book of ASTM Standards.

Saouter E, Hare L, Campbell PGC, Boudou A, Ribeyre F. 1993. Mercury accumulation in the burrowing mayfly *Hexagenia rigida* (Ephemeroptera) exposed to CH₃HgCl or HgCl₂ in water and sediment. *Water Res.* 27:1041-1048.

Schetagne R, Verdon R. 1999. Mercury in fish of natural lakes of northern Quebec. In Lucotte M, Schetagne R, Thérien N, Langlois C, Tremblay A, editors. Mercury in the biogeochemical cycle: natural environments and hydroelectric reservoirs of northern Quebec. Springer-Verlag. p. 115–30.

Schetagne R, Doyon J-F, Verdon R. 1997. Summary report: evolution of fish mercury levels at the La Grande Complex, Québec (1978–1994). Montréal (QC): Joint report of the Direction principale, Communication et Environnement, Hydro-Québec, and Groupe conseil Genivar Inc.

Schneider, JC, editor. 2000. Manual of fisheries survey methods II: with periodic updates. Fisheries Special Report 25. Ann Arbor (MI): Michigan Department of Natural Resources.

Scott WB. 1967. Freshwater fishes of Eastern Canada. Toronto (ON) University of Toronto Press.

Scott WB, Crossman EJ. 1973. Freshwater fishes of Canada. Fisheries Research Board of Canada Bulletin 184. Ottawa (ON): Fisheries and Oceans Canada.

Secor DH, Henderson AR, Zapalo A, Piccoli PM. 1995. Can otolith microchemistry chart patterns of migration and habitat utilization in anadromous fishes. *J Exp Mar Biol Ecol* 192:15-33.

Shuter BJ. 1990. Population-level indicators of stress. Amer Fish Soc Symposium 8:145-166.

Swanson SM. 1993. Wapiti/Smoky river ecosystem study. Prepared for Procter and Gamble Ltd./Weyerhaeuser Canada Ltd., Grand Prairie, Alta. by Sentar Consultants Ltd. Calgary (AB).

Tremblay G, Doyon JF, Schetagne R. 1998. Réseau de suivi environnemental du complexe La Grande. Démarche méthodologique relative au suivi des teneurs en mercure des poissons. Rapport conjoint direction principale Communication et Environnement d'Hydro-Québec et Groupe-conseil Génivar inc.

Tyus HM, Starnes WC, Karp CA, Saunders III JF. 1999. Effects of invasive tissue collection on rainbow trout, razorback and bonytail chub. Nor Am J Fish Manage 19:848-855.

Ward SM, Neumann RM. 1999. Seasonal variation in concentrations in mercury in axial muscle tissue of largemouth bass. Nor Am J Fish Manage 19:89-96.

Table of Contents

4.	Effects on Fish Habitat: Benthic Invertebrate Community Survey	4-1
4.1	Overview	4-1
4.2	EEM Phases	4-2
4.2.1	First and Subsequent EEM Phases	4-2
4.2.2	Magnitude and geographic Extent	4-2
4.2.3	Investigation of Cause	4-3
4.3	Study Design Considerations for the Benthic Invertebrate Community Survey	4-3
4.3.1	Power analysis and sample sizes	4-4
4.3.2	Confounding Factors	4-4
4.3.3	Standard Nomenclature	4-6
4.3.4	Reporting of Field Station Positions	4-9
4.3.5	Recommended Sampling Program Designs	4-9
4.3.6	Reference and Exposure Area Consideration for the EEM program phases	4-20
4.3.7	Selection of Ecologically Relevant Habitats	4-21
4.3.7.1	General Guidance for Habitat Selection	4-21
4.3.7.2	Habitat Considerations for the EEM program	4-22
4.3.7.3	Habitat Considerations for Marine/Estuarine Habitats	4-22
4.3.8	Selection of Ecologically Relevant Sampling Seasons	4-25
4.3.8.1	General Guidance for Sampling Season Selection	4-25
4.3.8.2	Sampling Season Considerations for the EEM Program	4-25
4.4	Statistical Considerations for Study Design	4-26
4.4.1	Determination of Sampling Effort for RCA Designs	4-26
4.4.2	Determination of Sampling Effort for Field Sub-sampling	4-27
4.4.3	The Use of Ordination Probability Ellipses for RCA Designs	4-30
4.5	Field Methods for Benthic Invertebrate Community Monitoring	4-31
4.5.1	Sampler Mesh Sizes	4-31
4.5.2	Sampling Equipment	4-33
4.5.3	Artificial Substrates	4-35
4.5.4	Marine/Estuarine Habitat Sampling Equipment	4-35
4.5.5	Sample Containers	4-36
4.5.6	Specimen Fixation and Preservation	4-37
4.5.7	QA/QC for Benthic Invertebrate Field Operations	4-37
4.6	Laboratory Methods	4-38
4.6.1	NABS Certification Program	4-38
4.6.2	Taxonomic Level of Identification	4-38
4.6.3	Reference Collections	4-40
4.6.4	QA/QC for Benthic Invertebrate Laboratory Operations	4-44
4.6.4.1	Sorting Efficiency	4-44
4.6.4.2	Sub-sampling	4-45

4.7	Data Assessment and Interpretation	4-45
4.7.1	Data Handling Methods	4-45
4.7.1.1	QA/QC for Data Input and Verification	4-45
4.7.1.2	Dealing with Outliers	4-46
4.7.1.3	Unknown, Immature and Non-benthic Organisms	4-46
4.7.1.4	Data Reduction and Transformation	4-47
4.8	Data Reporting Guidelines	4-48
4.9	Effect Endpoints and Supporting Endpoints for the Benthic Invertebrate Community	4-49
4.10	Evaluation of Results	4-53
4.10.1	Effect on the Benthic Invertebrate Community	4-53
4.10.2	Next Step	4-55
4.11	Additional Tools for Focused Monitoring, Weight-of-Evidence Approaches and/or Investigation of Cause	4-56
4.11.1	Use of Weight-of-Evidence Approaches to Establish Cause of Effects	4-56
4.11.2	Lethal and Sublethal Toxicity Tests	4-56
4.11.3	Analysis of Sediment Cores for Historic Trends	4-57
4.11.4	Other Benthic Invertebrate Measures and Organisms	4-57
4.12	References	4-59

List of Tables

Table 4-1: Recommended sampling program designs	4-16
Table 4-2: Taxonomic keys for benthic invertebrate taxonomic identification in freshwater environments	4-41
Table 4-3: Recommended level of taxonomic precision for benthic invertebrates in marine environment (for lowest practical taxonomic level approach)	4-42
Table 4-4: List of marine and estuarine taxonomic benthic invertebrate keys for Canada	4-43

List of Figures

Figure 4-1: Examples of area, replicate station and field sub-sample spatial scales for a basic control-impact design	4-17
Figure 4-2: Control-impact designs	4-18
Figure 4-3: Gradient designs	4-19
Figure 4-4: Reference condition approach	4-20
Figure 4-5: Impairment stress levels derived for reference sites in hybrid multidimensional scaling ordination space	4-31

4. Effects on Fish Habitat: Benthic Invertebrate Community Survey

4.1 Overview

The objectives of a benthic invertebrate community survey for environmental effects monitoring (EEM) are to delineate the magnitude and geographic extent of habitat degradation due to effluent discharge, and to provide an evaluation of the aquatic food resources available for fish selected for the fish survey (see Chapter 3 of the present document). However, without a direct comparison between fish diet and the benthic invertebrate fauna, the benthic community survey is mainly aimed at examining habitat degradation. Therefore, the goal of the benthic community survey is to determine if there are structural differences (i.e., total invertebrate density, number of taxa, shifts in the kinds of dominance of taxa) in invertebrate communities in the vicinity of the mine effluent discharge points relative to reference communities. Design considerations will differ depending on whether the mines discharge into freshwater, estuarine or marine receiving waters; this issue is addressed in Section 4.3. It is also recognized that benthic invertebrate surveys will not always use the same study design as the adult fish or water quality surveys because of the different criteria and challenges inherent in the different sampling protocols.

If the benthic invertebrate community survey is conducted in an area where this is possible, sediment samples shall be collected and assessed for particle-size distribution and total organic carbon (*Metal Mining Effluent Regulations* [MMER], Schedule 5, subparagraph 16(a)(iii)). Water samples shall be taken from the sampling areas when the benthic invertebrate community survey is conducted (MMER, Schedule 5, subparagraph 7(a)(ii)). For more information on water and sediment sampling, see Chapters 5 and 7 of the present document.

The objective of this chapter is to provide guidance on the study design and interpretation of results of a benthic invertebrate community survey in relation to EEM requirements. Specifically, this document expands upon 1) study design considerations, 2) standardization of methodologies and 3) data analyses appropriate to the study design. The *Metal Mining Effluent Regulations* (MMER, Schedule 5) set the requirements and timelines for the benthic invertebrate community surveys. The overall framework of the EEM program is presented in Chapter 1 of this guidance document.

The benthic invertebrate community descriptors used to determine effects (effect endpoints) include total benthic invertebrate density, taxa richness, evenness index (Simpson's), and similarity index (Bray-Curtis) (MMER, Schedule 5, section 16 (iii)). Additional community descriptors that could be calculated and reported to assist in data interpretation but that are not used in the determination of effects (supporting endpoints) include Simpson's diversity index, taxon (i.e., family) density, taxon (i.e., family) proportion, and taxon (i.e., family) presence/absence. For more information on benthic

invertebrate community effect endpoints and supporting endpoints refer to Section 4.9 of the present document.

4.2 EEM Phases

4.2.1 First and Subsequent EEM Phases

The first phase of EEM is intended to characterize the benthic communities in major habitats that may be affected by mine effluent and to establish a baseline against which data from future phases can be compared. This phase will also allow for a critical assessment of the need to refine the study design in future phases or the need for the introduction of alternative monitoring techniques. To address the stated objectives of the benthic invertebrate community survey for Phase 1 mines, study design guidelines are presented below.

One specific objective of a Phase 1 survey is to define areas that are relatively homogeneous in terms of habitat class and that have specific ranges in level of exposure to mine effluent.

The study design for the first benthic invertebrate community survey should include:

- 1) Sampling during an ecologically relevant season
- 2) Sampling in both reference and high-exposure areas (e.g., area closest to effluent discharge point)
- 3) Sampling in ecologically relevant habitats
- 4) One of 7 site-specific sampling designs (Table 4-1)
- 5) Site-specific supporting variables
- 6) Standardization of field and laboratory methods

Subsequent EEM phases are intended to confirm the results of the previous phases, help refine monitoring techniques as needed, and determine the factors leading to any detected effect.

4.2.2 Magnitude and geographic Extent

The objective of magnitude and geographic extent studies is to determine the spatial extent of effects on the benthic invertebrate community that are related to mine effluent. Chapter 1 provides information on mines conducting magnitude and geographic extent studies and the critical effects sizes that have been developed by Environment Canada to focus additional monitoring.

Magnitude and geographic extent study designs should include:

- 1) Study and sampling design elements similar to those of previous monitoring, but with more extensive geographic coverage (additional sampling areas)

- 2) An evaluation of the adequacy of previously sampled areas. The new geographic extent may include additional habitats and substrata such as higher-order streams and lakes or marine/estuarine areas ranging from intertidal to subtidal. If these new habitats were not represented in the reference areas used in previous monitoring, a re-assessment of the adequacy of these reference areas is recommended
- 3) The sampling of additional ecologically relevant habitats, seasons or invertebrate life stages, if this is appropriate for assessing the magnitude of the effect
- 4) A consideration of other biotic indicators as tools to assess the magnitude of the effect if their use is appropriate and adds value. The list of optional indicators includes biomass and taxonomic composition of periphyton, phytoplankton, macrophyte or zooplankton communities; sampling of other invertebrate life stages, lower-level invertebrate identification, invertebrate biomass, secondary production, additional sensitive habitats or seasons; and toxicity tests on sediment and water

Magnitude and geographic extent surveys may ask the following questions:

Magnitude:

- 1) How many taxonomic groups are affected?
- 2) What is the magnitude (e.g., the amount of change in density) of the effect on the taxonomic groups affected?
- 3) Is there an effect on other benthic community members, such as periphyton or macrophytes, present in the reference area and expected to be present in the exposure area? Note that this is not a requirement of EEM but could be included in a study of investigation of cause.

Geographic extent:

- 1) What is the geographic area affected?
- 2) Are the benthic invertebrate communities at the sampling stations furthest from the effluent discharge similar to those living under reference conditions?

4.2.3 Investigation of Cause

For information on investigations of cause (IOC), see Chapters 12 of the present guidance document.

4.3 Study Design Considerations for the Benthic Invertebrate Community Survey

Discussed below are various considerations and recommendations which should be examined during the study design process. Benthic invertebrate community survey study designs will be site-specific. The 7 recommended study designs are outlined in section 4.3.5. They attempt to take into consideration factors and possible constraints related to the availability and spatial distribution of suitable reference areas and the spatial extent

and heterogeneity of potential impact areas. It should be emphasized that these guidelines, although considered the most applicable generic designs available, are *not* an exhaustive list of the possible means and ways of conducting a benthic invertebrate community survey. It is assumed that each study leader has sufficient knowledge to apply these recommendations in a sound scientific manner and to determine if unique conditions exist which would warrant modification of the study designs.

4.3.1 Power analysis and sample sizes

For detailed information on power analysis, refer to Chapter 8 of this guidance document.

For the first EEM phase it is recommended that the survey consist of the following:

- 1) At least 2 study areas: reference and high effluent exposure area
- 2) At least 5 replicate stations in each of the 2 study areas
- 3) A minimum of 3 field sub-samples to be taken at each station.

Note that, without a priori information on invertebrate density and variability within a station, the number of field sub-samples required to accurately reflect the true density at each station is arbitrarily set at 3. This amounts to a total recommended sampling effort for mines conducting their first monitoring (Phase 1) of 30 benthic samples. Where study designs other than the control/impact design are appropriate, the same minimum sampling effort should be used, although the distribution of areas, stations and samples may differ.

A further recommendation is that the stations be located such that only the dominant habitat class (see section 4.3.7) is sampled. Restricting sampling to the dominant habitat class reduces data variation. Study areas that have extremely heterogeneous habitats, or two habitats that are equally dominant, may require a greater sampling effort than the minimum previously suggested. Further increases in sampling effort, beyond the minimum, are recommended and could include any of the following: addition of one or more reference areas, addition of a low effluent exposure or a very low effluent exposure area, addition of more stations per area, or the addition of more field sub-samples per station. Increases in sampling effort should be determined in consultation with the Regional Coordinator.

4.3.2 Confounding Factors

Note that the Metal Mining EEM Program does not mandate the Metal Mining Industry to investigate effects of other industries or pollution sources on the benthic invertebrate community under multiple discharge situations.

Are there confounding factors that can be resolved by modifying the study design?

The interpretation of benthic invertebrate community effects may be difficult if

confounding factors exist within the study area. A careful review of historical or existing data and site characterization information to inform decisions about study and sampling design elements can often resolve problems with confounding factors. For additional information on confounding factors, see Hauer and Lamberti (1996), Culp et al. (2000), and Lowell et al. (2000).

Four categories of such factors include:

Environmental variables: Environmental variables can confound the interpretation of benthic invertebrate community effects if it is not possible to separate the effect of the mine effluent from the effects of differences in natural habitat variables. Augmenting the design to better characterize reference conditions with representation of all habitat types sampled may reduce the problem. This could include locating reference areas in adjacent or further-afield watersheds or by sampling additional reference areas (e.g., moving from a simple control/impact design to a more appropriate design; see figures 4-3 and 4-4 and Table 4-1). Some of the potentially confounding variables that may be dealt with by applying more appropriate study and sampling designs include depth gradients, substrate particle size, rapid effluent dilution, interannual and rare events, and seasonal and long-term variability in physical characteristics such as temperature and flow regimes. It may be possible to judge the influence of environmental or habitat variation by examining correlations between measurements of these factors and measurements of the benthic indicators.

Multiple discharges or historic effects: The potential for confounding effects exists if areas with varying levels of exposure to the mine effluent also have varying levels of exposure to other effluents or stressors, or from historic habitat modifications such as dams or impoundments. If feasible, changing the sampling design by modifying sampling locations may reduce the problem. The collection of sediment cores may also be useful in depositional environments to resolve confounding factors resulting from historic effects (see Chapter 7 for more details on sediment monitoring).

Time of sampling: The time of year or the particular year of sampling may confound the interpretation of benthic invertebrate community effects due to effluent. This can be assessed by knowledge of the phenology of benthic invertebrate community species (i.e., relation between climate and life history characteristics) and examination of data collected in previous years from reference areas.

Sampling methods: If standard methods (e.g., sampler types, mesh sizes, taxonomic levels) have not been used consistently within a study or in consecutive studies, any benthic invertebrate community response to the mine effluent may be obscured. It may be possible to examine the data in more detail and convert the data to a comparable level (i.e., convert all taxonomic identification levels to a higher common level). However, in many cases, a redesigned study ensuring that standard methods are consistently applied may be necessary to resolve these problems. Finally, if environmental or logistical conditions exist that preclude the safe and effective collection of samples, the applicability of alternative methods should be examined.

Currently, the only recommended alternative method for the benthic invertebrate component is the application of mesocosms to conduct on-site community bioassays. However, other scientifically defensible monitoring methods that can determine if the mine effluent is having an effect on the benthic invertebrate community may be proposed by the mine. Mesocosms are also useful as an investigation-of-cause tool (see Chapter 12), and their applicability and methodology are described in detail in Chapter 9. Other alternative methods are also described in Chapter 9.

4.3.3 Standard Nomenclature

Standardized definitions for sampling location nomenclature are essential to the EEM program because these will aid national and regional assessments. The following standard terminology for sampling locations should be adopted and applied in a consistent and rigorous manner for all EEM studies with a benthic invertebrate community survey. A schematic representation of these terms is provided in Figure 4-1.

This section defines the terms *field sub-sample* and *replicate station*. Reference and exposure areas are defined in Chapter 2. For the basic analysis of variance (ANOVA) study designs (i.e., control/impact or multiple control/impact), where the objectives are to detect differences between or among areas, each reference or exposure area consists of a number of replicate stations (i.e., the replicates for an ANOVA). Each replicate station consists of a number of pooled field sub-samples. Similarly, gradient or reference condition (i.e., reference condition approach [RCA]) study designs use the replicate station as the spatial scale of replication, with field sub-samples being collected as appropriate. See section 4.3.5 for a description of these approaches.

The concept of area is not directly transferable to the gradient or RCA study designs. When using or developing a gradient or RCA study design, a balanced design with similar numbers of replicate stations located within reference and exposure areas is not the basis for comparisons. For example, in RCA study designs, exposure stations are individually compared to a distribution of reference stations, which represent appropriate reference conditions. For gradient designs, the lack of suitable reference or exposure areas may be the direct cause for selecting this study design, and thus the ANOVA type terminology is not directly applicable for this approach. Detailed guidance on dealing with these study designs is included in section 4.3.5.

Further guidance regarding the number of replicates and their allocation for different spatial scales and study designs is provided in sections 4.3.5 and 4.4.2.

Field sub-sample

Field sub-samples consist of individual area or time-limited collections of benthic invertebrates (e.g., a grab, core, cylinder, quadrat, kick or U-net sample). To ensure adequate spatial placement of field sub-samples within a station, they should be collected in a random or stratified-random pattern. For many of the statistical analyses used to

assess effects in freshwater and marine environments (section 4.9), data from all field sub-samples within a station are pooled, providing a single value of each descriptor from each station.

Pooling of field sub-samples

The pooling of field sub-sample data can occur at several points in the monitoring program. The point at which pooling occurs will depend on several factors, including:

- 1) Field sample processing and storage efficiency (e.g., are field storage jars large enough to contain pooled samples?)
- 2) Laboratory sorting efficiency (e.g., is it more efficient to sort smaller samples?)
- 3) The potential to address study design issues

The first two factors, resulting in an actual physical pooling of the samples, are considered logistical in nature, and their applicability should be determined on a site-specific or method-specific basis. Note that once this physical pooling is done, the potential information from individual sub-samples is lost. In regards to factor 3, if there is a need for additional information to address study design issues (e.g., to examine species area curves or field sub-sampling precision), field sub-samples may be preserved and processed separately. The resulting unpooled data are then available to address the study design issues and can subsequently be pooled electronically for the appropriate statistical analyses. Electronic pooling for the endpoints should be done in such a manner as to be *equivalent* to results if field sub-samples were physically pooled. This is particularly important for the taxa richness endpoint. Sample calculations for pooled station density and richness are shown below.

For density endpoints, values should be calculated as follows:

Density from pooled field sub-samples = (# in field sub-sample a + # in sub-sample b + # in sub-sample c)/total area of field sub-samples a, b and c.

Note that the resulting number is also the same as calculating the density of each sub-sample and taking an average.

However, the calculation of taxa richness for a station is not equivalent to taking the average taxa richness for the three sub-samples. Station taxa richness should be calculated as follows:

Station taxa richness = *all* taxa observed at a station in all sub-samples, *not* the average number of taxa of the three sub-samples.

Replicate station

A replicate station is a specific, fixed sampling location within an area that can be recognized, re-sampled and defined quantitatively (e.g., latitude and longitude and a written description). For each habitat type, a number of replicate stations should be sampled, each resulting in a single composite sample, preferably consisting of ≥ 3 benthic invertebrate field sub-samples. Stations located within the exposure area should be positioned so as to ensure exposure to the effluent plume. Additionally, sufficient physical separation should exist between the replicate stations to allow them to be considered statistical replicates.

The recommended geographic extent of replicate stations for lakes, streams and rivers is as follows:

Lakes: The geographic extent of each replicate station should be at least 10 m x 10 m and separated by at least 20 m.

Rivers and streams: The geographic extent of each replicate station should encompass a longitudinal stretch of the river that includes one pool/riffle sequence. A river distance of six times the bankfull width should be adequate (Leopold et al. 1964; Newbury 1984; Leopold 1994) and allow a minimum separation of three times the bankfull width between stations of similar habitat. To ensure consistency of application for the EEM program, “bankfull width” is defined as in Newbury and Gaboury (1993) and in Chapter 5 of this guidance document. If it is not feasible to sample this length of river (e.g., large rivers or headwater streams with rapidly changing gradients), then an acceptable alternative approach would be to define the geographic extent of stations in a manner similar to that suggested for lakes (i.e., stations are re-visitable locations with predefined dimensions of at least 10 m x 10 m, with adequate separation).

Marine coastal environments: Each of the replicate stations should be a defined location with re-visitable dimensions (e.g., 10 m x 10 m). Replicate stations may be spaced 50 m apart or more, depending on the size of the area. In some estuaries, a replicate station should encompass a longitudinal stretch, which includes the major habitat to be sampled (e.g., a distance of 6 times the bankfull width). If this length of river is not feasible for large estuaries, an alternative definition would be similar to that suggested for coastal areas.

Area

General information and definitions of reference and exposure areas are presented in Chapter 2.

Sufficient geographic coverage for a single benthic invertebrate study area is recommended for lakes, streams and rivers, as follows:

Lakes: The spatial extent of the study area should be at least 100 m x 100 m and large enough to adequately accommodate the necessary number of replicate stations with sufficient separation.

Rivers and streams: The spatial extent of the study area is defined in terms of stream or river morphology and should encompass a length of river that is adequate to accommodate the necessary number of replicate stations with sufficient separation. The total length of river comprising an area would therefore be defined by the number of replicate stations multiplied by 6 times the bankfull width, the river length, on average, in which one pool riffle sequence is expected to occur (Newbury 1984).

Estuary: For low-salinity, relatively homogeneous estuaries, area is defined in the same way as for rivers. For long, narrow marine regions such as narrow bays or fjords in which a control/impact type design is to be used, the area should be large enough to encompass the homogeneous habitat being sampled, as well as the defined exposure range. This will be at least 100 m x 100 m and large enough to adequately accommodate the necessary number of replicate stations.

4.3.4 Reporting of Field Station Positions

Refer to Chapter 2 for general information on the reporting of field station positions.

4.3.5 Recommended Sampling Program Designs

The design of the benthic invertebrate community survey is site-specific, and one of 7 benthic sampling program designs listed below is recommended.

- 1) Control-impact design (C-I)
- 2) Multiple control-impact design (MC-I)
- 3) Before/after control-impact (BACI)
- 4) Simple gradient (SG) design
- 5) Radial gradient (RG) design
- 6) Multiple gradient design (MG)
- 7) Reference condition approach (RCA)

Examples of these designs are illustrated in Figures 4-2, 4-3 and 4-4.

These designs fall into three basic categories with different “philosophical” approaches, as follows:

- a) The C-I or MC-I designs (including BACI) are ANOVA-type designs used to detect differences between discrete exposure and reference areas.

- b) The gradient (SG, RG or MG) designs are intended to examine changes in community structure along a physical and/or effluent gradient, and are better suited to regression analyses or analysis of covariance (ANCOVA).
- c) The multivariate approach of the RCA compares potential “impaired” or test stations to a selection of appropriate reference stations.

It should be noted that there may be some circumstances where ANOVA analyses are applicable to b) and c) above. Alternative monitoring methods (e.g., mesocosms) are also recommended but must be scientifically defensible. A summary of the attributes, applicability and limitations of the sampling designs is presented in Table 4-1 and described in more detail below.

The following descriptions apply primarily to the design of the first and subsequent phases, although special applications for determining the magnitude and geographic extent of an effect are indicated, where applicable.

Control-impact design

The simplest study design for use in EEM is the control-impact (or reference-exposure) design (Green 1979). In rivers and estuaries, this consists of no less than one reference area and a series of downstream exposure areas. For regular monitoring, this should include, at a minimum, one high effluent exposure area. Levels of exposure to mine effluent differ between exposure and reference areas, but should be similar between the stations within each area. Habitat classes sampled should be consistent among areas and, with the exception of exposure level, these areas are to be as similar as possible in terms of substrate, depth, current velocity, water properties, environmental gradients, land use, etc. The first study design employs ANOVA comparisons among areas and is recommended for simple, homogeneous rivers and streams without confounding upstream or near-site discharges from other sources.

The mine may propose modifications to this basic ANOVA approach providing that the modified design is scientifically defensible and addresses the appropriate monitoring questions. For example, if a reference area cannot be located upstream or in adjacent watersheds due to a confounding factor, but a C-I design would otherwise be applicable, a modification of the C-I design may be appropriate. In this case, the design could be modified so that the reference area is “downstream” instead of “upstream” of the point source. The downstream reference area would have to be outside of the exposure area and meet the same reference area criteria as other designs.

This first study design is also recommended for simple, homogeneous estuaries or narrow inlets or bays without confounding upstream or near-site discharges from other sources or where the ecologically relevant habitat occurs in spatially discrete but homogeneous patches (i.e., intermittent rocky outcroppings).

Magnitude and geographic extent

The C-I design can be used to ascertain the geographic extent of an effect by first making use of rapid bioassessment protocols (Plafkin et al. 1989) or other available information to approximate how far the effect extends. Following this, a C-I monitoring program can be used that includes the high effluent exposure area and targets additional exposure areas in localities where the effect is suspected to be dissipating (e.g., additional exposure areas located so as to bracket the suspected furthest reach of the effluent effects, together with the previous reference and exposure areas). ANOVA comparisons among areas can then be made to determine the geographic extent of an effect at a given significance level.

Multiple control-impact design

Two of the major problems associated with the use of a single reference area are 1) it can be easily confounded by other factors, and 2) there is a lack of independence among the stations in a single reference area (pseudoreplication) (Hurlbert 1984). In systems where an appropriate reference area is not available due to confounding factors or where it is determined, after a review of historical information, that more reference areas are desirable, the MC-I design should be used. Schematic diagrams of this design for application in mines discharging to large rivers, lakes or coastal waters are presented in Figures 4-2d, e and f. Sampling schemes should be devised so that additional reference areas are located in adjacent watersheds or bays and that comparable habitat classes spanning the range of habitats found within the exposure area are selected.

The design philosophy of both the C-I and MC-I designs is that a specific difference in magnitude of effect between a series of areas is being examined. This lends itself to a classic ANOVA design with associated power analyses. These methods are statistically tractable and can provide indicators as to whether or not there is a biological effect from the mine effluent on the benthic invertebrate community. These designs assume that effluent exposure and habitat conditions are relatively homogeneous among all stations within a sampling area or that effluent exposure is within an acceptable range for a particular defined area.

Before/after control-impact designs

An improvement to the above C-I and MC-I designs is possible when data can be collected both before and after initiation of effluent discharge into the receiving water area. The same considerations discussed above apply for choice of reference (control) and exposure (impact) areas. But the design is further enhanced by collecting data both before and after the facility becomes operational. This kind of monitoring design has been termed a before/after control-impact (BACI) design (Schmitt and Osenberg 1996). Use of a BACI design helps to distinguish effluent effects from natural differences between reference and exposure areas that may have existed before the initiation of effluent discharge.

Detailed descriptions of several kinds of BACI designs and their statistical analyses are

available in Green (1979), Schmitt and Osenberg (1996), Underwood (1997), and references therein. In its simplest form, a BACI design entails collecting monitoring data at least once, both before and after initiation of effluent discharge, in both a reference and an exposure area, after which the data are analyzed using an area-by-time factorial ANOVA (Green 1979). In this situation, evidence for an effluent effect is inferred when the area-by-time interaction term in the ANOVA is significant. When the reference and exposure areas have been sampled repeatedly during both the before and after periods, it is possible to use a BACI paired-series analysis; in this case, potential effects are investigated by testing for a change in delta (difference between reference and exposure) from the before to the after period (Schmitt and Osenberg 1996). The design can be further improved by incorporating multiple reference areas (Schmitt and Osenberg 1996; Underwood 1997). Refer to Chapter 2, section 2.2.2.2.2 for additional information on baseline data.

Simple gradient and radial gradient designs

Simple and radial gradient designs (Figures 4-3a, b and c) are suitable for situations where rapid effluent dilution precludes the selection of an exposure area that is comparatively homogeneous in terms of effluent concentration. As with the C-I design, gradient designs can be used in cases where no suitable reference areas are available upstream or in adjacent watersheds or bays. Gradient designs are also useful for determining how far along an effluent path effects are observed (i.e., objective of magnitude and geographical extent).

Philosophically, the gradient approach examines departures from expected (non-impacted) “patterns” of correlated biotic and environmental factors over spatial gradients. This is more suited to a regression type of analysis (or equivalent) in which replication (i.e., five stations within an area) is less appropriate than expending a similar effort to obtain accurate measurements of biotic and habitat variables over a sufficiently broad range of the gradient conditions. In the simplest case, a statistically significant effect would be declared if the slope of the regression of a response variable against distance from the effluent source is significantly different than 0. In this approach, a point-source discharge is expected to have a “declining” gradient of effects away from the source, and it is not always feasible to make the judgment that there either “is” or “is not” an effect at a given station. At a certain point along the gradient it is necessary to judge that this effect is no longer measurable or important. Therefore, in gradient designs, reference information is obtained from the stations furthest away from the effluent source.

A gradient does not necessarily imply straight lines or the even spacing of stations within areas. The spacing of stations may be more or less continuous on a gradient away from the discharge, with less emphasis on distinctly different dilution zones and more on adequate geographic coverage, as compared to a C-I design. There are often no “blank” spaces between distinct sample areas, but rather a continuum of sampling stations along the gradient. However, if a change in effluent dilution within the receiving environment

is abrupt, more sampling effort may be desirable over these stretches to accurately track rapid changes in mine effects.

SG designs are particularly appropriate for narrow water bodies such as rivers and streams. In wider water bodies such as lakes or open coastal areas, a radial gradient design may be more appropriate. Sampling is conducted away from the effluent source along several gradient transects. As in the MC-I approach, the use of an RG design will provide a larger number of reference sites. Furthermore, a broader geographic area will be sampled, which can be important in non-homogeneous, open lakeshore or marine areas, which often have complex current and circulation patterns or a variety of equally important major habitat classes or gradients.

For RGs, a comparison of regression patterns for each gradient (e.g., regressions of faunal abundance versus distance from the outfall) may help to illuminate the direction and extent of effects. Alternatively, all data from all gradients can be included in one regression, if the comparison is between biotic and physical factors unrelated to geographic or natural habitat factors. If sufficient sampling is done (e.g., RGs), it may be possible to pick and choose unconfounded replicate stations (e.g., homogeneous habitat conditions) to regress a biotic versus a mine-related variable.

Wherever possible, the exposure gradient should be de-coupled or independent from any environmental gradients. A declining exposure gradient may fall along a path with varying depths, but an SG or RG approach may still be feasible if the exposure and depth gradients are not correlated and the differences in depth are not so great as to obscure any effluent effects. In cases where the exposure gradient is correlated with a co-occurring environmental gradient, an MG design may be more appropriate (see next section). Alternatively, a multivariate approach may be necessary to remove the confounding influence of varying depth.

Gradient designs and magnitude and geographic extent

Due to the layout of sampling stations, gradient designs are particularly well suited for determining the geographic extent of an effect. The simplest design for magnitude and geographic extent would be to allocate sampling stations along a gradient from more to less exposed, ensuring that the most distant stations are located well beyond the likely extent of effects. The geographic extent of effects could then be determined graphically by plotting the response variables against distance from the mine and inspecting the data for an inflection point where the response variable asymptotes to the reference condition. Data from sampling stations arrayed in this manner could also be used, together with measured physicochemical data, in a multivariate analysis (e.g., ordination or clustering) used to identify which distant stations tend to group with reference stations and which tend to group with clearly impacted stations. Both of these approaches (graphical plotting and multivariate analysis) look for patterns in the data to determine the approximate extent of an effect; that is, they do not entail hypothesis testing and therefore a power analysis would not be applicable in these cases (in contrast to the C-I approach to magnitude and geographic extent described above).

It is also possible to design a hypothesis-testing gradient program for examining the geographic extent of an effect. This would entail using field sub-samples as replicates (treating stations as areas) and making station-by-station ANOVA comparisons along a gradient to determine where an effect disappears at a given significance level. However, this latter approach might require extensive sampling effort, depending upon the number of stations along the gradient and the required (by power analysis) number of field sub-samples per station.

Multiple-gradient design

In some cases, it may also be useful to compare reference gradients to those exposed to mine effluent. This would be the case when a co-occurring environmental gradient confounds an effluent gradient in the exposure area. By using a MG design (see charts d) and e) of Figure 4-3), it is possible to make statistical comparisons of the exposure area gradient to a similar environmental gradient in a reference area. The reference gradients should be as similar as possible in depth and habitat to the exposure gradient. Potential effluent impacts would be tested for by using ANCOVA to factor out the influence of the co-occurring environmental gradient.

Reference condition approach

The fundamental concept of the RCA is to establish a database of sites that represents unimpaired conditions (reference stations) at which biological and environmental attributes are measured. This database is used to develop predictive models that match a set of environmental variables to biological conditions. These predictive models then allow a set of environmental measurements to be made at a new station and used in the model to predict the station's expected biological condition (i.e., the biological conditions of the group of reference stations with similar environmental attributes). A comparison of the actual biological condition at the new station with the predicted conditions allows an assessment of the condition of the new station to be made.

The RCA can reduce the need to find nearby comparable reference sites when studying an impacted system, which can be a problem in some traditional approaches. Rather than identifying and sampling upstream reference sites in a river system or next-bay-over reference sites in a lake, the RCA uses a set of biologically equivalent reference sites selected from an existing database to evaluate an exposure site. Provided that it is kept up-to-date, the reference condition database can be used over a number of EEM phases.

The reference condition database is established by an initial standardized sampling program at a wide variety of geographic scales. The same benthic invertebrate community sampling protocol is used in as many ecoregions and stream orders or lakes as are available in a catchment. A number of environmental variables are measured in conjunction with invertebrate sampling. The data are then subjected to a 3-step multivariate analysis in which:

- 1) a number of invertebrate groups are formed based on similarity of community structure;
- 2) biological data are correlated with environmental attributes and an optimal set of environmental variables is identified that can be used to predict group membership; and
- 3) the biological condition of test (exposure) stations is assessed by using the optimal set of environmental variables to predict group membership. How the test station fits, relative to the group to which it is predicted to belong, establishes whether, and to what degree, the station is different from the reference group. Assessment can be made by either the use of the community descriptors, by determining if the site is within the prescribed range of variation observed at reference sites (2 standard deviations [SDs]), or by the use of ordination methods and determining if the exposure site is within the 95% probability ellipse of the matched reference sites.

Depending on the timing and location of the sampling program, it may also be possible to use the resulting database to make ANOVA comparisons between reference and exposure areas.

Once the reference database is established, the RCA can be used as a rapid bioassessment method and to deal with national and local issues using the same database and software. Due to the intensive initial sampling effort required, the RCA would not be considered a practical approach for use by a single mine in a remote location if a reference database is not already available; however, it may be applicable in areas where multiple industries (including different EEM industrial sectors) are located. In this case it may be practicable and cost-effective for multiple users to collaborate in the development of the reference database. Additional information on the RCA can be found in Bailey et al. (2003).

To assist industry in locating suitable reference sites for the EEM program, the Cooperative Freshwater Ecology Unit of Laurentian University has led the Northern Ontario Benthic Invertebrate Reference Condition Approach (RCA) Biomonitoring Network (Northern Ontario RCA Network). For additional information on this network refer to the following website:

http://laurentian.ca/Laurentian/Home/Departments/Cooperative+Freshwater+Ecology+Unit/Research-Projects/Evaluating+Ecosystem+Health/RCA+Programs.htm?Laurentian_Lang=en-CA

The Canadian Aquatic Biomonitoring Network (CABIN) is a collaborative program developed and maintained by Environment Canada to establish a network of reference sites through the RCA. This information is available to all users interested in assessing the biological health of freshwater in Canada. For additional information on CABIN, please refer to their website at: http://cabin.cciw.ca/Main/cabin_about.asp.

Table 4-1: Recommended sampling program designs

Design Type	Receiving Environment	Reference/Control Area	Impact Area	Statistics
Control-impact (C-I) <i>Figure 4-2</i>	Freshwater rivers or lakes, homogeneous or low-salinity estuaries	A single reference area, upstream of mine effluent outfall	High effluent exposure area (additional exposure areas are added in magnitude and geographic extent)	ANOVA
Multiple control-impact (MC-I) <i>Figure 4-2 d,e,f</i>	Freshwater rivers or lakes with geographically homogeneous lake shores, homogeneous estuaries and coastal zones	Multiple reference areas in the same or environmentally similar adjacent watersheds or bays	High effluent exposure area (additional exposure areas are added in magnitude and geographic extent)	ANOVA
Before/after control-impact (BACI)	Same as C-I and MC-I	Same as C-I and MC-I, but with sampling done both before and after initiation of effluent discharge	Same as C-I and MC-I, but with sampling done both before and after initiation of effluent discharge	ANOVA
Simple gradient (SG) <i>Figure 4-3a, b</i>	Freshwater rivers or geographically restricted lakes, non-homogeneous, narrow estuaries or geographically restricted marine bays, inlets or fjords	A series of reference stations with little or no effluent, situated towards the end of a declining gradient of mine effluent	Single gradient through declining levels of effluent in the receiving environment	Regression/ ANCOVA
Radial gradient (RG) <i>Figure 4-3 c</i>	Lakes, non-homogeneous open marine bays and coastal areas	As above, but situated near the end of several radially oriented transects	As above, but repeated in a radially oriented design	As above
Multiple gradient (MG) <i>Figure 4-3 d, e</i>	Freshwater lakes or rivers Non-homogeneous open marine bays and coastal areas	A series of reference stations with no effluent situated on a transect along the same kind of environmental gradient observed in the exposure area	Gradient through declining levels of effluent and a co-occurring environmental gradient in the receiving environment	ANCOVA, with reference and exposure transects considered as treatment groups
Reference condition approach (RCA) <i>Figure 4-4</i>	Freshwater rivers or lakes, particularly for cooperative investigations or where there is an existing reference database	Multiple series of reference stations with little or no effluent situated in similar drainage basins within the same ecoregion	Series of stations within the exposure area which are tested individually against the reference station distribution	Multivariate/ ANOVA (if possible)

Note: Multivariate analyses can be performed on data collected using any of the above designs to look for patterns (i.e., not hypothesis tests) that may be useful for highlighting potential areas of concern.

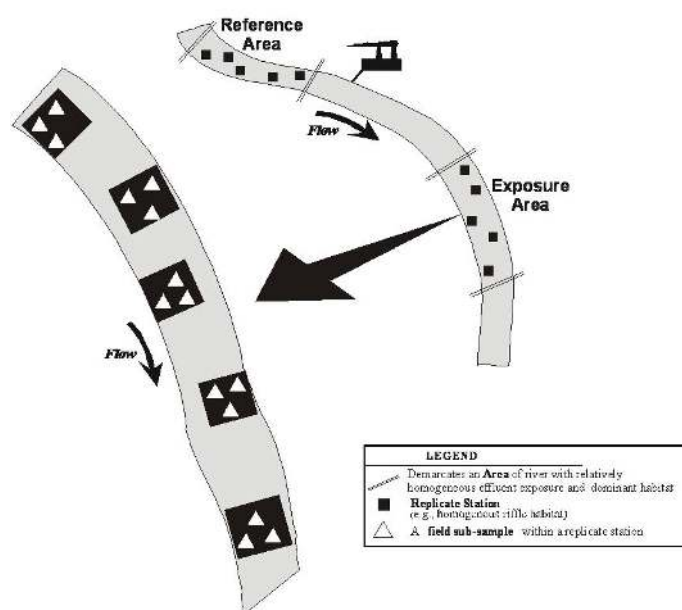


Figure 4-1: Examples of area, replicate station and field sub-sample spatial scales for a basic control-impact design

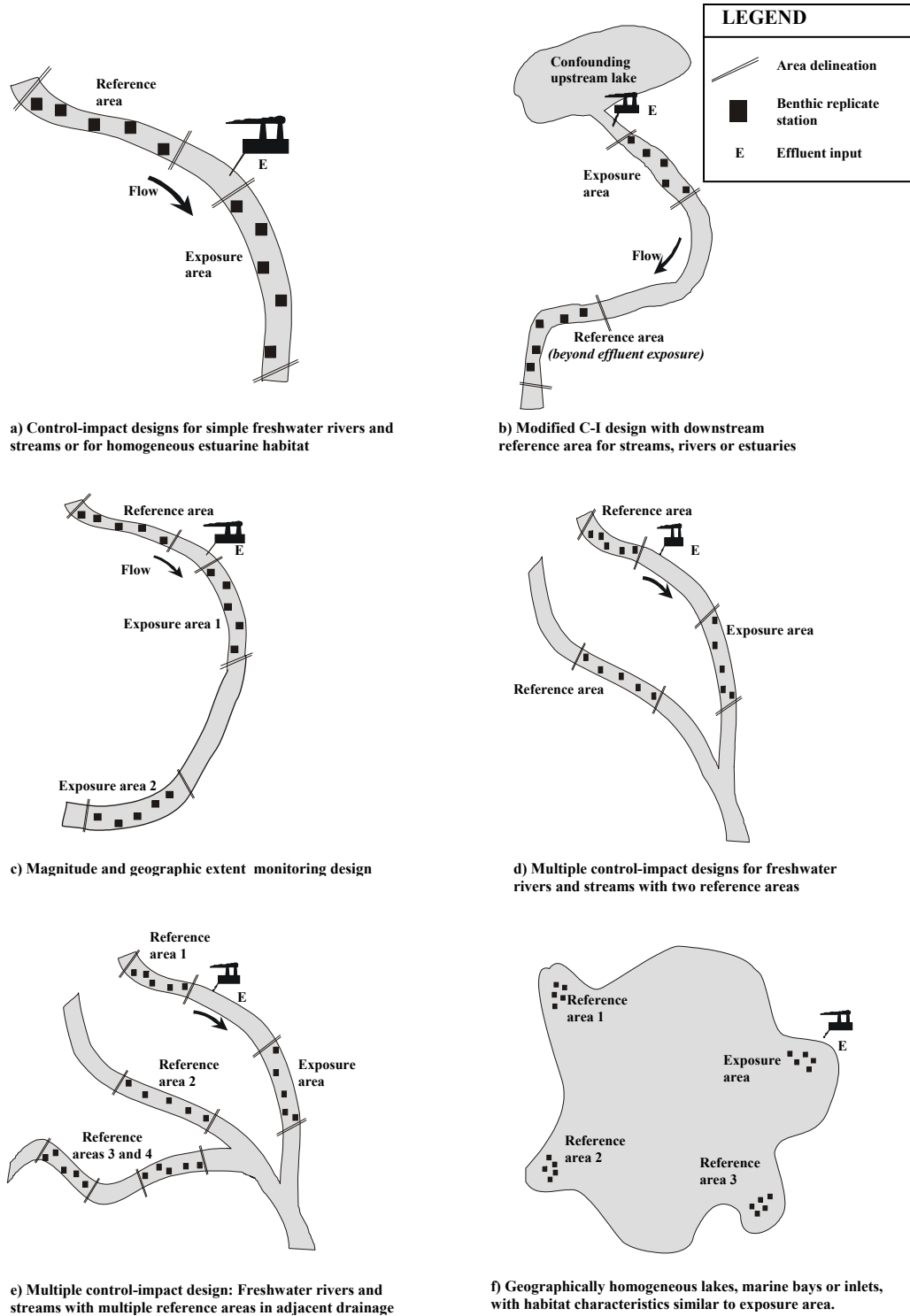
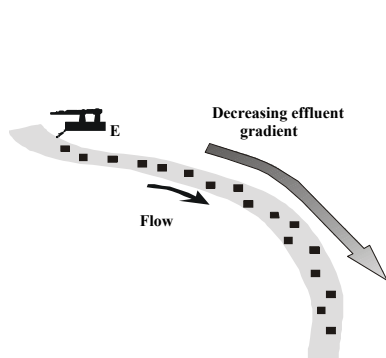
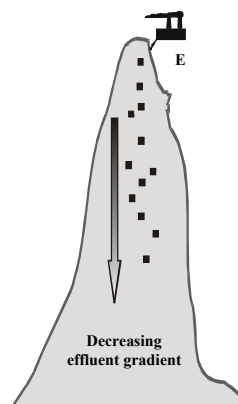


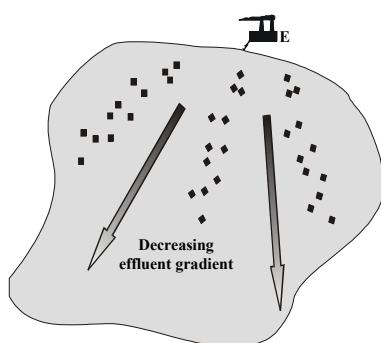
Figure 4-2: Control-impact designs



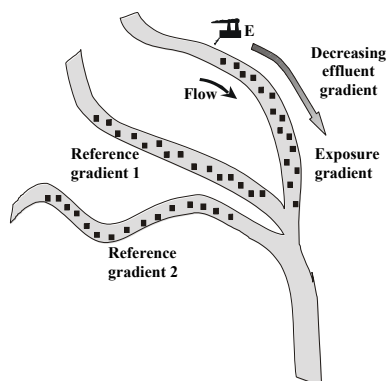
a) Simple gradient designs for freshwater rivers and streams and estuaries



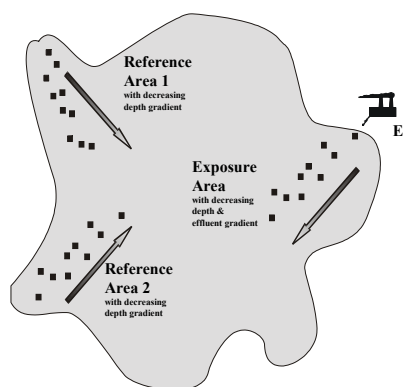
b) Simple gradient designs for lake or coastal sites situated in narrow bays or fjords



c) Radial gradient design for lake or coastal situations



d) Multiple gradient design for freshwater rivers or



e) Multiple gradient designs for lake or coastal sites

LEGEND	
	Benthic replicate station
E	Effluent input

Figure 4-3: Gradient designs

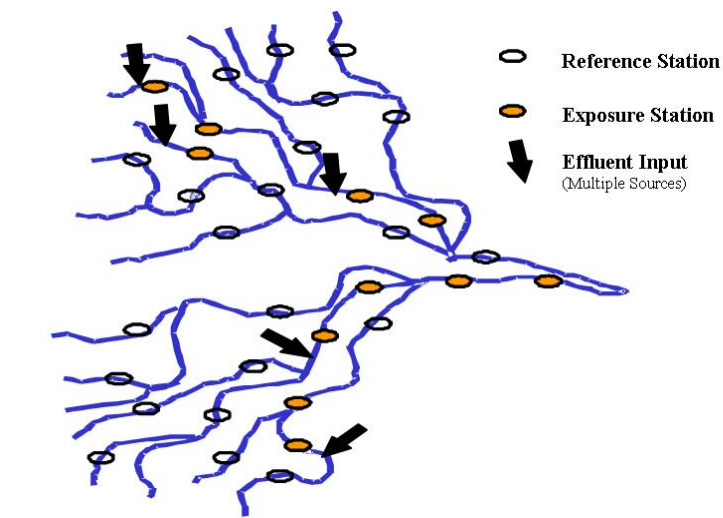


Figure 4-4: Reference condition approach

4.3.6 Reference and Exposure Area Consideration for the EEM program phases

The allocation of reference and exposure areas is dependent on the site-specific study design and the phase of the EEM program.

For Phase 1, the objective is to determine whether there is an effect on the benthic invertebrate community in the high effluent exposure area where an effect is more likely to occur. This spatial limitation is designed to concentrate sampling effort in a cost-effective manner. The study design and allocation of reference and exposure areas should be based on this objective.

For **subsequent phases**, the objectives are to confirm results, detect changes and allow for trend monitoring data. As these objectives are similar in geographic scale to Phase 1, the selection criteria for reference and exposure areas should be the same. However, as with any ongoing monitoring program, the appropriateness of reference and exposure area selection should be re-evaluated as additional information is gained.

For **magnitude and geographic extent** the objective is to determine the spatial extent of previously identified effects. Thus, sampling should be conducted at exposure areas located farther away from the mine effluent discharge point, until a return to reference conditions is reached. The physical allocation of multiple exposure areas and stations is dependent on the study design. If a confounding factor is encountered before the reference condition is reached in the low effluent exposure area, and this factor cannot be resolved by modifying the study design (see Table 4-1), then the exposure area may be

defined to extend only as far as the confounding factor is encountered. Alternative, cost-effective study designs or methods may be applicable (see the Chapter 9 and Table 4-1).

In addition, as part of the review of monitoring information, reference areas sampled in previous monitoring should be re-evaluated to assess whether they are adequate for the magnitude and geographic extent program. The new geographic extent may naturally include additional habitats such as higher-order streams or lakes. If these new habitats were not represented in the reference areas that were used for previous monitoring, a reassessment of the adequacy of these reference areas will be necessary. The addition of reference areas should also be considered to allow a more balanced design between the number of reference and exposure areas.

If an RCA study design was used during previous monitoring, additional reference areas may not be necessary (assuming they adequately represent the habitat types), but it is recommended that a subset be re-sampled to examine the effects of natural temporal variation.

4.3.7 Selection of Ecologically Relevant Habitats

4.3.7.1 General Guidance for Habitat Selection

The most ecologically relevant habitats should be sampled within the exposure areas, and similar habitats should be located and sampled within the reference areas. The selection of the appropriate habitat types requires consideration of the following questions:

- Which habitat type is present in the highest proportion in the exposure area?
- Which habitat, in the absence of human influences, supports the richest assemblages of invertebrates (benthic invertebrate diversity) within the study area?
- In which habitat are the invertebrates most likely to be exposed to sediment or water-borne contaminants for extended periods of time?
- Is historical information available for a particular habitat?

The first consideration is to sample the habitat that accounts for the greatest proportion of the exposure area. However, other factors can override the importance of geographically dominant habitat including the ecological relevance of sampling highly sensitive and diverse habitats, even if they comprise a lower proportion of the study basin. In streams, riffles can support a diverse assemblage of species that display a wide range of sensitivities to water-quality changes. Therefore, the community in this habitat has the potential for greater change than less species-rich communities. In contrast, the fauna of depositional areas, which are generally less rich taxonomically, are of interest during biomonitoring exercises because they may be directly exposed to concentrations of sediment-borne contaminants for longer periods. Consequently, communities in depositional areas may respond to contaminants differently than the more sensitive but

less exposed riffle habitats. For additional guidance regarding stream habitat selection, refer to Cuffney et al. (1993), Plafkin et al. (1989), and Meador et al. (1993).

4.3.7.2 Habitat Considerations for the EEM program

The decision about which habitat to sample should be based on site-specific considerations. Decisions about the sampling of more than one ecologically relevant habitat during the same survey depend upon the phase of the EEM program.

For **Phase 1** and subsequent phases, the objective is to determine if there is an effect on the benthic invertebrate community; therefore, the habitat most likely to exhibit these effects should be sampled. If more than one habitat is determined to be ecologically relevant, effort could be expended during magnitude and geographic extent monitoring to sample all ecologically relevant habitats. This may lessen the potential for missing an effect on a sensitive habitat and/or the necessity of expanding the survey to additional habitats during future monitoring efforts. If questions regarding magnitude and extent can be addressed by additional sampling during the same field trip, it may be cost-effective to do so.

Sampling of additional ecologically relevant habitat types should not be at the expense of reduced sampling effort in the primary habitat of interest. For most biomonitoring studies, sampling a single habitat is intended to reduce the variability inherent in sampling natural substrates. This variability would be even greater if the same level of effort were spread over a range of habitat types.

For magnitude and geographic extent monitoring, the most ecologically relevant habitats should be sampled within the exposure areas, and similar habitats should be located and sampled within the reference areas. The decision regarding number and type of habitats to sample is made based on a review of the previous monitoring results, site-specific considerations and the objectives of magnitude and geographic extent. For example, to determine the geographic extent of an observed effect, additional habitats such as higher-order streams or lakes may become important. On the other hand, if, during previous monitoring, a number of habitat types were sampled but one particular habitat appeared to show responses and the others did not, this habitat type could be targeted during magnitude and geographic extent monitoring.

4.3.7.3 Habitat Considerations for Marine/Estuarine Habitats

A decision needs to be made to sample either depositional or erosional habitats in the estuarine/marine receiving environment. In addition, decisions about sampling intertidal vs. subtidal substrates for estuarine/marine mines will depend on which is the appropriate receiving environment and on the feasibility of obtaining useful samples. In marine/estuarine habitats, the selection of the appropriate habitat types therefore requires consideration of the following questions:

- What habitats are feasible to sample?

The habitat that is most common geographically and most likely to be affected by the effluent should be selected. However, selection of major habitats is partly related to viability of sampling. For example, if the major habitat is a vertical rock cliff to a depth of 300 m immediately outside the effluent discharge outfall, this is rarely feasible for benthic invertebrate sampling without extraordinary equipment. As another extreme example, if the major habitat is intertidal but consists of a steep rock cliff with heavy wave and wind exposure or ice build-up, sampling may not be possible. When multiple habitats are available and appropriate, some choices need to be made. In some cases, more than one habitat may have to be sampled (radial gradient or similar design). Where there is a choice, sampling of soft substrates is preferred because methods are generally more quantitative.

- What is the habitat that is most biologically active or “important?”

When the subtidal environment most exposed to mine effluent consists of both consolidated and unconsolidated sediments, then either both substrates need to be sampled or a decision on which to sample must be made. All other factors being equal, the unconsolidated sediment is more efficiently sampled quantitatively. However, when it is obvious that a coarse sand substrate is almost devoid of macrofauna within the top 10 cm (or the depth of penetration of a sampling device), whereas the nearby rocky reef is extremely rich and an obvious haven for many fish, it is the most “active.” Similarly, if there is an important fishery resource in one major habitat type that is directly exposed to mine effluent, it may be considered the more biologically important.

- Can the ecologically relevant habitat be “classified” according to recognized physical type and characteristic species?

Habitat classification systems have been discussed and reviewed by many researchers and can be useful for comparing “expected” biotic factors with actual biotic factors present in the mine vicinity. Some marine examples include a comprehensive draft document presented to DFO for delineating the Strait of Georgia on the west coast of Canada and northwestern U.S. (Watson 1997). Some relevant references for marine classification for worldwide shorelines to deep coastal areas include Frith et al. (1993), Booth et al. (1996), Robinson and Levings (1995), Hay et al. (1996), and Robinson et al. (1996). Specifically, estuarine classification has been reviewed by Matthews (1993), Scott and Jones (1995), Finlayson and van der Valk (1995), and Levings and Thom (1994). In the U.S., the most widely used system is that of Cowardin et al. (1979) and Cowardin and Golet (1995), with expansions proposed by other authors.

- Is the effluent discharge depth and/or buoyancy most likely to affect the intertidal or subtidal regions?

If the effluent discharge is and remains mainly intertidal, then this should be the targeted habitat. However, if the effluent affects both intertidal and subtidal habitats, then the subtidal is the preferred habitat, because this area is most likely to show impacts in fish. If suitable, both habitats may be sampled. This question should also take into consideration seasonal water column stability changes, which can affect intertidal areas.

- What habitat type is present in the highest proportion?

In many cases, coastal shorelines will be mixed silt, sand, gravel and rock substrates. In bays near freshwater discharge points, there tends to be accumulations of sandy or silty sediment. Estuarine mine sediments will usually have dominantly soft substrates from river-borne material. If there are similar percentages of both depositional and erosional habitats, the preferred habitat to sample is depositional, because this type will accumulate the discharged material from mine effluents and is more likely to present deleterious effects. Erosional substrates tend to be kept “clean” by high current action or wave or ice scouring.

However, if the percentage of solid substrate habitat is much greater than the soft substrate habitat, or if a previously “clean” rocky shoreline has begun to accumulate sediment related to mine discharge, then this may be the preferred habitat to sample.

- Are there confounding factors that may affect benthic communities?

Benthic communities in naturally or anthropogenically confounded sampling areas are problematic to use for interpreting effects from mines. Obviously they should be avoided. For example, in situations where consolidated and unconsolidated substrates are present, only one of these may be outside the influence of the confounding factors. One source of confounding factors that is particularly important in Arctic areas is the seasonal or year-round effects of freezing or ice scour, particularly in intertidal or estuarine areas, which may seriously disrupt surficial communities.

- What is the environment affected by subtidal discharges?

Obviously, the environment most exposed to effluent should be the targeted sampling area and will also determine the type of sampling design used. In an estuary, if the discharge occurs at the surface where there is a strong and permanent surface freshwater layer with little intrusion of saltwater at high tide at depth, then the habitat to be sampled is downstream from the mine. However, where there is a strong tidal intrusion, sampling will have to go upstream and downstream. There are numerous other factors of this type to consider, all of which require detailed site-specific information about the habitats and the pattern of effluent dispersion.

In summary, if there is a choice of habitats to sample, it is recommended first that subtidal habitats be sampled because they tend to have higher diversity and less patchiness in fauna than the intertidal, due to less extreme or harsh habitat conditions.

This is particularly true in Arctic regions, where extreme wintertime conditions may eliminate most of the longer-lived fauna that tend to more clearly integrate the effects of contaminants. Second, if there are a variety of suitable habitats, depositional habitats should be chosen, particularly for subtidal areas, because the methodology allows easier and more quantitative sampling procedures. Depositional areas also tend to accumulate contaminants over time, whereas erosional areas may not.

4.3.8 Selection of Ecologically Relevant Sampling Seasons

4.3.8.1 General Guidance for Sampling Season Selection

All benthic invertebrate community surveys should be completed during the most ecologically relevant season. Sampling should occur during a period of effluent discharge and after the receiving environment has been exposed to the effluent for a sufficient period during which effects would reasonably be expected to occur (i.e., generally within 3-6 months).

The preferred seasonal period for sampling is when biological diversity is highest. In general, this corresponds with the seasonal recruitment cycles of benthic organisms (generally related to climate and food abundance). Many insects with freshwater stages reproduce in the spring and fall, although others have multiple cohorts throughout the open-water period. For many lotic habitat types, sampling is conducted during the fall (September/October), when the majority of taxa are present and/or large enough to be collected by the sampling equipment and flow regimes allow access for sampling. In large lakes where the benthic community is often dominated by annelids, crustaceans and molluscs, insect emergence periods and hydrologic regimes are of less importance in determining the sampling period (Rosenberg and Resh 1993).

If historic benthic invertebrate community surveys exist for the system under investigation, it is useful to examine the data and, if appropriate, conduct the survey during similar periods so that the surveys can be compared. Other factors that may influence the sampling period include seasonal flow disruption such as extreme high- or low-flow conditions, freezing and ice scour, mine effluent discharge conditions, type of sampling gear, and feasibility of sampling and field crew safety. Sampling during periods when effluent is not being discharged should be avoided. An understanding of the seasonal patterns and life cycles of the taxa along with changes in the hydrologic regime found within the specific system is also helpful to determine the appropriate timing for the survey. Rosenberg and Resh (1993), Johnson et al. (1993), Rees (1984), Malley and Reynolds (1979), Barber and Kevern (1974), and Jonasson (1955) provide information that may assist in the selection of a sampling period.

4.3.8.2 Sampling Season Considerations for the EEM Program

It is recommended that efforts be concentrated within a single seasonal sampling period, unless previous data indicate that there is more than one critical time period for the benthic community in the study basin. As this seasonal period should then be used in

subsequent studies, it is important to make this decision after compiling all available site-specific data regarding taxa life history characteristics and hydrologic discharge regimes.

Similarly, the sampling season for magnitude and geographic extent monitoring should be the same as previous monitoring unless, upon review of the previous results, there is scientific or logistical justification for a change. Furthermore, additional seasons may be warranted to help determine the magnitude of the response of the benthic community. For example, if the sampling is done at a time when the life stage of a particular invertebrate is not present, then an additional sampling season may be necessary to determine if effects are seen for this specific invertebrate. Bivalves, for example, are not easily sampled in the fall, which is often a critical period for many other invertebrates. In this case an additional season could be added to the monitoring program with the sampling program designed to answer this site-specific concern (i.e., an additional summer sampling trip where methods designed for bivalve sampling are used).

For most marine or estuarine areas the sampling season could be anytime from spring through mid-fall. For temperate marine environments, benthic sampling is usually conducted in late summer or fall as some benthic forms have planktonic larval stages that do not settle to the bottom until later in the season when populations with spring recruits have stabilized. For Arctic areas, the appropriate time period would likely be late summer or early fall, when the long day-length and warmer temperature have allowed some time for growth and development of flora and fauna and there is no sea-surface ice to contend with. In general, reproductive periods and patterns of abundance of benthic species are related to tidal cycles, season and abundance of food supply.

4.4 Statistical Considerations for Study Design

General statistical guidance (e.g., selecting α and β levels and determining sampling effort) is discussed in Chapter 8. This section provides specific guidance on benthic invertebrate statistics including sampling effort for RCA designs and the use of ordination probability ellipses for RCA designs. In addition, a discussion is included which provides guidance on determining the number of field sub-samples which should be taken at a given station and how this field sub-sample data could be used to improve future study designs.

It should be noted here that although a RCA can be used to present results of a benthic invertebrate community survey, mines should also submit information related to the effect endpoints required under the MMER (see section 4.9).

4.4.1 Determination of Sampling Effort for RCA Designs

The issue of replication is somewhat different when using the RCA. Replication is at the station scale and, since variation within a station is often much lower than among stations, single samples are taken at stations and variation among stations is used to

describe the reference condition. The number of reference replicates is determined by the number of stations in the group to which the exposed station is predicted to belong using the RCA. This is determined when forming the groups of reference stations in the initial classification (see section 4.3), but has been set to a minimum of 10 stations. The variation among the reference stations forming the reference group determines the Type I error, which has been set at 0.1 by using a 90% probability ellipse. Because, in this approach, single-exposure stations are compared to multiple reference stations (minimum of 10), it is not possible to set Type II error, which requires an estimate of the variance associated with a single station. A surrogate can be applied by taking more than one sample at the exposed station, but this is estimating within-station error rather than the appropriate variation at the among-station level. Clearly, Type II station error cannot be determined when there is only one member of the population of exposed stations. Therefore, the power analyses referred to above would not be applicable for the RCA study design. Consequently, RCA studies should be designed in a way that provides an accurate and precise determination of reference conditions so as to maximize the likelihood of detecting departures from reference conditions at exposure stations, when they exist.

4.4.2 Determination of Sampling Effort for Field Sub-sampling

The objective of multiple field sub-samples at each replicate station is to ensure that the sampling effort will produce an accurate reflection of all the metrics of interest (e.g., taxa richness, density) for each station that is sampled. This is necessary because species may not be homogeneously distributed throughout a station (which is much bigger than the size of the physical sampling apparatus being used). Inadequate sub-sampling 1) gives an imprecise estimate of the true mean for each station and 2) can contribute to an inflated estimate of the true among-station variance, thereby decreasing power.

Therefore, the allocation of field sub-samples within a replicate station depends on the following two inter-related factors that should be considered during any benthic sampling design exercise. However, in the absence of background information, the recommended minimum number of field sub-samples to obtain from each station is 3.

1. The abundance (or density) and degree of aggregation of organisms in relation to the desired level of precision for station estimates

For a given station, the number of field sub-samples needs to be sufficient to give a mean and variance that provide confidence that a representative number of animals has been captured (for a review, see Burd et al. 1990). The more aggregated a community, the higher the variance of mean abundance for each replicate station. Elliott (1977) and Holme and McIntyre (1984) suggested the same simple method of determining the number of field sub-samples to obtain a predefined level of precision. Elliott (1977) suggests that toleration of an index of precision (D) of 20% (i.e., that the standard error is equal to 20% of the mean) is acceptable for most bottom samples. The number of field sub-samples can then be calculated as follows:

$$n = s^2 / D^2 \bar{X}^2$$

where

$\bar{X}$ = the sample mean

n = the number of field sub-samples

s^2 = the sample variance

D = the index of precision (i.e., 0.20)

Thus, to determine how many field sub-samples (i.e., grabs) per replicate station will provide an estimate with 20% precision, previous data can be used to determine the mean and variance and, thus, the appropriate number of field sub-samples. This determination may vary from location to location along with changes in the mean-to-variance ratio. It is recommended that the number of field sub-samples be calculated for locations that exhibit the highest variability and that the resulting sample size be applied equally to all areas to standardize sampling effort. Although this recommendation will produce better precision in the less variable habitats, it is a conservative approach and maintains equal sampling effort between areas and replicate stations. Also of note is that, with aggregated populations, although the overall mean should remain the same, depending on the scale of the aggregation in relation to sampler size, variance will change with the size of the sampler. Therefore, sample size estimates using preliminary data are only relevant to a sampling program that would employ the same type and size of sampler with which the preliminary data were obtained. In cases where this sampling effort cannot be determined from a previous phase's data, counting organisms in field sub-samples from the current survey as they are processed and calculating means and variances will allow determination of how many grabs should be processed in the laboratory. However, this a posteriori approach necessitates that a sufficient number of grabs were obtained during the field survey in the first place, so the effort to calculate the sample size within a replicate station a priori should minimize problems due to insufficient sampling effort.

A related approach uses abundance and variance to determine sub-sampling effort and precision and can be used for determining the number of field sub-samples at all replicate stations. It is derived from the relationship between within-station mean abundance and variance across all replicate stations in the area or gradient being evaluated. Downing (1979, 1986) used Taylor's power law (1961) to estimate aggregation in a freshwater benthic community and thus determine the sampling effort required to reduce variance to an acceptable level. In effect, a given number of organisms are required for each replicate station in order to produce the precision of within-station mean abundance from sampling. Vezina (1988) used the same approach to determine empirically the degree of aggregation inherent in marine benthic communities. This approach entails calculating a power regression equation that describes the log/log relationship between within-station mean abundance and variance across all stations; this provides a formula that is then used to determine the estimated variance expected for a given abundance of organisms in that survey region. From this, the estimated variance for each mean abundance at each replicate station is calculated and then used in the same way as Elliott (1977) to estimate the number of field sub-samples at that replicate station. The difference between the

methods of Elliott (1977) and Downing (1979) is that in Elliot's method the variance used in the equation to determine the number of field sub-samples is based on sample variance, while in Downing's method the variance used in the equation is based on the variance calculated for the sample mean from the power regression equation for all the samples in the survey region. Furthermore, the index of aggregation (slope) from the power regression equation can then be used to determine the most appropriate data transformation for parametric statistical analyses. Unfortunately, this method assumes that the overall assemblage has a uniform aggregation throughout the study area, which may or may not be true when an external environmental stressor is applied. However, the degree of goodness of fit of the mean and variance data to the log regression equation provides a good indication of how true the homogeneous aggregation assumption is. If there are extreme outliers, they should be taken out of the analysis to avoid skewing the results. Because the aggregation of benthic communities can change as environmental conditions change either naturally or unnaturally, it is wise to review the relationship between mean and variance every time benthic samples are collected. Finally, it should be noted in the above discussion that the "power regression equation" used here to calculate the number of field sub-samples is unrelated to the "power analyses" used to determine the number of replicate stations discussed in the previous sections.

2. The number and distribution of different species in relation to obtaining a representative collection

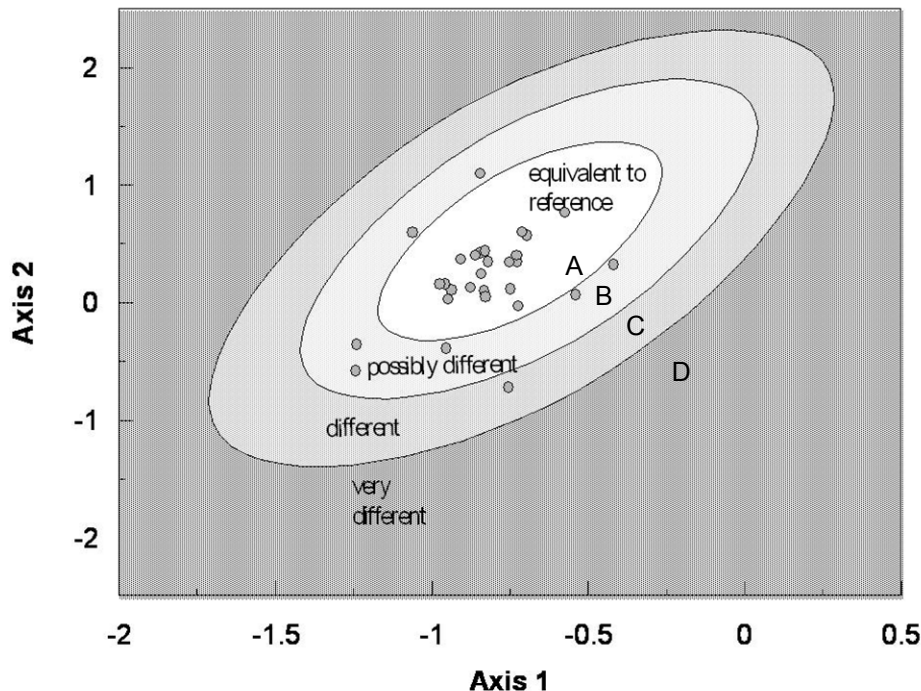
To determine if sufficient species have been sampled, simple rarefaction methods such as the "species abundance curve" or species/sampling area curve can be used (for a review, see Burd et al. 1990), which compare the number of species obtained vs. number of individuals for different numbers of pooled replicates. This analysis is particularly important in Arctic areas, where diversity may be high, but only on a geographic scale much larger than is feasible to sample (i.e., number of species relative to abundance is high, but abundance is quite low—this can also occur in the deep sea). Because of assumptions inherent in the underlying distribution of fauna related to logarithmic species abundance curves, a more sophisticated approach is the "similarity/sampling area" curve, which uses similarity indices on presence/absence data to determine the sampling effort to obtain an acceptable overall faunal similarity between replicate stations (Weinberg 1978; Kronberg 1987).

If preliminary data are unavailable or unsuitable for determination of the number of field sub-samples to obtain a representative collection of species, a check on sampling effort could be performed very simply. If it is estimated that X number of grab samples per replicate station is sufficient to achieve a data quality objective of retrieving 95% of benthic species present at any replicate station, more grabs can be collected at a few select replicate stations and analyzed. Determination of a taxa richness plateau from these extra samples determines whether the number of grabs were sufficient to achieve the 95% objective (using a species area curve).

4.4.3 The Use of Ordination Probability Ellipses for RCA Designs

A large-scale water quality survey on rivers conducted in the U.K. in 1990 provided the impetus for the development of methods to circumscribe the continuum of responses into a series of bands that represented grades of biological quality (Clarke et al. 1992). The study produced a simplification of the continuum of responses in sites ranging from good to poor biological quality. It was seen as an appropriate mechanism for obtaining a simple statement of biological quality, which allows broad comparisons in either space or time that are useful for management purposes. From a management perspective it is desirable to assign a degree of impairment. This can be done by setting response categories from mild to severe impairment. In the study by Clarke et al. (1992), a number of schemes for categorizing the response were considered and tested. The threshold between unstressed and stressed sites (band A) was set at the 90% probability level ($SD = 1.64$) for number of taxa and the biological monitoring working party (BMWP) score and 95% for the average score per taxon (ASPT). In Australia, the threshold is set at 2 SD from the reference site mean for the number of taxa. Finally, 95% is frequently set as the limit for determining a biological effect for univariate data and single community descriptors (Lowell 1997). The strategy employed in the U.K. (Wright 1995) to discriminate between degrees of impairment was to quantify the thresholds for stressed and non-stressed sites via the setting of 3 equal-sized bands, as Wright (1995) argued that there was no logical basis for an alternative scheme for dividing up the continuum of sites.

A similar approach can be adopted for defining degrees of impact using multivariate ordination. The reference invertebrate assemblage can be described by its distribution in ordination space, and the assemblage at any given site is characterized by its position in that XY space (Figure 4-5). The greater the similarity between sites, the closer together they are in XY space. Using this approach to set effect size for an invertebrate assemblage, all the reference sites are plotted in XY space together with a test site. The likelihood of the test site being the same as the reference site is quantified by constructing probability ellipses for the reference site *only*. Reynoldson et al. (1995) selected the 90% probability ellipse as representing the first band, the threshold for a site being considered equivalent to reference. The rationale for using the 90% ellipse rather than the more typical 95% was based on the fact that a multivariate approach will tend to be noisier than univariate measures and therefore a more conservative threshold was deemed appropriate. Sites located in ordination space inside this smallest ellipse (90% probability) would be considered as equivalent to reference and therefore unstressed. Two other probability ellipses are used (Figure 4-5), which are equal in width, to describe further divergence from the reference state, following the argument used by Wright and co-workers (Clarke et al. 1992; Wright 1995). Sites between the smallest (90%) and next ellipse (99% probability) would be considered possibly different; there is a 1 in 10 chance that sites will fall in this band through normal variability. Sites between the 99% and the largest ellipse (99.9% probability) are considered different: there is a 1 in 100 chance that these sites would incorrectly be described as different. And finally, sites located outside the 99.9% ellipse are designated as very different.



Note: Bands, based on 90, 99 and 99.9% probability ellipses, are identified as A (unstressed), B (possibly stressed), C (stressed) and D (severely stressed).

Figure 4-5: Impairment stress levels derived for reference sites in hybrid multidimensional scaling ordination space

4.5 Field Methods for Benthic Invertebrate Community Monitoring

4.5.1 Sampler Mesh Sizes

Benthic samples typically contain varying amounts of fine sediment and debris. To expedite transfer to sample containers, storage, and shipping, these samples should be reduced in the field by sieving. Field sieving should be done, wherever possible, immediately after sample retrieval and before preservation, as many organisms become fragile and brittle after preservation. Various techniques for sieving are available, but most involve washing the sample with a sieve or sieve bucket device.

The recommendation for sieve and/or mesh size for all freshwater mines is 500 μm .

In fresh water, macroinvertebrates are defined as those retained by mesh sizes of 200–500 micrometres (μm) (Slack et al. 1973; Weber 1973; Wiederholm 1980; Suess 1982), although immature life stages of some taxa may be smaller and some adult life stages may be larger.

Note that these mesh sizes are applicable to all equipment used in the field and laboratory (i.e., both the Nitex mesh on the benthic samplers and sieving apparatus).

In some site-specific circumstances it may be desirable for the field samples to be screened for smaller organisms by using a smaller sieve size (less than 500 μm). For example:

- 1) for comparative purposes, where historic benthic surveys for the system under investigation utilized smaller mesh sizes, or
- 2) if sampling needs to be conducted, for logistical reasons, at times when organisms are very small. Rees (1984), Barber and Kevern (1974), and Jonasson (1955) provide information on seasonal effects of mesh size.

In these aforementioned cases, it is highly recommended that a stack of screens be used which minimally have the mandatory sieve sizes and then any other smaller sizes, as appropriate. This procedure simultaneously allows site-specific concerns to be addressed and fulfills EEM objectives by allowing for national or regional comparisons to be conducted on the standardized mesh sizes. Sieving with the finest-scale sieve can be done in the field so long as the appropriate fractionation of the sample is performed in the laboratory before processing.

For **marine organisms**, samples should be sieved with seawater rather than freshwater, since the osmotic shock of freshwater may cause cell bursting and gross distortion of the animals. Where appropriate, field water used to sieve should be screened for ambient organisms with a mesh smaller than the required minimum screen size used for the study. In addition, extreme care should be taken during washing of samples to avoid breakage of specimens, which can greatly reduce taxonomic efficiency and cost-effectiveness. Methods have been described to reduce breakage, particularly in marine samples (Gray et al. 1990).

In marine systems, it is recommended to use a stacked set of 1000- μm and 500- μm screens in the field, with the 500- μm samples being archived and processed only if appropriate. Marine macrobenthos are typically those retained by sieves with 500–1000- μm mesh (Reish 1959; Thiel 1975; Pearson 1975; Holme and McIntyre 1984; Gray et al. 1990). It is estimated that a 1000- μm sieve will retain about 95% of the biomass of marine macrofauna (Reish 1959), while reducing the numbers of juvenile taxa and meiofauna present in samples that respond functionally differently to environmental perturbation than do adult macrofauna (Schwinghamer 1981, 1983; Warwick 1986).

Studying smaller benthic organisms for magnitude and geographic extent in marine systems may include assessment of meiofauna such as nematodes, copepods and smaller oligochaetes or it may include assessment of living and dead foraminifera (Schwinghamer 1981, 1983) or it may include more detailed assessment of juvenile forms of macrofauna. All of these approaches require the use of smaller mesh sizes and/or different samplers (cores may be more appropriate than grabs: see Holme and McIntyre 1984) than are currently recommended here. However, if smaller forms are

important, simply adding an additional sieve may not fulfill this function. The appropriateness of the sampling techniques should be assessed for smaller forms. For marine environments, Gray et al. (1990) noted that meiobenthos are most appropriately collected with core samplers, which are not recommended sampling devices for the EEM program. Thus, before simply screening for smaller organisms, appropriate protocols should be implemented.

4.5.2 Sampling Equipment

Two major considerations for benthic surveys are mesh size (see previous section) and quantitative sampling equipment. Quantitative sampling of benthic communities is carried out using devices that sample a known area or volume of habitat, such as grab samplers or stream net samplers. Each sampling device should be non-selective and suitable for a particular substrate. Benthic samples collected from natural substrates provide an indication of past and current stressors. Therefore, samplers that collect benthic communities from the bottom sediments are recommended unless this is not possible due to physical constraints. Samplers are to be consistent within a habitat class among all stations and areas. However, different samplers may be used in the same survey if they are used to sample different habitat classes. For example, if both erosional and depositional habitat classes are sampled, it would be reasonable to use one of the recommended grabs for the exposure and reference depositional habitat, but a Hess-type sampler for the exposure and reference erosional habitat. It is recommended that grab samplers with screens on the top and top-opening gates be used so that the bow-wave ahead of them is reduced and less substrate is lost, and for examination (and possibly sediment chemical analysis) of undisturbed surface layer in sediment samples.

Standardization of benthic samplers facilitates regional and national comparison of benthic invertebrate survey data. Recommendations of samplers appropriate to the various habitat classes encountered during EEM benthic surveys are provided below. Eleftheriou and Holme (1984), Klemm et al. (1990), and Scrimgeour et al. (1993) discuss the options as they pertain to different receiving environments and summarize the advantages and disadvantages of the recommended samplers. The selection of a sampler may also be influenced by the type used in previous surveys of particular systems. To ensure that surveys can be compared with previous historical surveys, it would be useful to use similar sampling equipment. For more detailed information, the reader is referred to the bibliographies on quantitative samplers and appropriate methodologies prepared by Klemm et al. (1990), Eleftheriou and Holme (1984), Elliott and Tullett (1978, 1983), Rosenberg (1978), Downing (1984), and Mason (1991). See also Rabeni and Gibbs (1978) and Alberta Environment (1990).

The standardization of techniques applies not only to the sampling equipment but also to the level of expertise required to correctly deploy the sampler. Crew members should be properly trained in the use of sampling equipment to minimize variation introduced by operator error. For example, when sampling erosional zones in rivers, the depth to which the substrate is disturbed within the net-sampler should be standardized since some individuals may be more energetic in regards to stirring up the substrate than others. The

study leader should be well versed in benthic invertebrate sampling and conduct effective training sessions with the crew members performing the field sampling. If training is done effectively, operator error can be eliminated (Reynoldson and Rosenberg 1996).

Depositional habitats: freshwater

Grab samplers are devices with spring-loaded or gravity-activated jaws that “bite” into unconsolidated substrates (sand, silt, mud, etc.) to enclose a defined surface area of the bottom. These devices are generally lowered on a line or cable from a survey vessel to the bottom, sometimes with the aid of a winch. If the sampler type chosen is not suited to the substratum present, it can affect sampling efficiency. Factors that may affect grab sampling include depth of penetration, completeness of closure of jaws, and subsequent loss of material during retrieval. In depositional zones of freshwater rivers or lakes, Ponar or Ekman grabs are suggested as standard samplers for EEM benthic invertebrate surveys. See Eleftheriou and Holme (1984), Klemm et al. (1990), and Scrimgeour et al. (1993) for additional information on samplers.

Erosional habitats: freshwater

Stream-net samplers are devices used for collecting benthic invertebrates in erosional riverine environments. They use mesh of various sizes (but see Section 4.5.1 for discussion on mesh size) to sieve organisms from water flowing through the mesh after disturbance of a known area of the substrate. It is recommended that erosional habitats in freshwater environments be sampled with Neill-Hess cylinder-type samplers that allow unit area (typically 0.1 m²) estimates to be made. One drawback to cylinder samplers in streams is a potential incompatibility with size of substrate. In some systems, mean particle size may be too great for the Neill-Hess cylinder to effectively sample the benthic invertebrates. In such cases, a U-net sampler (Scrimgeour et al. 1993) can provide area-limited samples and be adjusted accordingly to the size of the substrate. This sampler has been used successfully for a range of substrate sizes (Glozier 1989) and can sample either individual stones or a defined area. Kick-net samplers do not provide an area-delimited estimate, but have been used widely in the United Kingdom, the United States, Australia and Canada in large-scale monitoring programs (Reynoldson et al. 1995). Kick sampling is particularly appropriate for the reference condition approach, where many stations are sampled. A timed kick sample is taken at each station to estimate benthic community descriptors. Standardization of kick-sampling techniques is essential for comparative purposes and can be obtained with minimal training (Reynoldson and Rosenberg 1996). The kick-sampling method involves a single composite sample collected at each station by a 3-minute travelling kick method (Reynoldson et al. 1997). Note that separately preserved field sub-samples are not required for the kick-sampling technique recommended for RCA.

For difficult habitats (e.g., very deep, slow-flowing areas or areas with hard substrates) alternatives such as the metal quadrat or airlift system may be available. However, for national or regional comparative purposes, the list of recommended samplers should be

sufficient for sampling the majority of ecologically relevant habitats. If habitats are extremely difficult to sample, alternative approaches may be considered.

4.5.3 Artificial Substrates

The use of artificial substrates for benthic invertebrate collections is generally not recommended as a sampling protocol in EEM.

There is no advantage to be gained from using artificial substrates where conventional sampling techniques provide at least as reliable data without the many drawbacks and difficulties of artificial substrates (AETE 1995). Artificial substrates do not collect a representative sample of the indigenous benthic invertebrate community at the site where they are placed, but rather select for mobile, drift-prone species of hard substrata. In addition, artificial substrates do not effectively monitor the effects of sediments or sediment-bound contaminants on aquatic biota because sediment-dwelling taxa tend to be under-represented in artificial substrate samples. The invertebrate community represented by artificial substrates indicates conditions during the period of exposure only and does not integrate long-term effects. Therefore, the use of artificial substrates for benthic invertebrate collections may fail to indicate the effects from effluents, particularly where non-mobile species, sediment-bound contaminants or longer-term integration of effects are important. However, it is recognized that there may be a limited number of cases where there is either a long history of artificial substrate use in a particular ecosystem or extreme habitat conditions (e.g., very deep, fast-water systems) where the use of artificial substrates is the only feasible *field* method available. In these cases, the use of artificial substrates may be considered along with other alternatives—provided this method can determine if there are effects on the benthic invertebrate community in a scientifically defensible manner.

4.5.4 Marine/Estuarine Habitat Sampling Equipment

Depositional Habitats: Marine/Estuarine

Depositional habitats in marine environments can be sampled with the Smith-McIntyre grab, a modified Van Veen grab, which is suitable and available in Canada. A good review of marine sampling methods is available in Eleftheriou and Holme (1984). However, in shallow subtidal areas where there is not enough water depth to allow the deployment of the larger grabs, a smaller or mini (petite) Ponar grab can be used. This grab is deployable from small inflatable boats and can be retrieved by hand.

Intertidal soft substrates may be sampled using any device that demarcates an area of at least 0.1 m². The soft substrate is then removed to a standard depth of 10 cm using an appropriate device. Note that, in general, the lowest intertidal level available for sampling is preferred because less harsh physical conditions promote higher species richness and abundance.

Erosional Habitats: Marine/Estuarine

In marine/estuarine environments, large unconsolidated sediments such as gravel may be sampled with grab samplers. If not, hard substrates in erosional habitats (intertidal and subtidal) should be sampled using quadrats with a minimum area of 0.1 m². However, some other quantitative techniques may be recommended to collect marine shellfish and other large species. These may include hand collection by divers, remote sensing techniques from defined surface areas (Eleftheriou and Holme 1984; Gray et al. 1990), and collection from defined boundaries along transects. An outline of a marine sampling protocol is described in a series of *Puget Sound Estuary Program Reports* (Tetra Tech 1986a, 1986b, 1987). When done properly, photographic surveys can be quantitative, at least for larger epibenthic organisms (c.f. Burd et al. 1990). Processing costs tend to be considerably less than for soft-bottom surveys using grabs or cores.

The intertidal zone should be sampled if the effluent plume impinges substantially on it. Determining the tidal level of greatest interest for examining mine impacts will involve logistical considerations. Basically, the lower in the intertidal area the surveys can be conducted, the better, since less harsh conditions create less patchiness and higher diversity in flora and fauna (inter-sample variability). Coastal plants and animals in this habitat typically exhibit vertical distributions that reflect gradients in environmental parameters such as air exposure, temperature (including freezing), salinity, light intensity and daylength, abrasion due to logs or ice, and wave shock. These gradients should be considered in planning and undertaking biological surveys in the intertidal environment. Sampling protocols for this area will be somewhat different from those described in the earlier sections (for review see Gray et al. 1990). Wherever possible, semi-quantitative surveys using quadrat areas of 0.1 m² should be done. Determining the substrate or habitat type to be sampled depends on sampling limitations and the dominant habitat present (see section 4.3.7 for discussion of dominant habitat selection). However, if it is not feasible or ecologically sound to collect samples, then visual surveys are recommended. If approved, a visual survey would include approaches such as recording and mapping (at a gross scale of 1:5000) the major biological features for assessment of gross changes in the biological community.

4.5.5 Sample Containers

Environment Canada's *Guidelines for Monitoring Benthos in Freshwater Environments* (EVS Environment Consultants 1993) specify that sample containers should:

- be large enough to ensure the sample takes up no more than 50% of the container volume, with the remainder of the space allocated for preservative;
- be sturdy enough for routine handling and transportation;
- be leak-proof;
- have physical and chemical properties that are not affected by the fixative/preservative; and
- conform to regulations concerning the transportation of dangerous goods.

4.5.6 Specimen Fixation and Preservation

All samples should be fixed in the field in a 10% buffered formalin solution to prevent damage to freshwater and marine worms. Formalin is also important for the proper preservation of most aquatic insects. Preservation directly in ethanol often results in soft, difficult-to-handle specimens. After preservation in the field, samples should be gently mixed several times to ensure that the preservative has thoroughly penetrated any fine material that may be present in the sample. Because formalin is a carcinogen and an irritant to workers, gloves and protective eye gear are needed and should be considered mandatory safety equipment. Furthermore, unbreakable sample jars should be sealed with parafilm, double-bagged for transport back to the laboratory facilities and adequately labelled. The samples should be preserved as soon as is practical after sampling to prevent predatory invertebrates from preying on others in the samples.

4.5.7 QA/QC for Benthic Invertebrate Field Operations

An outline of the quality assurance/quality control (QA/QC) recommendations for the field components of the benthic invertebrate community survey is presented below.

Field sampling is the first stage of data collection. QA/QC procedures for the benthic invertebrate survey are outlined in the study design and should be followed precisely to maintain high data quality. Field standard operating procedures (SOPs) should specify sampling equipment and protocols appropriate to the study. A QA/QC plan for field sampling has many components. Some of the main procedures are listed below:

- 1) All personnel involved in the field sampling should have appropriate training and experience with field equipment and objectives.
- 2) All safety measures should be identified, understood and adhered to.
- 3) Collection equipment should be appropriate for the specific water body and selected invertebrate group, and should be checked frequently and maintained regularly.
- 4) There should be some a priori criteria for acceptability of samples obtained and clear directions if acceptability guidelines fail (i.e., when to retake a sample; grab sample penetrations of 10-cm depth would be considered an acceptable sample, Gray et al. 1990). Also, sampling methods need to be consistent throughout the study.
- 5) A visual description of benthic grab samples should be recorded to describe sediment color, odour, texture and debris.
- 6) Contamination during chemical sampling should be checked by means of trip blanks and equipment rinsates.
- 7) Field sieving, if necessary, should be done as soon as possible after retrieval of samples.
- 8) Samples should be stored in appropriate containers with appropriate preservative to prevent breakage and spoilage.
- 9) All sample containers should be appropriately labelled.
- 10) Detailed field notes should be maintained in a bound waterproof notebook.
- 11) Chain-of-custody forms and appropriate shipping and storage procedures should be applied.

For further information regarding all aspects of QA/QC procedures for benthic invertebrate programs, refer to the 1999 AETE report (Beak 1999).

4.6 Laboratory Methods

For information pertaining to sample sorting and sub-sampling, please refer to the *Revised Guidance for Sample Sorting and Subsampling Protocols for EEM Benthic Invertebrate Community Surveys*, which can be obtained from the EEM website (<http://www.ec.gc.ca/esee-eem/default.asp?lang=En&n=B9DBF4CC-1>).

4.6.1 NABS Certification Program

The accurate identification of aquatic benthic invertebrates is crucial to monitoring programs like the metal mining EEM program. The North American Benthological Society (NABS) implemented a certification program for benthic invertebrate identification. The program tests the candidate's knowledge and skills in aquatic invertebrate taxonomy and ensures that individuals are providing high-quality identifications. It is recommended that the identification of aquatic benthic invertebrates be conducted by an individual who has completed the NABS certification program. For additional information, please refer to the following website: <http://www.nabstcp.com/>.

4.6.2 Taxonomic Level of Identification

Identification of the benthic invertebrates sampled should be adequate to meet the objectives of the assessment program. Research indicates that family-level identification provides sufficient taxonomic resolution to detect community responses to human disturbances (Warwick 1988a, 1988b; Bowman and Bailey 1997). As discussed below, the level of taxonomic resolution used may vary across the different monitoring phases, with finer taxonomic resolution needed to detect more subtle environmental impacts.

The recommended level of taxonomic identification is family for the first and subsequent monitoring of freshwater systems. All summary statistics and descriptive metrics should be calculated and reported at the family level for submission to the first and subsequent monitoring interpretative reports. Organisms that cannot be identified to the desired level of taxonomic precision should be reported as a separate category in the fundamental data set. It is recommended that investigators use taxonomic keys appropriate to the geographic region of study. Table 4-2 lists taxonomic references typically used for various groups of freshwater organisms.

For some phases, a lower taxonomic level may be recommended, depending on the questions and objectives of the study. The lowest practical level (LPL) has been defined as genus for most insects and the lowest level possible without special procedures (dissection, microscopy) or reliance on specialist for other groups (Taylor 1997). This

definition can be used as a guide if lower level identifications are desired for magnitude and extent and investigation of cause monitoring.

There may be site-specific conditions that warrant a lower taxonomic level for some or all familial groups. For example, historic benthic invertebrate information may be identified to lower taxonomic levels, and it may be desirable to identify subsequent surveys to a similar level for comparative purposes. If a lower taxonomic level has been used, either in historic data or during a current survey, the summary statistics and descriptive metrics can be reported at this level, provided a summarized data set at the family level is also included.

Two objectives of the magnitude and geographic extent survey may require different levels of taxonomic resolution. Determination of the geographical extent of the effect may be addressed adequately with *family* identification. Family-level identification would provide the information necessary for calculating and reporting the required summary and descriptive statistics related to extent of the effect. This first objective is similar in scope to Phase 1 and subsequent monitoring, the major difference being the addition of low effluent exposure areas further from the effluent discharge.

The second objective of magnitude and geographic extent monitoring, when determining the magnitude of the effect, may use family identification or it could warrant investigation at a lower taxonomic level. The question of magnitude of effect with regard to taxonomic level can be addressed using the following question:

- What is the magnitude of the effect on specific taxonomic groups that may be sensitive to the site-specific mine effluent characteristics (e.g., how many groups within a sensitive family are affected)?

Addressing the magnitude of effect during magnitude and geographic extent monitoring can be accomplished by using one of the options outlined below:

- Identify all samples collected to the lowest practical level. Establishing the magnitude of effect in this way provides additional information that may be useful for the study design exercise at the outset of the investigation of cause.
- Re-analyze families that were significantly affected during the first monitoring to identify indicator taxa that can be used to assess the magnitude of effect at stations farther afield. For example, if an effect during the first monitoring was observed for the family Baetidae (order Ephemeroptera), all Baetidae could be identified to a lower level (e.g., genus) for the magnitude and geographic extent monitoring program. This approach would catalog the “sensitive” taxa within the family, and the magnitude of effects would be established by examining this subset of sensitive taxa.
- Other scientifically defensible approaches may be used to identify magnitude of effect as required.

In **marine/estuarine** environments, it is recommended that all benthic invertebrate organisms be identified to the *family* level. In interpretative reports, all summary statistics are calculated and reported to the family level. Various authors have examined the utility of using higher taxonomic classifications for environmental monitoring of organically polluted sites in Europe (cf. Warwick and Clarke [1993] and references therein). For marine benthos, juvenile or non-adult fauna should be identified and enumerated separately from adults, as they show different patterns of response to environmental effects.

Though mines may proceed with benthic invertebrate identifications to a lower level, the recommended level of identification for data reporting and determination of effects is the family level. There may also be site-specific conditions that warrant a lower taxonomic level for some or all familial groups. For example, historic benthic invertebrate information may be identified to lower taxonomic levels and it may be desirable to identify subsequent surveys to a similar level for comparative purposes. If a lower taxonomic level has been used, either in historic data or during a current survey, the summary statistics and descriptive metrics can be reported at this level—provided a summarized data set at the family level is also included.

For marine samples it is suggested that, if sufficient numbers of specimens are available in the reference collections, they could be used for a further purpose: to develop a size and biomass database for each mine as another indicator or tool (see section 4.11.4). For these purposes, 5 to 10 representative specimens per taxa are recommended, with mean width, lengths and blotted wet weights recorded for each group.

4.6.3 Reference Collections

Consistency in taxonomic identifications within and between surveys is essential to obtaining useful information on environmental effects monitoring. Therefore, for comparative purposes and quality control of taxonomic identification, the maintenance of a reference collection of organisms is recommended. In addition, it is recommended that an independent professional taxonomist verify the identifications in the collection. Museums are sometimes prepared to perform this service when remote areas are included in the study and new specimens or distribution records are likely. Reference collections have several benefits, including their use in confirming identifications, ensuring consistent taxonomy between surveys, and the training of personnel. Protocols for establishing and maintaining reference collections for benthic invertebrates are detailed in a report prepared for Environment Canada's Fraser River Action Plan (Green 1994). Following the recommendations of this report, each mine (or group of mines) should compile and archive a complete reference collection with several specimens of representative-sized individuals for each taxon. The collection should encompass representative organisms from each area in the survey, be labelled according to the location and date of collection, and updated as appropriate (i.e., when a taxon is collected). This type of reference collection will not occupy a large space: a small cupboard should be sufficient and should be in the custodianship of the mine. If a mine does not have the facilities or personnel to maintain their own reference collection,

universities or museums may be willing to fulfill this function. However, since considerable effort is involved in the long-term maintenance of preserved biological material, the quantity of material submitted should be minimized.

Table 4-2: Taxonomic keys for benthic invertebrate taxonomic identification in freshwater environments

Taxon	Taxonomic Reference Typically Used
General Keys	Merritt and Cummins 1984, 1996; Peckarsky et al. 1990; Pennak 1978; Thorp and Covich 1991
Regional Keys	Clifford 1991 (Alberta) Lehmkuhl 1975a, 1975b, 1976, 1979 (Saskatchewan) Laplanche et al. 1991 (Quebec)
Taxon-specific Keys	
ANNELIDA	
Oligochaeta	Brinkhurst 1986
Hirudinea	Klemm 1972, 1985, 1991
Crustacea	Bousfield 1958
Amphipoda	Brandlova et al. 1972
Decapoda	Dussart 1969
Cladocera	Crocker and Barr 1968
Copepoda	Fitzpatrick 1983
INSECTA	Chu and Cutkomp 1992; Hilsenhoff 1995
PLECOPTERA (stoneflies)	Fullington and Steward 1980; Harper and Stewart 1984; Hitchcock 1974; Stewart and Stark 1993
EPHEMEROPTERA (mayflies)	Bednarik and McCafferty 1979; Edmunds et al. 1976; Lewis 1974; Morihara and McCafferty 1979; McCafferty and Waltz 1990; Waltz 1994
ODONATA (dragonflies and damselflies)	Hilsenhoff 1995; Westfall and May 1996; Walker 1933, 1953, 1958; Walker and Corbet 1978
TRICHOPTERA (caddisflies)	Scheffer and Wiggins 1986; Schuster and Etnier 1978; Wiggins 1996
COLEOPTERA (beetles)	Hilsenhoff and Schmude 1992
DIPTERA (flies)	Hilsenhoff 1995; Johannsen 1977; Oliver et al. 1978; Saether 1975, 1977; Simpson and Bode 1980; Wiederholm 1983, 1986; Wood et al. 1963
GASTROPODA (snails)	Burch 1989; Clarke 1981
PELECYPODA (clams, mussels)	Mackie et al. 1980; Clarke 1981; Burch 1975a, 1975b; Mackie and Huggins 1983

Table 4-3 lists recommended levels of taxonomy desirable for major taxonomic groups of marine benthic organisms. In general, the level of taxonomy should be consistent in each major group for all samples from a survey and also from survey to survey. Organisms that cannot be identified to the desired level of taxonomic precision should be reported as

a separate category in the fundamental data set at the finest level of taxonomic resolution possible. Since the accuracy of the taxonomic work depends on the availability of up-to-date taxonomic literature, a basic library of identification keys is essential. Keys appropriate to the geographic region of study are recommended. A detailed list of taxonomic references for marine and estuarine habitat is found in Table 4-4. Microscope slide mounts should be prepared for taxa requiring detailed microscopic examination for identification. This may involve various steps, including dissection, clearing and staining. Slide preparation techniques are listed in Klemm et al. (1990). For marine benthos, juvenile or non-adult fauna should be identified and enumerated separately from adults, as they show different patterns of response to environmental effects. All identifications should be carried out or verified by a qualified and experienced taxonomist. Existing reference collections may be useful as well. An example is the Atlantic Reference Centre at Huntsman Marine Station in St. Andrews, New Brunswick. Photographic iconographs have been used to advantage (Camburn et al. 1984–1986).

Table 4-3: Recommended level of taxonomic precision for benthic invertebrates in marine environment (for lowest practical taxonomic level approach)

Taxon	Level
Porifera	Class
Cnidaria	Genus
Turbellaria	Genus
Nemertea	Genus
Nematoda	(not to be included in analyses*)
Sipuncula	Species
Priapulida	Species
Brachiopoda	Genus
Bryozoa	Family
Mollusca	
• Aplacophora	Genus
• Gastropoda	Species
• Bivalvia	Species
• Polyplacophora	Genus
• Scaphopoda	Species
Annelida	
• Polychaeta	Species (except some immature)
• Oligochaeta	Genus
Arthropoda	
• Pycnogonida	Family
• Cephalocarida	Sub-class
• Malacostraca	Species
• Copepoda	(remove from analyses*)
• Cirripedia	Species
Ascidiacea	Family
Echinodermata	Species

*Nematodes and copepods (e.g. harpacticoida) are meiofauna, and only a fraction of specimens will be captured by a 500 µm or 1000 µm screen. Therefore, numbers are not representative and should be excluded from analyses (Holme and McIntyre 1984).

Table 4-4: List of marine and estuarine taxonomic benthic invertebrate keys for Canada

References

- Abbot 1974 (seashells)
 Abbott et al. 2001 (Mollusca)
 Appy et al. 1980 (Bay of Fundy polychaetes)
 Austin 1985 (Pacific invertebrates)
 Baker 1980 (Tubificid species)
 Banse 1972; Banse and Hobson 1974 (polychaetes)
 Berkeley and Berkeley 1952 a,b (Pacific Annelida)
 Blake 1971 (Polydora, East Coast)
 Blake 1991 (Polychaeta, North Atlantic)
 Blake 1988 (Phyllodocidae [Polychaeta], Atlantic)
 Bousfield 1960 (Atlantic seashells)
 Bousfield and Hendryks 1994, 1995 a, 1995b (Pacific Amphipoda)
 Bousfield and Hoover 1995 (Pacific amphipods)
 Bousfield and Kendall 1994 (Pacific amphipods)
 Bousfield 1973 (Amphipoda, Atlantic)
 Brinkhurst 1982 (Oligochaetes)
 Brinkhurst and Baker 1979 (Marine Tubificidae) (Oligochaeta)
 Brunel et al. 1998 (Catalogue of the invertebrates of the Gulf of St. Lawrence)
 Butler 1983 (Pacific shrimps)
 Clark 1924 (Holothuroidea)
 Clark 1915 (Ophiuroidea)
 Coates 1980 (B.C. Enchytraeidae)
 Coe 1912 (Echinodermata, Atlantic)
 Coe 1943 (Nemertea, Atlantic)
 Cutler 1973 (Sipuncula)
 Fauchald 1977 (Polychaeta)
 Fournier and Petersen 1991 (Polychaeta)
 Gibson 1994 (Nemertea)
 Gosner 1971
 Graham 1988 (Gastropoda)
 Hart 1982 (B.C. crabs)
 Hobson and Banse 1981 (B.C. polychaetes)
 Hyman 1940 (Polycladida [Turbellaria, Atlantic])
 Hyman 1944 (Turbellaria, Atlantic)
 Keen and Coan 1974 (Mollusca)
 Knight-Jones 1978 (Spirorbidae [Polychaeta], Pacific and Atlantic)
 Knight-Jones 1983 (Sabellidae [Polychaeta])
 Kozloff 1987 (Pacific N.W. invertebrates)
 Lambert 1981 (B.C. sea stars)
 Laubitz 1972 (Caprellidae)
 Light 1977 (Spionidae [Polychaeta], Pacific)
 Morris 1951 (Mollusca, Atlantic)
 Pettibone 1963 (Polychaeta, Atlantic)
 Pettibone 1992 (Pholoidea, Polychaeta)
 Pettibone 1993 (Polynoidae, Polychaeta)
 Pohle 1990 (Decapoda, Atlantic)
 Sars 1895 (Amphipoda)
 Sars 1899 (Isopoda)
 Sars 1900 (Cumacea)
 SBMNH 1994a,b,c; 1995a,b,c; 1996a,b,c; 1997a,b
 Schultz 1969 (Isopod crustaceans)
 Smith 1964 (Marine invertebrate keys, Atlantic)
 Squires 1990 (Decapoda, Atlantic)
 Steele and Brunel 1968 (Amphipoda)
 Tattersall and Tattersall 1951 (Mysidacea)
 Thorp and Covich 1991 (Freshwater invertebrate keys)
 Ushakov 1955 (Polychaeta)
 Wallace 1919 (Bay of Fundy Isopoda)
 Watling 1979 (Cumacea, Atlantic)
 Weiss 1995 (Marine macrofauna)

4.6.4 QA/QC for Benthic Invertebrate Laboratory Operations

In the laboratory, invertebrate samples are processed and counts of the various taxa are made. It is recommended that the sorted, preserved samples from each survey be retained in an appropriate storage facility for at least 6 years, or until it is determined that no further information will be required from the samples. Samples should be processed in a consistent manner to minimize experimental error in counts. To minimize processing error, the following items should be included in the QA/QC program:

1. All personnel involved in the sample processing and analyses should have appropriate training. NABS implemented a certification program for aquatic invertebrate taxonomists. For additional information see <http://www.nabstcp.com/>.
2. The effects of sub-sampling (if done) on abundance estimates should be examined on a minimum of 10% of the samples, and the effects of sub-sampling on the sample estimates should be documented.
3. Re-sorting of randomly selected samples should be done to determine the success of the initial sorting (see detailed discussions below).
4. Appropriate taxonomic references should be used for the type of habitat and geographic location.
5. A complete reference collection for each mine should be compiled and verified by an external taxonomic expert and updated as appropriate (i.e., when new taxa are recorded).
6. A system for archiving samples should be outlined.
7. Detailed sample processing and laboratory notes should be maintained.

Ecological sample processing involves, as a first step, sorting organisms from debris and, possibly, sub-sampling sorted organisms for detailed identification. Inevitably, processing errors are associated with these activities and should be estimated (e.g., Kreis 1986, 1989).

4.6.4.1 Sorting Efficiency

Verification of sorting efficiency is easily performed on a spot-check basis if the leftover debris from a sample is retained. It is recommended that at least 10% of all samples be re-sorted and that the criterion for an acceptable sort be that $\leq 10\%$ of the total number of organisms were missed. This estimate should be reported in the interpretative report. If $\geq 10\%$ of the total number were missed during the re-sort, then all the samples within that group of samples should be re-sorted.

A re-sort would also be required if an entire group of benthic invertebrates was missed by the sorter (i.e., not recognized as an organism), even if the missed organism constituted $< 10\%$ of the total. The factors to consider when determining similar groups of samples include: 1) sampling area, 2) habitat class and 3) individual sorters. The QA/QC guidelines apply independently to each group of samples sorted. Sorted and unsorted fractions are to be retained until taxonomy and sorting efficiency are confirmed.

4.6.4.2 Sub-sampling

Sub-sampling of invertebrate samples in the laboratory is acceptable, providing that the quantitative method is used. Large samples or samples with large amounts of sediment debris may require laboratory sub-sampling prior to sorting. Readers are referred to the *Revised Guidance for Sample Sorting and Subsampling Protocols for EEM Benthic Invertebrate Community Surveys* (Environment Canada 2002), which can be obtained from the EEM website (www.ec.gc.ca/eseee-eem). The detailed reporting of sub-sampling accuracy and precision for all methods is essential to the QA/QC of EEM benthic invertebrate programs. The criterion for an acceptable sub-sampling protocol is that the estimates of each group of samples should be within 20% of the true counts. If the error exceeds 20% for a particular sub-sampling technique or type of samples (i.e., type and amount of organic matter), the technique should be modified to achieve this level of precision, or all samples within that group should be completely sorted to ensure the sub-sampling process is not compromising data integrity. The estimates are then compared to the actual counts from the sample, and the accuracy of the estimates and the precision between sub-samples can be calculated using the following equation:

$$\% \text{ error in the estimate} = [1 - (\text{estimated \# in sample} / \text{actual \# in sample})] \times 100$$

The accuracy should be reported in the interpretative report.

It is recommended that a minimum number of 300 organisms be removed from a sample in any sub-sampling program to provide additional standardization. If any sampling stations have not reached the recommended minimum number of organisms during sub-sampling (i.e. 300) or have poor accuracy, the sample should be flagged when reported.

For further information regarding all aspects of QA/QC procedures for benthic invertebrate programs, readers are referred to the 1999 AETE report (Beak 1999).

4.7 Data Assessment and Interpretation

4.7.1 Data Handling Methods

4.7.1.1 QA/QC for Data Input and Verification

After data entry, the first step in data analysis is to check for transcription errors. Failure to do this invalidates further analyses. All computer entries should be verified by checking a hard copy of the file against the raw data sheets. Someone other than the person who originally entered the data should do this cross-checking. Double entry systems and transcription checks against the original data records are useful QC techniques. Missing data should be clearly distinguished from taxon absence by use of unique non-zero missing value codes with code definitions built into each file. Read-only

files help to ensure data integrity. QA/QC concerns regarding data analysis include data verification and validity, repeatability and robustness of statistical analyses, and rigour and defensibility of analysis. EVS Environment Consultants (1993) suggest that other investigators should be able to arrive at the same conclusions if they were to use the methods and data set found in the report. Other considerations regarding the data verification and analyses are listed below:

- 1) Use trained and experienced personnel.
- 2) Conduct screening exercises to identify transcription errors, outliers and other suspicious data points.
- 3) Provide raw data in an electronic database format and appendices to reports that summarize the data.
- 4) Document the methods (specific statistical tests) and software (if applicable) used for analysis.
- 5) Maintain detailed notes regarding the analyses of the survey data.

For further information regarding all aspects of QA/QC procedures for benthic invertebrate programs, readers are referred to the 1999 AETE report (Beak 1999).

4.7.1.2 Dealing with Outliers

Assuming the data are entered correctly, data should be summarized, screened for erroneous values and outliers, assessed for normality, and transformed if necessary (EVS Environment Consultants 1993). Visual screening techniques such as box-and-whisker plots, normal-probability plots and stem-leaf diagrams can be used to identify extreme values (true outliers and/or data entry errors) (see Tukey 1977). Norris and Georges (1993) recommend examining abundance estimates for each taxon to determine if numbers are reasonable. They also recommend calculating means and standard deviations because aberrantly high or low values can indicate errors. Extreme values or outliers that are not errors of some kind should not be removed from the data set because this will result in the loss of an observation and a loss of power to the benthic invertebrate community survey. Instead, extreme values should be identified in the report and the influence of the extreme value on the results should be determined by reanalyzing the data minus the extreme value.

4.7.1.3 Unknown, Immature and Non-benthic Organisms

There have been several instances where non-benthic organisms have been submitted as part of the metal mining EEM program. If it is documented that a given family of organisms can at some point become benthically attached (e.g., *Simocephalus*), then it is acceptable to include the organism within the benthic invertebrate community. However, species such as planktonic *Daphnia* should be removed from the data set.

Some samples may contain immature individuals that cannot be identified to the recommended level of taxonomic precision. A similar situation could also occur when samples are improperly preserved and identifying features are destroyed (e.g., mollusc

shells dissolve due to unbuffered formalin). For the purposes of correctly reporting the raw data, these unidentified taxa and their abundances should be provided within the electronic raw data and report appendices. However, for data analysis, investigators need to decide whether or not to apportion the unknown individuals according to the ratio of known specimens. This assumes that the ratio of unidentified specimens is similar to the ratio of identified specimens, which may or may not be true. The choices include:

1. not incorporating immature or damaged forms at all
2. pooling all specimens (i.e., mature/immature, identified/unidentified) and lumping them into one category at the next highest taxonomic level
3. keeping unidentified taxa as a separate category in the analysis

Option (1) is not preferred if the “problem” taxa represent a large proportion of the total benthic invertebrate community. Option (2) assumes that all taxa within a higher taxonomic level respond the same way to effluent-related stressors, which may or may not be true. Option (3) will have variable effects on data interpretation depending on the abundance of unidentified taxa. Whatever choice is made will depend on the expertise and experience of the individual investigator; however, it should be fully documented in the Methods section of the interpretative report.

For **marine** surveys, it is recommended that immature and juveniles be counted and enumerated separately from adults, whether or not they can be identified to the species level, so that the adult assemblage can be analyzed without the confounding influence of transient juveniles. Thus, data analyses should show results both with and without immatures included. This is because newly settled benthic forms have different survival characteristics than adults, which have been present in the sediment much longer and integrate the effects of habitat perturbations over time. Depending on the timing of sampling, newly settled juveniles may be abundant in samples, but may all die within days due to habitat stressors, predation or competition. This is not to say that data on immatures are not important. Dramatic variations in immature settlement between nearby samples within physically homogeneous habitats may be indicative of varying levels of stress. It is just important to avoid confounding the results by mixing groups together for analysis.

4.7.1.4 Data Reduction and Transformation

Data transformation is often performed without consideration of the effects it has on the interpretation of results. For general information on transformations, see Chapter 8. Transformation should only be applied with a complete understanding of its effect on the data and their interpretation, and only if it is necessary to aid in statistical analyses. Transformations should:

1. make heterogeneous variances homogeneous or make the variance independent of the mean for parametric analyses
2. normalize distributions
3. linearize relationships among variables

4. reduce the effects of extremely dominant taxa within a data set on a multivariate analysis (or ordination)
5. reduce the analytical problem of too many zeros in a data matrix (see Clarke and Green 1988).

Data reductions should be done only to aid in statistical or multivariate analyses, and for the same reasons as data transformation. Data reductions can include eliminating or rolling-up rare taxa or reducing field sub-samples by pooling or averaging. Protocols for data reductions for marine communities are varied, but subsequent interpretations of data analyses should take these reductions into account. For example, elimination of rare taxa may result in the elimination of 90% or more of the biomass within a given station if those rare taxa are large. In some cases, rare taxa are rolled up into higher groups, which prevents loss of information but adds assumptions about the uniform behaviour of mixed taxonomic groups. Reviews of standard methods of data reductions are given in Stephenson and Cook (1980), with some ecological consideration in Burd et al. (1990).

Logarithmic transformations have often been used for benthic invertebrate data because organism abundance typically varies exponentially (Green 1979). A log transformation will reduce the importance of the numerically dominant members and improve the likelihood of resolving structure when differences are due to medium-abundance or rare taxa. However, a log transformation is quite extreme. Other researchers have advocated the use of other geometric conversions such as square root, cube root, fourth-root, natural log, etc. (for reviews, see Hoyle 1973; Tukey 1977; Hoaglin et al. 1983; Downing 1981). Downing (1979) showed empirically that the best overall transformation for stabilizing variance in freshwater benthos was the fourth-root ($x^{0.25}$), because this greatly improves the performance of parametric multivariate methods such as ordinations. Vezina (1988) repeated the exercise for marine subtidal communities, concluding that they were empirically less aggregated than their freshwater counterparts and require a less extreme transformation (e.g., $x^{0.4}$). However, both researchers emphasize that the mean and variance relationships of any given community need to be analyzed to determine the most appropriate transformation. In this way it is possible to check whether or not the transformation used has stabilized the variance.

4.8 Data Reporting Guidelines

Data are submitted in the electronic database format and in hard copies (the interpretative report), as outlined and provided by Environment Canada (see Chapter 10 of the present document for additional information on electronic reporting). The complete fundamental data set, including rare and highly variable taxa and ambiguous identifications, should be stored in this manner, even if data filtering has been applied prior to calculation of community descriptors. Other approaches to data filtering, calculation of community descriptors, and analysis can be employed in reanalysis or meta-analysis. A list of the relevant details for the field, laboratory and data analysis components of the EEM benthic invertebrate survey is provided below; these details should be included and submitted with the interpretative report.

Field reporting

1. field sheets should be retained for six years
2. replicate station location (grid coordinates)
3. date and time of sampling
4. field crew members
5. habitat descriptions, including measures of the supporting environmental variables
6. sampling method used, including type and size of sampler and sieve or mesh size

Laboratory reporting

1. bench sheets should be retained for six years
2. raw data reported for each individual or pooled field sub-sample, listing taxa present and numbers of individuals
3. method and level of sub-sampling applied in the laboratory sorting process
4. sorting efficiency achieved
5. taxonomic authorities used
6. location of reference collection and report on taxonomic verification

Data analysis reporting

1. tabular listing of the number of individuals per taxon in each sample as an appendix
2. tabular summaries of calculated descriptors with variance estimates
3. estimates of power obtained for the survey
4. effects of outliers or extreme values on the results (if any)
5. a summary of adherence to data quality objectives, standard operating procedures and sampling protocols, and identification of any QA/QC problems

4.9 Effect Endpoints and Supporting Endpoints for the Benthic Invertebrate Community

Total invertebrate density: The total number of individuals of all taxonomic categories collected at the station expressed per unit area (e.g., numbers/m²). Values should be reported for each station, as well as the arithmetic mean $\pm$ standard error (SE), $\pm$ standard deviation (SD), median, minimum and maximum for the area.

Taxa (i.e., family) richness: The total number of different taxonomic categories collected at the station, and the arithmetic mean $\pm$ SE, $\pm$ SD, median, minimum and maximum for the area.

Evenness index (Simpson's Evenness Index) (equitability): Evenness (E) can be quantified for each station, and mean E $\pm$ SE, $\pm$ SD, median, minimum and maximum for the area should be reported. Evenness is calculated as in Smith and Wilson (1996):

$$E = 1 / \sum_{i=1}^s (p_i)^2 / S$$

where:

E = evenness

p_i = the proportion of the i^{th} taxon at the station

S = the total number of taxa at the station

Similarity index (Bray-Curtis [B-C] Index): The B-C Index is a distance co-efficient that reaches a maximum value of 1 for two sites that are entirely different and a minimum value of 0 for two sites that possess identical descriptors. Distance coefficients measure the amount of association between sites, and the B-C Index is a member of the class of distance coefficients known as a semimetric that some prefer to call dissimilarity coefficients. The B-C Index measures the percentage of difference between sites (Legendre and Legendre 1983), where the distance statistic is calculated as below:

$$B - C = \frac{\sum_{i=1}^n |y_{i1} - y_{i2}|}{\sum_{i=1}^n (y_{i1} + y_{i2})}$$

where:

B-C = Bray-Curtis distance between sites 1 and 2

Y_{i1} = count for taxon i at site 1

Y_{i2} = count for taxon i at site 2

n = total number of taxa present at the two sites

The Bray-Curtis distance (B-C) from a calculated reference median will be reported for each station, and the arithmetic mean $\pm$ SE, $\pm$ SD, minimum and maximum B-C distance is reported for the area. As the use of this index for determination of effects may be novel to some, a brief literature summary and a detailed example is provided below.

Most of the invertebrate community statistics discussed above are measures of total density and taxa richness and provide no quantitative information on what kind of organisms are present. A similarity index is also recommended, as it summarizes the overall difference in community structure between reference and exposed sites in a single number, requires no preconceived assumptions about the nature of the community and only varies in one direction (Taylor and Bailey 1997). Of the various indices available, many reviewers have indicated that the Bray-Curtis Index (Bray and Curtis 1957) is the most reliable (Pontasch et al. 1989; Jackson 1993; Bloom 1981). The Bray-Curtis Index is also unaffected by the nature of the communities being compared (Bloom 1981), and differences contribute the same to the Bray-Curtis (B-C) Index regardless of whether the taxon is rare or abundant. Bloom (1981) showed that, of 4 indices examined, only the B-C Index accurately reflected the true resemblance over its range.

Example of Bray-Curtis Index for use in the EEM program

The following steps use an example data set to illustrate how the Bray-Curtis Index should be used for the evaluation of effects in the EEM program. In this example, 5 stations were sampled from an exposure area and a reference area, with a total of 5 taxa found to be present.

- 1) Taxa density is entered into a table.
- 2) For the reference stations, the median taxa density is determined (see example below).

	Taxa Density				
Reference Stations	Taxon 1	Taxon 2	Taxon 3	Taxon 4	Taxon 5
Ref 1	2	3	2	3	1
Ref 2	3	5	2	4	3
Ref 3	9	1	1	1	1
Ref 4	4	6	3	4	1
Ref 5	5	4	2	3	2
Reference Median	4	4	2	3	1

- 3) A similar table is constructed for the exposure stations without the median calculation.

	Taxa Density				
Exposure Stations	Taxon 1	Taxon 2	Taxon 3	Taxon 4	Taxon 5
Exp 1	23	4	2	10	1
Exp 2	12	2	2	8	3
Exp 3	14	6	1	6	2
Exp 4	13	1	3	12	2
Exp 5	15	3	2	4	1

- 4) The distance of each station (reference and exposure) from the reference median is calculated as illustrated by the following example for reference station 1.

For this approach, the reference median for particular taxa becomes y_{i2} , the taxon count for site 2 in the above equation.

	Taxa 1	Taxa 2	Taxa 3	Taxa 4	Taxa 5
Ref 1 (y_{i1})	2	3	2	3	1
Reference median (y_{i2})	4	4	2	3	1

 y_{i1}-y_{i2} or Ref 1- reference median	2	1	0	0	0

(y_{i1}+y_{i2})	6	7	4	6	2

Substituting into the B-C equation gives:

$$B-C_{Ref 1} = \frac{2+1+0+0+0}{6+7+4+6+2} = \frac{3}{25} = 0.12$$

- 5) The B-C distance from the reference median is calculated for each station in this manner.
- 6) The result of this calculation should be reported for each station, along with the mean ($\pm$ SE) for the area. The sample data set would result in the following B-C distances:

Station	$\Sigma y_{i1}-y_{i2} $	$\Sigma (y_{i1}+y_{i2})$	B-C distance from median	Mean $\pm$ SE
Ref 1	3	25	0.12	0.18 $\pm$ 0.06
Ref 2	5	31	0.16	
Ref 3	11	27	0.41	
Ref 4	4	32	0.13	
Ref 5	2	30	0.07	
Exp 1	26	54	0.48	0.43 $\pm$ 0.03
Exp 2	17	41	0.41	
Exp 3	17	43	0.40	
Exp 4	23	45	0.51	
Exp 5	13	39	0.33	

- 7) Finally, for the purposes of determining an effect at the exposure area, the mean B-C distance between the reference stations and the reference median (0.18 $\pm$ 0.06) can be compared statistically to the mean distance between the exposure stations

and the reference median (0.43 ± 0.03).

Simpson's Diversity Index: Simpson's Diversity Index (D) takes into account both the abundance patterns and taxonomic richness of the community. This is calculated by determining, for each taxonomic group at a station, the proportion of individuals that it contributes to the total in the station. D for each station and mean ($\pm$ SE, $\pm$ SD), median, minimum and maximum D for the area should be reported. Simpson's Diversity Index is calculated as (Krebs 1985):

$$D = 1 - \sum_{i=1}^s (p_i)^2$$

where:

D = Simpson's index of diversity

S = the total number of taxa at the station

p_i = the proportion of the i^{th} taxon at the station

Taxa (i.e., family) density: The number of individuals of each family expressed per unit area (e.g., numbers/m²). Values should be reported for each taxon at each station and as the mean ($\pm$ SE) of each taxon for the area.

Taxa (i.e., family) proportion: The percentage abundance for each taxon at each station and the mean ($\pm$ SE) percentage abundance of each taxon for the area.

Taxa (i.e., family) presence/absence: A matrix indicating the presence and absence of each taxon at the sampling stations should be reported. The matrix will consist of stations (columns) and taxa (rows).

In addition to the benthic invertebrate endpoints, the sediment monitoring variables are also to be reported (see Chapter 7).

4.10 Evaluation of Results

4.10.1 Effect on the Benthic Invertebrate Community

The objective of the benthic invertebrate component of the EEM program is to answer the following question:

“Is there an effect on the benthic invertebrate community?”

The definition of effect is described in Schedule 5, section 1 of the *Metal Mining Effluent Regulations*.

During the first phases and in magnitude and geographic extent phases of the monitoring program, the following effect endpoints are calculated, reported and used to determine if there is an effect on the benthic invertebrate community:

1. Total benthic invertebrate density
2. Taxa (i.e., family) richness
3. Evenness index (Simpson's)
4. Similarity index (Bray-Curtis)

For the benthic invertebrate component, it is recommended that the following supporting endpoints also be calculated and reported:

Simpson's Diversity Index

Taxa (i.e., family) density

Taxa (i.e., family) proportion

Taxa (i.e., family) presence/absence

All these endpoints, described into details in the previous section are largely summary metrics selected to encompass the range of effects that may be a result of mine effluent.

Many other benthic invertebrate descriptive metrics are available in the literature and serve to address a wide range of questions regarding benthic invertebrate communities. If desired, additional site-specific descriptors may be calculated and used to support the interpretation of effects. For guidance on selecting these optional descriptive metrics and discussion of their applicability, readers are referred to the reviews by Resh et al. (1995).

For the statistical analyses and determination of sufficient power, the recommendations for setting of effect size, α and β are also applicable. The recommendation in this previous section was to set α and β equally at 0.10 or less. The appropriate method of analysis for each of the study design options (e.g., ANOVA, ANCOVA, regression, multivariate analysis) is indicated in Table 4-1.

A final caveat regarding effects on the benthic invertebrate community: it is essential for the mine to select a site-specific study design to allow for an appropriate evaluation. Critical to the study design is the selection of an appropriate reference area or areas. The importance of proper reference area selection is underscored by the following, potentially frequent, example. If a mine performs a simple control-impact design with the reference area placed upstream, then differences between upstream and downstream communities will be those determining the presence or absence of effects. However, if the downstream benthic communities are modified due to a factor such as the restoration of an upstream flow disruption (e.g., from a dam), then these communities, although different from upstream communities, may be more similar to (but perhaps not exactly the same as) the communities at a reference area chosen in a drainage basin adjacent to (or even further afield than) the mine drainage basin. In this example, selection of an additional reference area (see Figure 4-2d for an example) may well be worth the extra cost involved so that site-specific interpretation and the appropriate assessment of effects can be

accomplished. Note that this example of significant upstream-downstream differences may not necessarily be considered an effect if sufficient additional evidence suggests otherwise.

4.10.2 Next Step

Once the monitoring data have been analyzed, decisions regarding the next step in the EEM program are made. The next step in the monitoring program is dependent on the relationship between several key factors, which are briefly discussed below.

The statistical outcome of the previous benthic invertebrate survey

There are 3 possible statistical outcomes of the benthic invertebrate community survey:

- a) no effect is detected but power is not sufficient (i.e., power < 0.90)
- b) no effect is detected and power is sufficient (i.e., power $\geq$ 0.90)
- c) an effect is detected

If any of the effect endpoints (total benthic invertebrate density, taxa richness, evenness index (Simpson's) and similarity index (Bray-Curtis)) demonstrate a statistical difference between exposure and reference areas (or along a gradient), then the conclusion is that there is an effect on the benthic invertebrate community. This result can be obtained by various statistical methods; the choice of methods depends on the study design of the monitoring program.

If the power was insufficient, the mine may reconsider the number of sampling stations or the sampling design that was used, in order to design a study with sufficient power in the next survey.

EEM program options after an effect has been established

If an effect on the benthic invertebrate community is found, the next question to be addressed is:

Is the effect mine-related?

An assessment of whether the effect is mine-related could include asking the following questions:

- Is the cause of the effect known or suspected?
- Can the effect be related to a natural change in the aquatic receiving environment?
- Can the effect be reasonably correlated to an anthropogenic cause other than the mine effluent?
- Is there a weight-of-evidence approach that can indicate a causal link? (See section 4.11)

This series of questions is provided as an example of the type of approach that may allow for the determination of whether or not the observed effect is mine-related. If the presence of confounding factors makes it difficult to determine the effect of mine effluent on the benthic invertebrate community, the mine should reconsider the study design for the next phase. If the effect has been confirmed, and the cause of the effect is unknown, the mine proceeds to the next step of data assessment and interpretation: determining the magnitude and geographic extent of the effect.

Are the magnitude and geographic extent known?

If an effect has been confirmed (see Chapter 1 for details on confirmed effects), and the cause of the effect is unknown, then the mine should proceed to the next step and determine the magnitude and geographic extent of the effect. For additional information refer to section 4.2.2.

4.11 Additional Tools for Focused Monitoring, Weight-of-Evidence Approaches and/or Investigation of Cause

There are a number of alternative approaches and tools possible for investigations of cause in the EEM program. Methods provided in this guidance document are not meant to be exhaustive, and mines may propose additional scientifically defensible approaches. Tools should be cost-effective, recognized in the primary literature, readily available from consulting, academic or government laboratories, and applicable to the EEM program.

Additional information can be found in Chapters 9 and 12 of this document.

4.11.1 Use of Weight-of-Evidence Approaches to Establish Cause of Effects

Distinguishing among the cumulative impacts of multiple stressors (which sometimes have confounding effects) requires the establishment of a definitive causal link to the mine effluent under evaluation. The environmental assessment of an aquatic ecosystem is particularly prone to impediments because such ecosystems often receive multiple, interactive effluent discharges. Assessments of monitoring results often rely, in large part, on field monitoring data that can only show correlations rather than clear cause and effect between mine effluent and a presumed effect. Establishing a strong causal link, however, can benefit from a weight-of-evidence approach that combines information from a variety of sources. For additional information on the use of weight-of-evidence approaches, readers are referred to chapters 9 and 12 of the present document.

4.11.2 Lethal and Sublethal Toxicity Tests

Lethal and sublethal toxicity test methods can be applied during magnitude and geographic extent and investigation-of-cause surveys when an effect has been identified

or when previous work failed to provide a satisfactory explanation of cause. These methods provide a direct determination of lethal or sublethal toxicity and can verify that alterations in benthos are due to the toxicity of the mine effluent rather than confounding factors. For example, adverse effects on benthic community structure may be due to factors other than effluent toxicity, including differences in environmental regime. Concurrent impairment of benthic community structure and toxicity implicates the effluent itself as the cause of changes in the benthos. These methods also provide important information for interpreting field effects in situations where benthic community data are inconclusive, or if only pollution-tolerant species are present in both impacted and reference sites.

4.11.3 Analysis of Sediment Cores for Historic Trends

Sedimentary records from depositional areas of water bodies can be used to indicate limnological conditions in recent and ancient history (Frey 1988). Precise dating of sediments, combined with an inventory of the remains of certain organisms and plant material (e.g., diatoms, zooplankton, insects), provides a chronology of changes that often can be linked to the period of anthropogenic influence. In addition to the water body itself, the history of the watershed and airshed may be deduced, and the influences of natural events may be distinguished from anthropogenic impacts. A substantial volume of literature is available on the subject, with a useful synthesis of the science provided by Frey (1988). Due to the level of expertise needed to undertake this type of analysis, the availability of paleolimnological services is limited. In addition, the analyses are restricted to resolving trends over longer time frames (multiple years to decades) as a result of sedimentation processes such as bioturbation. The costs of the technique will be site-specific.

4.11.4 Other Benthic Invertebrate Measures and Organisms

Benthic invertebrates are recommended as the primary indicator organisms for use in an EEM program for monitoring effects on fish habitat. However, the level of identification and measures recommended in the main text of the guidance document are not an exclusive list of measures for which benthic invertebrates can be evaluated. Additional measures include biomass, lower level of identification, secondary production, and population fitness parameters.

Benthic invertebrate biomass in **marine environments** can provide additionally useful information because it is related to the availability of energy to other trophic levels (e.g., fish). For marine communities, some investigators suggest that an analysis of benthic abundance and biomass together provide a sensitive indicator of changes in the composition of the benthic community (e.g., Warwick 1986; Warwick et al. 1987; Clarke 1990; Burd et al. 1990). For example, in marine samples, it is in the measurement of distributions of biomass that the three main functional groups of benthic organisms—microfauna (grain surface dwellers), meiofauna (interstitial organisms) and macrofauna (burrowers and epifauna)—can be distinctly separated (Schwinghamer 1981, 1983).

Because these three groups of organisms have different reproductive modes, metabolic rates, life histories and habitat adaptations, they respond differently to habitat perturbation. This could be particularly important in Arctic subtidal habitats, where abundance may be low, but individuals may be large. However, because precise biomass measurements are time-consuming and problematic (cf. Crisp 1984) unless collected in more detail and more often than is feasible for EEM requirements, it is only possible to determine relative changes in biomass of samples for the EEM surveys. This is easily done by taking blotted wet-weight measurements of representative-sized adult specimens of each species for each survey. Since the method is non-destructive, the reference collection may be used for this purpose prior to external verification or archiving. The mean weight of a given species can then be used to transform species abundance data to relative species biomass data for further summary or statistical analyses. These data show relative, large-scale changes only, and cannot be used to infer production or trophic flow rates within benthic communities.

In addition to benthic invertebrates, several other types of aquatic biota were considered for use in the EEM program. The most relevant ones were 1) phytoplankton, 2) macrophytes, and 3) periphyton.

4.12 References

Abbott RT. 1974. American seashells. 2nd ed. New York (NY): Van Nostrand Reinhold Co.

Abbott RT, Zim HS, Sandström GF. 2001. Seashells of North America: a guide to field identification. Golden Field Guides from St. Martin's Press. New York (NY): St. Martin's Press.

[AETE] Aquatic Effects Technology Evaluation Program. 1995. Field evaluation of aquatic effects monitoring methods - pilot study. Volume 1. AETE Project 4.1.1. Ottawa (ON): Canada Centre for Mineral and Energy Technology.

Alberta Environment. 1990. Selected methods for the monitoring of benthic invertebrates in Alberta rivers. Environmental Quality Monitoring Branch, Alberta Environment.

Appy TD, Linkletter IF, Dadswell MJ. 1980. A guide to the marine flora and fauna of the Bay of Fundy: Annelida: Polychaeta. St. Andrews (NB): Fisheries and Marine Service. Technical Report No. 920.

Austin WC. 1985. An annotated checklist of marine invertebrates in the cold temperate northeast Pacific. Vol. 1, 2 and 3. Cowichan Bay (BC): Khoyatan Marine Laboratory. Report for Fisheries and Oceans Canada.

Bailey RC, Norris RH, Reynoldson TB. 2003. Bioassessment of freshwater ecosystems. Boston (MA): Kluwer.

Baker HR. 1980. Key to the common Tubificid species of the northeast Pacific. Manuscript from Oligochaeta workshop.

Banse K. 1972. Redescription of some species of *Chone* Kröyer and *Euchone* Malmgren, and three new species (Polychaeta: Sabellidae). Fish Bull 70(2):459-495.

Banse K, Hobson KD. 1974. Benthic errantiate polychaetes of British Columbia and Washington. (Bulletin of Fisheries Research Board of Canada 185). Ottawa (ON): Fisheries and Marine Service.

Barber WE, Kevern NR. 1974. Seasonal variation of sieving efficiency in a lotic habitat. Freshwat Biol 4:293-300.

Beak International Inc. 1999. Quality assurance program for assessing mine-related effects using benthic invertebrate communities. Ottawa (ON): Natural Resources Canada, Canada Centre for Mineral and Energy Technology. Aquatic Effects Technology Evaluation Program project 2.1.4.

Bednarik AF, McCafferty WP. 1979. Biosystematic revision of the genus *Stenonema* (Ephemeroptera, Heptageniidae). Can Fish Aquat Sci Bull 201. Ottawa: Department of Fisheries and Oceans.

Berkeley C, Berkeley E. 1952a. Canadian Pacific fauna, 9. Annelida 9b (1) Polychaeta Errantia. (Bulletin of Fisheries Research Board of Canada). Toronto (ON): University of Toronto Press.

Berkeley C, Berkeley E. 1952b. Canadian Pacific fauna 9. Annelida 9b (2) Polychaeta Sedentaria. (Bulletin of Fisheries Research Board of Canada). Toronto (ON): University of Toronto Press.

Blake JA. 1971. Revision of the genus *Polydora* from the east coast of North America (Polychaeta: Spionidae). *Smithson Contr Zool* 75:1-32.

Blake JA. 1988. New species and records of Phyllodocidae (Polychaeta) from Georges Bank and other areas of the western North Atlantic. *Sarsia* 73:245-257.

Blake JA. 1991. Revision of some genera and species of Cirratulidae (Polychaeta) from the western North Atlantic. In Petersen ME, Kirkegaard JB, editors. Systematics, biology and morphology of world Polychaeta. Proceedings of the Second International Polychaete Conference, Copenhagen, August 18-23, 1986. *Ophelia Supplement* 5: 1-723. p. 17-30.

Bloom SA. 1981. Similarity indices in community studies: potential pitfalls. *Mar Ecol Prog Ser* 5:125-128.

Booth J, Hay D, Truscott, J. 1996. Standard methods for sampling resources and habitats in coastal subtidal regions of British Columbia: Part I - review of mapping with preliminary recommendations. (Canadian Technical Report of Fish and Aquatic Sciences 2118). Nanaimo (BC): Fisheries and Oceans Canada.

Bousfield EL. 1958. Freshwater amphipod crustaceans of glaciated North America. *Can Field Nat* 72(2):April-June.

Bousfield EL. 1960. Canadian Atlantic sea shells. Ottawa (ON): Department of Northern Affairs and Natural Resources, National Museum of Canada.

Bousfield EL. 1973. Shallow-water gammaridean Amphipoda of New England. Ithaca (NY): Cornell University Press.

Bousfield EL, Hendryks EA, 1994. The amphipod superfamily Leucothoidea on the Pacific coast of North America. Family Pleustidae: subfamily Pleustinae, systematics and biogeography. *Amphipacifica* I:2.

Bousfield EL, Hendryks EA. 1995a. The amphipod family Pleustidae on the Pacific coast of North America. Part III. Subfamilies Parapleustinae, Dactylopleustinae, and Pleusirinae: systematics and distributional ecology. *Amphipacifica* II:1.

Bousfield EL, Hendryks EA. 1995b. The amphipod superfamily Eusiroidea in the North Pacific region. I. Eusiridae: systematics and distributional ecology. *Amphipacifica* I:4.

Bousfield EL, Hoover PM. 1995. The amphipod superfamily Pontoporeioidea on the Pacific coast of North America. II. Family Haustoriidae. Genus *Eohaustorias* J.L. Barnard: systematics and distributional ecology. *Amphipacifica* II:1.

Bousfield EL, Kendall JA. 1994. The amphipod superfamily Dexaminoidea of the North American Pacific coast, families Atylidae and Dexaminidae: systematics and distributional ecology. *Amphipacifica* I:3.

Bowman MF, Bailey RC. 1997. Does taxonomic resolution affect the multivariate description of the structure of freshwater benthic macroinvertebrate communities? *Can J Fish Aquat Sci* 54:1802-1807.

Brandlova J, Brandl Z, Fernando, CH. 1972. The Cladocera of Ontario with remarks on some species and distribution. *Can J Zool* 50:1373-1403.

Bray JR, Curtis JT. 1957. An ordination of the upland forest communities of southern Wisconsin. *Ecol Monogr* 27:325-349.

Brinkhurst RO. 1982. British and other marine and estuarine oligochaetes. (Synopses of the British fauna 21). Cambridge (UK): Cambridge University Press.

Brinkhurst RO. 1986. Guide to the freshwater aquatic microdrile oligochaetes of North America. (Canadian Special Publication of Fisheries and Aquatic Sciences 84). Ottawa (ON): Fisheries and Ocean Canada, Scientific Information and Publications Branch.

Brinkhurst RO, Baker HR. 1979. A review of the marine Tubificidae (Oligochaeta) of North America. *Can J Zool* 57:1553-1569.

Brunel P, Bosse L, Lamarche G. 1998. Catalogue des Invertébrés marins de l'estuaire et du golfe du Saint-Laurent / Catalogue of the marine invertebrates of the estuary and Gulf of Saint Lawrence. Ottawa (ON): National Research Council of Canada Press.

Burch JB. 1975a. Freshwater Sphaeriacean clams (Mollusca: Pelecypoda) of North America. Hamburg (MI): Malacological Publications.

Burch JB. 1975b. Freshwater Unionacean clams (Mollusca: Pelecypoda) of North America. Hamburg (MI): Malacological Publications.

Burch JB. 1989. North American freshwater snails. Hamburg (MI): Malacological Publications.

Burd BJ, Nemec A, Brinkhurst RO. 1990. The development and application of analytical methods in benthic marine infaunal studies. *Adv Mar Biol* 26:169-247.

Butler TH. 1983. Shrimps of the Pacific coast of Canada. (Bulletin 202). Ottawa (ON): Department of Fisheries and Oceans.

Camburn KE, Kingston JC, Charles DF, editors. 1984-1986. PIRLA diatom iconograph.

Contains 53 photographic plates and 1059 figures, plus figure legends. (PIRLA Unpublished Report Series No. 3). Bloomington (IN): Indiana University.

Chu HF, Cutkomp LK. 1992. How to know the immature insects. 2nd edition. Dubuque (IA): Wm. C. Brown Publishers.

Clark HL. 1915. Catalogue of recent ophiurans. *Mem Mus Comp Zool Harv* 25:165-376.

Clark HL. 1924. Some holothurians from British Columbia. *Can Field Nat* 38:54-57.

Clarke AH. 1981. The freshwater molluscs of Canada. Ottawa (ON): National Museum of Natural Sciences.

Clarke KR. 1990. Comparisons of dominance curves. *J Exp Mar Biol Ecol* 138:143-157.

Clarke RT, Furse MT, Wright JF. 1992. A comparison of single, paired and three season combined macro-invertebrate samples for the biological banding of river quality. Bristol (UK): National Rivers Authority.

Clarke KR, Green RH. 1988. Statistical design and analysis for a “biological effects” study. *Mar Ecol Prog Ser* 46:213-226.

Clifford HF. 1991. Aquatic invertebrates of Alberta. Edmonton (AB): University of Alberta Press.

Coates KA. 1980. Keys to intertidal genera and species of Enchytraeidae found in British Columbia. Manuscript from Oligochaeta Workshop. Victoria (BC) University of Victoria.

Coe WR. 1912. Echinoderms of Connecticut. Connecticut State Geological and Natural History Survey 19:1-147.

Coe WR. 1943. Biology of the nemerteans of the west coast of North America. *Trans Conn Acad Arts Sci* 35:129-328, pls. 1-4.

Cowardin LM, Carter V, Golet FC, LaRoe ET. 1979. Classification of wetlands and deepwater habitats of the United States. Washington (DC): U.S. Fish and Wildlife Service. FWS/OBS-79/31.

Cowardin LM, Golet FC. 1995. U.S. Fish and Wildlife Service 1979 wetland classification: a review. *Vegetatio* 118:139-152.

Crisp DJ. 1984. Energy flow measurements. In Holme NA, McIntyre AD, editors. *Methods for the study of marine benthos*. IBP Handbook 16. Oxford (UK): Blackwell Scientific Publications. p. 284-370.

Crocker DW, Barr DW. 1968. Handbook of the crayfishes of Ontario. Toronto (ON): University of Toronto Press.

Cuffney TF, Gurtz ME, Meador MR. 1993. Guidelines for the processing and quality assurance of benthic invertebrate samples collected as part of the national water-quality assessment program. U.S. Geological Survey Open-File Report 93-407.

Culp JM, Lowell RB, Cash KJ. 2000. Integrating mesocosm experiments with field and laboratory studies to generate weight-of-evidence risk assessments for large rivers. *Environ Toxicol Chem* 19:1167-1173.

Cutler EB. 1973. Sipuncula of the western North Atlantic. *Bull Am Mus Nat Hist* 152(3):1-204.

Downing JA. 1979. Aggregation, transformation and the design of benthos sampling programs. *J Fish Res Bd Can* 36:1454-1463.

Downing JA. 1981. How well does the fourth-root transformation work? *Can J Fish Aquat Sci* 38:127-129.

Downing JA. 1984. Sampling the benthos of standing waters. In Downing JA, Rigler FH, editors. *A manual on methods for the assessment of secondary productivity in fresh waters*. 2nd. edition. (IBP Handbook 17). Oxford (UK): Blackwell Scientific Publications. p. 87-130.

Downing JA. 1986. Spatial heterogeneity: evolved behaviour or mathematical artefact? *Nature* 323: 255-257.

Dussart B. 1969. *Les copépodes des eaux continentales*. Paris (FR): N. Boubée & Cie.

Edmunds GF Jr, Jensen SL, Berner L. 1976. *The mayflies of North and Central America*. Minneapolis (MN): University of Minnesota Press.

Eleftheriou A, Holme NA. 1984. Macrofauna techniques. In Holme NA, McIntyre AD, editors. *Methods for the study of marine benthos*. (IBP Handbook 16). Oxford (UK): Blackwell Scientific Publications. p. 140-216.

Elliott JM. 1977. Some methods for the statistical analysis of samples of benthic invertebrates. (Freshwater Biological Association Scientific Publication No. 25). Ambleside (UK): Freshwater Biological Association.

Elliott JM, Tullett PA. 1978. *A bibliography of samplers for benthic invertebrates*. (Freshwater Biological Association, Occasional Publication No. 4). Ambleside (UK): Freshwater Biological Association.

Elliott JM, Tullett PA. 1983. *A supplement to a bibliography of samplers for benthic macroinvertebrates*. (Freshwater Biological Association Occasional Publication No. 20.) Ambleside (UK) Freshwater Biological Association.

Environment Canada. 2002. Revised guidance for sample sorting and sub-sampling protocols for EEM benthic invertebrate community surveys. Ottawa (ON): Environment Canada, National EEM Office.

Fauchald K. 1977. The Polychaeta worms: definitions and keys to orders, family and genera. (Science Series 28). Los Angeles (CA): Natural History Museum of Los Angeles County.

Finlayson CM, van der Valk AG. 1995. Wetland classification and inventory: a summary. *Vegetatio* 118(1-2):185-192.

Fitzpatrick JF. Jr. 1983. How to know the freshwater crustacea. Dubuque (IA): Wm.C. Brown Publishers.

Fournier JA, Petersen ME. 1991. *Cossura longocirrata*: redescription and distribution, with notes on reproductive biology and a comparison of described species of *Cossura* (Polychaeta: Cossuridae). *Ophelia Suppl* 5:63-80.

Frey DG. 1988. What is paleolimnology? *J Paleolim* 1:5-8.

Frith HR, Seraring G, Wainwright P, Harper H, Emmett B. 1993. Review of habitat classification systems and an assessment of their suitability to coastal B.C. Unpublished report. Sidney (BC): L.G.L. Ltd for Environment Canada, Environmental Emergency Services Branch.

Fullington KE, Stewart KW. 1980. Nymphs of the stonefly genus *Taeniopteryx* (Plecoptera: Taeniopterygidae) of North America. *J Kansas Entomol Soc* 53:237-259.

EVS Environment Consultants. 1993. Guidelines for Monitoring Benthos in Freshwater Environments. Prepared for: Environment Canada, 224 West Esplanade, North Vancouver, B.C.

Gibson R. 1994. Nemerteans. 2nd edition.. (Synopsis of the British Fauna No. 24). Shrewsbury (UK): Field Studies Council.

Glozier N. 1989. The effects of biotic and abiotic factors on the foraging success of a lotic minnow, *Rhinichthys cataractae* [master's thesis]. Calgary (AB): University of Calgary.

Gosner KI. 1971. Guide to the identification of marine and estuarine invertebrates. New York (NY): J. Wiley and Sons.

Graham A. 1988. Molluscs: prosobranchs and pyramellid gastropods (2nd ed.). Kermack DM, Barnes RSK, editors. Synopsis of the British Fauna (New Series) No. 2. London (UK): The Linnean Society of London.

Gray JS, Clarke KR, Warwick RM, Hobbs G. 1990. Detection of initial effects of marine pollution on marine benthos: an example from the Ekofisk and Eldfisk oilfields, North

Sea. Mar Ecol Prog Ser 66:285-299.

Green RH. 1979. Sampling design and statistical methods for environmental biologists. Toronto (ON): John Wiley and Sons.

Green G. 1994. Protocols for reference and voucher collections of aquatic invertebrates stored at the Royal British Columbia Museum. DOE-FRAP 1994-15.

Harper PP, Stewart KW. 1984. 13: Plecoptera. In Merritt RW, Cummins, KW, editors. An introduction to the aquatic insects of North America. 2nd ed. Dubuque (IA): Kendall/Hunt Publ.Co.

Hart JFL. 1982. Crabs and their relatives of British Columbia. Handbook No. 40. Victoria (BC): British Columbia Provincial Museum.

Hauer FR, Lamberti GA, editors. 1996. Methods in stream ecology. San Diego (CA): Academic Press Inc.

Hay DE, Waters RD, and Boxwell TA, editors. 1996. Proceedings, Marine Ecosystem Monitoring Network Workshop, Nanaimo, B.C. March 28-30. 1995. (Canadian Technical Report of Fisheries and Aquatic Sciences 2108). Nanaimo (BC): Fisheries and Oceans Canada.

Hilsenhoff WL. 1995. Aquatic insects of Wisconsin: keys to Wisconsin genera and notes on biology, habitat, distribution and species. Publication Number 3. Madison (WI): Natural History Museums Council, University of Wisconsin.

Hilsenhoff WL, Schmude KL. 1992. Riffle beetles of Wisconsin (Coleoptera: Dryopidae, Elmidae, Lutrochidae, Psephenidae) with notes on distribution, habitat, and identification. Great Lakes Entomol 25(3):191-213.

Hitchcock SW. 1974. Guide to the insects of Connecticut. Part VII. The Plecoptera or Stoneflies of Connecticut. Bulletin of the State Geological and Natural History Survey of Connecticut 107.

Hoaglin DC, Mosteller F, Tukey JW. 1983. Understanding robust and exploratory data analysis. New York (NY): John Wiley.

Hobson KD, Banse K. 1981. Sedentariate and Archiannelid Polychaetes of British Columbia and Washington. Bulletin 209. Ottawa (ON): Department of Fisheries and Oceans.

Holme NA, McIntyre AD, editors. 1984. Methods for the study of marine benthos. (IBP Handbook 16). Oxford (UK): Blackwell Scientific Publications.

Hoyle MH. 1973. Transformations: an introduction and a bibliography. Inter Stat Rev 41:203-223.

Hurlbert SH. 1984. Pseudoreplication and the design of ecological field experiments. *Ecol Monog* 54:187-211.

Hyman LH. 1940. The polyclad flatworms of the Atlantic coast of the United States and Canada. *Proc US Nat Mus* 89:449-495.

Hyman LH. 1944. Marine Turbellaria from the Atlantic coast of North America. *Am Mus Novitates* No. 1266.

Jackson DA. 1993. Multivariate analysis of benthic invertebrate communities: the implication of choosing particular data standardizations, measures of association, and ordination measures. *Hydrobiol* 268:9-26.

Johannsen OA. 1977. Aquatic Diptera. Los Angeles (CA): Entomological Reprint Specialists. Fourth reprinting.

Johnson RK, Wiederholm T, Rosenberg DM. 1993. Freshwater biomonitoring using individual organisms, populations, and species assemblages of benthic macroinvertebrates. In Rosenberg DM, Resh VH, editors. *Freshwater biomonitoring and benthic macroinvertebrates*. New York (NY): Chapman and Hall. p. 40-125.

Jonasson PM. 1955. The efficiency of sieving techniques for sampling freshwater bottom fauna. *Oikos* 6:183-207.

Keen AM, Coan E. 1974. Marine molluscan genera of western North America. Palo Alto (CA): Stanford University Press.

Klemm DJ. 1972. Freshwater leeches (Annelida: Hirudinea) of North America. U.S. Environmental Protection Agency Identification Manual No 8. Washington (DC): U.S. Environmental Protection Agency.

Klemm DJ. 1985. A guide to the freshwater Annelida (Polychaeta, Naidid and Tubificid Oligochaeta, and Hirudinea) of North America. Dubuque (IA): Kendall/Hunt Publ. Co.

Klemm DJ. 1991. Taxonomy and pollution ecology of the Great Lakes region leeches (Annelida: Hirudinea). *Michigan Academician* 24:37-103.

Klemm DJ, Lewis PA, Fulk F, Lazorchak JM. 1990. Macroinvertebrate field and laboratory methods for evaluating the biological integrity of surface waters. Cincinnati (OH): U.S. Environmental Protection Agency, Environmental Monitoring Laboratory. EPA 600/4-90/030.

Knight-Jones P. 1978. New Spirorbidae (Polychaeta: Sedentaria) from the East Pacific, Atlantic, Indian and southern oceans. *Zool J Linnean Soc.* 64:201-240.

Knight-Jones P. 1983. Contributions to the taxonomy of Sabellidae (Polychaeta). *Zool J Linnean Soc* 79:245-295.

Kozloff EN. 1987. Marine invertebrates of the Pacific Northwest. Seattle (WA): University of Washington Press.

Krebs CJ. 1985. Ecology: the experimental analysis of distribution and abundance. 3rd edition. New York (NY): Harper and Row.

Kreis RG Jr. 1986. Variability study. Section 17. In Charles DF, Whitehead DR, editors. Paleocological investigation of recent lake acidification (PIRLA): methods and project description. Palo Alto (CA): Electric Power Research Institute.

Kreis RG Jr. 1989. Variability study: interim results. Section 4. In Charles DF, Whitehead DR, editors. Paleocological investigation of recent lake acidification (PIRLA): 1983-1985: interim report. October 1989. Palo Alto (CA): Electric Power Research Institute.

Kronberg I. 1987. Accuracy of species and abundance minimal areas determined by similarity area curves. Mar Biol 96:555-561.

Lambert P. 1981. The sea stars of British Columbia. (B.C. Provincial Museum Handbook No. 39). Victoria (BC): British Columbia Provincial Museum.

Laplanche S, Bousquet Y, Bélanger P, Chantal C. 1991. Liste des espèces de coléoptères du Québec. (Fabriques, Supplément 6). Sillery (QC): Association des entomologistes amateurs du Québec.

Laubitz DR. 1972. The Caprellidae (Crustacea, Amphipoda) of Atlantic and Arctic Canada. (Publications in Biological Oceanography No. 4). Ottawa (ON): National Museums of Canada, National Museum of Natural Sciences.

Legendre L, Legendre P. 1983. Numerical ecology. Amsterdam (NL): Elsevier.

Lehmkuhl DM. 1975a. Field guide to aquatic insect families. Blue Jay 33:199-219.

Lehmkuhl DM. 1975b. Saskatchewan damselflies and dragonflies. Blue Jay 33:18-27.

Lehmkuhl DM. 1976. Mayflies. Blue Jay 34:70-81.

Lehmkuhl DM. 1979. How to know the aquatic insects. Dubuque (IA): Wm. C. Brown Company Publishers.

Leopold LB, Wolman GM, Miller JP. 1964. Fluvial processes in geomorphology. San Francisco (CA): W.H. Freeman and Co.

Leopold LB. 1994. A view of the river. Cambridge (MA): Harvard University Press.

Lewis PA. 1974. Taxonomy and ecology of *Stenonema* mayflies (Heptageniidae: Ephemeroptera). Cincinnati (OH): U.S. Environmental Protection Agency. EPA-670/4-74-006.

Levings CD, Thom RM. 1994. Habitat changes in the Georgia Basin: implications for resource management and restoration. In Wilson RCH, Beamish RJ, Aitkens F, Bell J, editors. Review of the marine environment and biota of Strait of Georgia, Puget Sound and Juan de Fuca Strait: Proceedings of the BC/Washington Symposium on the Marine Environment, Jan. 13-14, 1994. Canadian Technical Report of Fisheries and Aquatic Sciences 1948). p. 330-351.

Light WJ. 1977. Spionidae (Annelida: Polychaeta) from San Francisco Bay, California: a revised list with nomenclatural changes, new records, and comments on related species from the northeastern Pacific Ocean. *Proc Biol Soc Wash* 90(1):66-88.

Lowell RB. 1997. Discussion paper on critical effect size guidelines for EEM using benthic invertebrate communities. Report to the Environmental Effects Monitoring Program. EEM/1997/9.

Lowell RB, Culp JM, Dubé MG. 2000. A weight-of-evidence approach for northern river risk assessment: integrating the effects of multiple stressors. *Environ Toxicol Chem* 19:1182-1190.

Mackie GL, White DS, Zdeba TW. 1980. A guide to freshwater mollusks of the Laurentian Great Lakes with special emphasis on the genus *Pisidium*. Duluth (MN): U. S. Environmental Protection Agency, Office of Research and Development, Environmental Research Laboratory. EPA-600/3-80-068.

Mackie GL, Huggins DG. 1983. Sphaeriacean clams of Kansas. (Technical Publications of the State Biological Survey of Kansas). Lawrence (KS): University of Kansas.

Malley DF, Reynolds JB. 1979. Sampling strategies and life history of non-insectan freshwater invertebrates. *J Fish Res Bd Can* 36:311-318.

Mason CF. 1991. Biology of freshwater pollution. 2nd edition. New York (NY): Longman Scientific & Technical.

Matthews GWT. 1993. The Ramsar Convention: its history and development. Gland (CH): Ramsar Convention Bureau.

McCafferty WP, Waltz RD. 1990. Revisionary synopsis of the Baetidae (Ephemeroptera) of North and Middle America. *Trans Amer Entomol Soc* 116:769-800.

Meador MR, Hupp CR, Cuffney TF, Gurtz ME. 1993. Methods for characterizing stream habitat as part of the national water-quality assessment program. Raleigh (NC): U.S. Geological Survey Open-File Report 93-408.

Merritt RW, Cummins KW. 1984. An introduction to the aquatic insects of North America. 2nd edition. Dubuque (IA): Kendall/Hunt.

Merritt RW, Cummins KW. 1996. An introduction to the aquatic insects of North America. 3rd edition. Dubuque (IA): Kendall/Hunt.

Moriyara DK, McCafferty WP. 1979. The *Baetis* larvae of North America (Ephemeroptera: Baetidae). Trans Amer Entomol. Soc 105:139-221.

Morris PA. 1951. A field guide to the shells of our Atlantic and Gulf coasts. (The Peterson Field Guide Series). Boston (MA): Houghton Mifflin Company.

Newbury RW. 1984. Hydrolic determinants of aquatic insect habitats. In Resh VH, Rosenberg DM, editors. The ecology of aquatic invertebrates. New York (NY): Praeger Publishers.

Newbury RW, Gaboury M. 1993. Stream analysis: fish habitat and design: a field manual. Gibson (BC): Newbury Hydraulics.

Norris RH, Georges A. 1993. Analysis and interpretation of benthic macroinvertebrate surveys. In Rosenberg DM, Resh VH, editors. Freshwater biomonitoring and benthic invertebrates. New York (NY): Chapman & Hall. p. 234-286.

Oliver DR, McClymont D, Roussel ME. 1978. A key to some larvae of the Chironomidae (Diptera) from the Mackenzie and Porcupine river watersheds. (Fisheries and Marine Service Technical Report 791). Ottawa (ON): Agriculture Canada.

Pearson TH. 1975. The benthic ecology of Loch Linnhe and Loch Eil, a sea-loch system on the west coast of Scotland. IV: Changes in the benthic fauna attributable to organic enrichment. J Exp Mar Biol Ecol 20:1-41.

Peckarsky BL, Fraissinet PR, Penton MA, Conklin D, editors. 1990. Freshwater macroinvertebrates of northeastern North America. Ithaca (NY): Cornell University Press.

Pennak RW. 1978. Freshwater invertebrates of the United States: Protozoa to Mollusca. 3rd ed. New York (NY): John Wiley & Sons Inc.

Pettibone MH. 1963. Marine polychaete worms of the New England Region. 1. Aphroditidae through Trochochaetidae. (U.S. National Museum Bulletin 227:1). Washington (DC): Smithsonian Institute.

Pettibone MH. 1992. Contribution to the polychaete family Pholoidae Kinberg. Smithsonian Contributions to Zoology 532:1-22.

Pettibone MH. 1993. Scaled polychaetes (Polynoidae) associated with ophiuroids and other invertebrates and review of species referred to *Malmgrenia* McIntosh and replaced by *Malmgreniella* Hartman, with descriptions of new taxa. Smithsonian Contributions to Zoology 538:1-92.

Plafkin JL, Barbour MT, Porter KD. 1989. Rapid bioassessment protocols for use in streams and rivers: Benthic macroinvertebrates and fish. Washington (DC): U.S. Environmental Protection Agency, Office of Water Assessment and Watershed Protection Division. EPA/444/4-89-001.

Pohle GW. 1990. A guide to decapod Crustacea from the Canadian Atlantic: Anomura and Brachyura. (Canadian Technical Report of Fisheries and Aquatic Sciences 1771). St. Andrews (NB): Fisheries and Oceans Canada.

Pontasch KW, Smith EP, Cairns J Jr. 1989. Diversity indices, community comparison indices and canonical discriminant analysis: Interpreting the results of multispecies toxicity tests. *Wat Res* 23:1229-1238.

Rabeni CF, Gibbs KE. 1978. Comparison of two methods used by divers for sampling benthic invertebrates in deep rivers. *J Fish Res Bd Can* 35:332-336.

Rees HL. 1984. A note on mesh selection and sampling efficiency in benthic studies. *Mar Poll Bull* 15:225-229.

Reish DJ. 1959. A discussion of the importance of screen size in washing quantitative marine bottom samples. *Ecology* 40:307-309.

Resh VH, Norris RH, Barbour MT. 1995. Design and implementation of rapid assessment approaches for water resource monitoring using benthic macroinvertebrates. *Austral J Ecol* 20:108-121.

Reynoldson TB, Bailey RC, Day KE, Norris RH. 1995. Biological guidelines for freshwater sediment based on benthic assessment of sediment (the BEAST) using a multivariate approach for predicting biological state. *Austral J Ecol* 20:198-219.

Reynoldson TB, Rosenberg DM. 1996. Sampling strategies and practical considerations in building reference data bases for the prediction of invertebrate community structure. In Bailey RC, Norris RH, Reynoldson TB, editors. *Study design and data analysis in benthic macroinvertebrate assessments of freshwater ecosystems using a reference site approach*. Technical Information Workshop, North American Benthological Society, 44th Annual Meeting, Kalispell, Montana. pp. 1-31.

Reynoldson TB, Norris RH, Resh VH, Rosenberg DM. 1997. The reference condition: a comparison of multimetric and multivariate approaches to assess water-quality impairment using benthic macroinvertebrates. *J North Am Benthol Soc* 16:833-852.

Robinson CLK, Levings CD. 1995. An overview of habitat classification systems, ecological models and geographic information systems applied to shallow foreshore marine habitats. (Canadian Manuscript Report of Fisheries and Aquatic Sciences 2322). West Vancouver (BC): Fisheries and Oceans Canada.

Robinson CLK, Hay DE, Booth J, Truscott J. 1996. Standard methods for sampling resources and habitats in coastal subtidal regions of British Columbia: Part 2 - Review of sampling with preliminary recommendations. (Canadian Technical Report of Fisheries and Aquatic Sciences 2119). Nanaimo (BC): Fisheries and Oceans Canada.

Rosenberg DM. 1978. Practical sampling of freshwater macrozoobenthos: A

bibliography of useful texts, reviews, and recent papers. (Canadian Fisheries and Marine Service Technical Report 790).

Rosenberg DM, Resh VH, editors. 1993. Freshwater biomonitoring and benthic macroinvertebrates. New York (NK): Chapman & Hall.

Saether Ole A. 1975. Nearctic and palaearctic *Heterotrissocladius* (Diptera: Chironomidae). (Bulletin of the Fisheries Research Board of Canada 193). Ottawa (ON): Fisheries and Marine Service.

Saether Ole A. 1977. Taxonomic studies on Chironomidae: *Nanocladius*, *Pseudochironomus* and the *Harnischia* complex. Bulletin of the Fisheries Research Board of Canada 196). Ottawa (ON): Fisheries and Marine Service.

Sars GO. 1895. An account of the Crustacea of Norway with short descriptions and figures of all the species. Vol. 1 Amphipoda. Christiana and Copenhagen (DK): A.L.B. Cammermeyers Forlag.

Sars GO. 1899. An account of the Crustacea of Norway with short descriptions and figures of all the species. Volume II. Isopoda. Bergen (NO): Bergen Museum.

Sars GO. 1900. An account of the Crustacea of Norway with short descriptions and figures of all the species. Vol. III. Cumacea. Bergen (NO): Bergen Museum.

[SBMNH] Santa Barbara Museum of Natural History. 1994a. Taxonomic atlas. Vol. 1: Platyhelminthes, and Nemertea. Santa Barbara (CA): Santa Barbara Museum of Natural History.

[SBMNH] Santa Barbara Museum of Natural History. 1994b. Taxonomic atlas. Vol. 2: Porifera. Santa Barbara (CA): Santa Barbara Museum of Natural History.

[SBMNH] Santa Barbara Museum of Natural History. 1994c. Taxonomic atlas. Vol. 4: Oligochaeta and Polychaeta: Phyllodocidae to Paralacydoniidae. Santa Barbara (CA): Santa Barbara Museum of Natural History.

[SBMNH] Santa Barbara Museum of Natural History. 1995a. Taxonomic atlas. Vol. 5: The Annelida, Part 2, Polychaeta: Phyllodocida (Syllidae and scale-bearing families), Amphinomida and Eunicida. Santa Barbara (CA): Santa Barbara Museum of Natural History.

[SBMNH] Santa Barbara Museum of Natural History. 1995b. Taxonomic atlas. Vol. 12: the Crustacea, Part 3, The Amphipoda. Santa Barbara (CA): Santa Barbara Museum of Natural History.

[SBMNH] Santa Barbara Museum of Natural History. 1995c. Taxonomic atlas. Vol. 13: The Bryozoa. Santa Barbara (CA): Santa Barbara Museum of Natural History.

[SBMNH] Santa Barbara Museum of Natural History. 1996a. Taxonomic atlas. Vol. 6:

The Annelida, Part 3, Polychaeta: Orbiniidae to Cossuridae. Santa Barbara (CA): Santa Barbara Museum of Natural History.

[SBMNH] Santa Barbara Museum of Natural History. 1996b. Taxonomic atlas Vol. 9: The Mollusca, Part 2: The Gastropoda. Santa Barbara (CA): Santa Barbara Museum of Natural History.

[SBMNH] Santa Barbara Museum of Natural History. 1996c. Taxonomic atlas. Vol. 14: Miscellaneous Taxa. Santa Barbara (CA): Santa Barbara Museum of Natural History.

[SBMNH] Santa Barbara Museum of Natural History. 1997a. Taxonomic atlas. Vol 10: The Pycnogonida, The Crustacea Part 1: The Decapoda and Mysidacea. Santa Barbara (CA): Santa Barbara Museum of Natural History.

[SBMNH] Santa Barbara Museum of Natural History. 1997b. Taxonomic atlas. Vol. 11: The Crustacea Part 2, The Isopoda, Cumacea and Tanaidacea. Santa Barbara (CA): Santa Barbara Museum of Natural History.

Scheffer P, Wiggins GB. 1986. A systematic study of the nearctic larvae of the *Hydropsyche morosa* group (Trichoptera: Hydropsychidae). Toronto (ON): Royal Ontario Museum.

Schmitt RJ, Osenberg CW. 1996. Detecting ecological impacts: concepts and applications in coastal habitats. San Diego (CA): Academic Press.

Schultz GA. 1969. The marine isopod crustaceans. Dubuque (IA): Wm. C. Brown Co.

Schuster GA, Etnier DA. 1978. A manual for the identification of the larvae of the caddisfly genera *Hydropsyche* Pictet and *Symphitopsyche* Ulmer in eastern and central North America. Cincinnati (OH): U.S. Environmental Protection Agency.

Schwinghamer P. 1981. Characteristic size distributions of integral benthic communities. Can J Fish Aquat Sci 38:1255-1263.

Schwinghamer P. 1983. Generating ecological hypotheses from biomass spectra using causal analysis: a benthic example. Mar Ecol Progr Ser 13:151-166.

Scott DA, Jones TA. 1995. Classification and inventory of wetlands. Vegetatio 118:1-16.

Scrimgeour GJ, Culp JM, Glozier NE. 1993. An improved technique for sampling lotic invertebrates. Hydrobiologia 254:65-71.

Simpson K, Bode R. 1980. Common larvae of Chironomidae (Diptera) from New York state streams and rivers, with particular reference to the fauna of artificial substrates. Bull New York State Mus 439:1-105.

Slack KV, Averett RC, Greeson PE, Lipscomb RG. 1973. Methods for collection and analysis of aquatic biological and microbiological samples. U.S. Geol Surv Tech Water Resour Invest Book 5.

Smith RI. 1964. Keys to marine invertebrates of the Woods Hole region: a manual for the identification of the more common marine invertebrates. Contribution No. 11. Woods Hole (MA): Systematics-Ecology Program, Marine Biological Laboratory.

Smith B, Wilson JB. 1996. A consumer's guide to evenness indices. *Oikos* 76:70-82.

Squires HJ. 1990. Decapod crustacea of the Atlantic coast of Canada. Ottawa (ON): Department of Fisheries and Oceans.

Steele DH, Brunel P. 1968. Amphipoda of the Atlantic and Arctic coasts of North America: *Anonyx* (Lysianassidae). *J Fish Res Bd Can* 25(5).

Stephenson W, Cook SD. 1980. Skewness of data in the analysis of species-in-sites-in-times. *Proc. Royal Soc. Queensland* 91:37-52.

Stewart KW, Stark BP. 1993. Nymphs of North American stonefly genera (Plecoptera). Denton (TX): University of North Texas Press

Suess MJ, editor. 1982. Examination of water for pollution control. A reference handbook. Vol. 3. Biological, bacteriological and virological examination. Oxford (UK): Pergamon Press.

Tattersall WM, Tattersall OS. 1951. The British Mysidacea. London (UK): Ray Society.

Taylor LR. 1961. Aggregation, variance and the mean. *Nature* 189:732-735.

Taylor BR. 1997. Optimization of field and laboratory methods for benthic invertebrate monitoring. Final report. Ottawa (ON): Natural Resources Canada, Canada Centre for Mineral and Energy Technology. Aquatic Effects Technology Evaluation Project 2.1.2.

Taylor BR, Bailey RC. 1997. Technical evaluation on methods for benthic invertebrate data analysis and interpretation. Final report. Natural Resources Canada, Canada Centre for Mineral and Energy Technology. Aquatic Effects Technology Evaluation Project 2.1.3.

Tetra Tech, Inc. 1986a. General QA/QC considerations for collecting environmental samples in Puget Sound. Final report. Seattle (WA): Report prepared for U.S. EPA, Region 10. Report TC-3991-04.

Tetra Tech, Inc. 1986b. Recommended protocols for station positioning in Puget Sound. Final report. Seattle (WA): Report prepared for U.S. EPA, Region 10. Report TC-3090-05.

Tetra Tech, Inc. 1987. Recommended protocols for sampling and analyzing subtidal benthic macroinvertebrate assemblages in Puget Sound. Final report. Seattle (WA): Report prepared

for U.S. EPA, Region 10. Report TC-3991-04.

Thiel TH. 1975. The size structure of the deep-sea benthos. *Int Rev Ges Hydrobiol* 60:575-606.

Thorp JH, Covich AP. 1991. Ecology and classification of North American freshwater invertebrates. San Diego (CA): Academic Press.

Tukey J. 1977. Exploratory data analysis. Reading (MA): Addison-Wesley.

Underwood AJ. 1997. Experiments in ecology: their logical design and interpretation using analysis of variance. Cambridge (UK): Cambridge University Press.

Ushakov PV. 1955. Polychaeta of the far eastern seas of the USSR. *Acad. Sci. USSR*. Translated 1965. Jerusalem (IL): Israel Program for Scientific Translations.

Vezina AF. 1988. Sampling variance and the design of quantitative surveys of the marine benthos. *Mar Biol* 97:151-155.

Walker EM. 1933. The nymphs of the Canadian species of *Ophiogomphus* Odonata, Gomphidae. *Can Ent* 65:217-229.

Walker EM. 1953. The Odonata of Canada and Alaska. Part I, General, Part II. The Zygoptera (damselflies). Vol. I. Toronto (ON): University of Toronto Press.

Walker EM. 1958. The Odonata of Canada and Alaska. Anisoptera. Vol. 2. Toronto (ON): University of Toronto Press.

Walker EM, Corbet PS. 1978. The Odonata of Canada and Alaska. Anisoptera, Macromiidae, Corduliidae, Libellulidae. Vol. 3. Toronto (ON): University of Toronto Press.

Wallace NA. 1919. The Isopoda of the Bay of Fundy. (University of Toronto Studies, Biol. Series No. 18). Toronto (ON): University of Toronto Library.

Waltz RD. 1994. Field recognition of adult *Acentrella* and *Heterocloeon* (Ephemeroptera: Baetidae). *Great Lakes Entomologist* 26:321-323.

Warwick RM. 1986. A new method for detecting pollution effects on marine benthic communities. *Mar Biol* 92:557-562.

Warwick RM, Clarke KR. 1993. Increased variability as a symptom of stress in marine communities. *J Exp Mar Biol Ecol* 172:215-226.

Warwick RM. 1988a. Analysis of community attributes of the macrobenthos of Frierfjord/Langesundfjord at taxonomic levels higher than species. *Mar Ecol Prog Ser* 46:167-170.

Warwick RM. 1988b. The level of taxonomic discrimination required to detect pollution effects on marine benthic communities. *Mar Pollut Bull* 19:259-568.

Warwick RM, Pearson TH, and Ruswahyuni. 1987. Detection of pollution effects on marine macrobenthos: Further evaluation of the species abundance/biomass method. *Mar Biol* 95:193-200.

Watling L. 1979. Marine fauna and flora of the northeastern United States: Cumacea. National Marine Fisheries Service, circular 423:1-22.

Watson J. 1997. A review of ecosystem classification: delineating the Strait of Georgia. Vancouver (BC): Fisheries and Oceans Canada, Pacific Region, Science Branch.

Weber CI, editor. 1973. Biological field and laboratory methods for measuring the quality of surface waters and effluents. Cincinnati (OH): U.S. Environmental Protection Agency. EPA-670/4-73-001.

Weinberg S. 1978. The minimal area problem in invertebrate communities of Mediterranean rocky substrata. *Mar Biol* 49:33-40.

Weiss HM. 1995. Marine animals of southern New England and New York: identification keys to common nearshore and shallow water macrofauna. Hartford (CT): State Geological and Natural History Survey of Connecticut, Department of Environmental Protection.

Westfall MJ Jr, May ML. 1996. Damselflies of North America. Gainesville (FL): Scientific Publishers.

Wiederholm T. 1980. Use of benthos in lake monitoring. *J Wat Poll Cont Fed* 52:537-547.

Wiederholm T, editor. 1983. Chironomidae of the holarctic region. Keys and diagnosis. Part 1 - Larvae. *Ent. Scand. Suppl. No. 19*. Sandby (SE): Scandinavian Society of Entomology.

Wiederholm T, editor. 1986. Chironomidae of the holarctic region. Keys and diagnosis. Part 2 - Pupae. *Ent. Scand. Suppl. No. 28*. Sandby (SE): Scandinavian Society of Entomology.

Wiggins GB. 1996. Larvae of the North American caddisfly genera (Trichoptera). 2nd Edition. Toronto (ON): University of Toronto Press.

Wood DM, Peterson, BV, Davies, DM, Gyorkos, H. 1963. The black flies (Diptera: Simuliidae) of Ontario. Part II: larval identification with descriptions and illustrations. *Proc. Entomol. Soc. Ontario*.

Wright JF. 1995. Development and use of a system for predicting the macroinvertebrate fauna in flowing waters. *Austral J Ecol* 20:181-197.

Table of Contents

5.	Effluent Characterization and Water Quality Monitoring	5-1
5.1	Overview	5-1
5.2	Sampling Frequency	5-1
5.3	Variables Measured	5-2
5.3.1	Calculation of Loadings	5-3
5.4	Sampling Locations	5-4
5.5	Reporting	5-5
5.6	Effluent Characterization	5-7
5.6.1	Sampling Methods and Laboratory Analysis	5-7
5.6.2	Method Detection Limit Change for Mercury	5-7
5.6.3	Methods for Determination of Thiosalts	5-8
5.7	Water Quality Monitoring	5-9
5.7.1	Preparation for the Field	5-9
5.7.2	Field Measurement of Water Quality Parameters	5-10
5.7.3	Collection of Water Samples for Laboratory Analyses	5-12
5.7.4	Sample Handling, Storage and Analysis for Water Quality Monitoring	5-13
5.7.4.1	Sample Handling and Preservation	5-13
5.7.4.2	Sample Shipping and Storage	5-14
5.7.4.3	Laboratory Analyses of Samples	5-14
5.7.5	Comparison of Water Quality Data in Exposure and Reference Areas	5-15
5.7.6	Estimation of Extent of Elevated Concentrations	5-17
5.8	Quality Assurance and Quality Control for Water Quality Monitoring	5-17
5.8.1	General Aspects of Quality Control in the Field	5-17
5.8.2	Field Aspects of Quality Assurance	5-18
5.8.3	Quality Assurance during Sample Handling, Shipping and Storage	5-18
5.8.4	Use of Blanks and Duplicate Samples	5-19
5.8.5	Quality Control in the Laboratory	5-22
5.8.5.1	Details on Quality Control Aspects of Laboratory Analyses	5-22
5.8.5.2	Good Laboratory Practices	5-22
5.8.5.3	Calibration Control	5-23
5.8.5.4	Precision	5-24
5.8.5.5	Accuracy and Bias	5-25
5.8.5.6	Recovery	5-25
5.8.5.7	Detection Limits	5-27
5.8.5.8	Data Reporting Conventions	5-28
5.8.5.9	Analytical Precision and Accuracy	5-28
5.8.6	Quality Assurance in the Laboratory	5-29
5.8.7	Recording and Reporting of QA/QC Information	5-29
5.8.7.1	Documentation	5-29

5.8.7.2	Reporting of QA/QC Information.....	5-30
5.9	References.....	5-32
Appendix 5-1: Justifications for Parameters for Effluent Characterization and Water Quality Monitoring		5-35

List of Tables

Table 5-1: Analytical parameters measured for effluent characterization and water quality monitoring.....	5-3
Table 5-2: Summary of recommended use of blanks and duplicate samples in the field and laboratory	5-21

5. Effluent Characterization and Water Quality Monitoring

5.1 Overview

The purpose of effluent characterization and water quality monitoring is to answer the following question: “What is the estimated mine-related change in contaminant concentrations in the exposed area?” Data generated from effluent characterization and water quality monitoring are used to:

- monitor changes in the quality of the effluent and environmental conditions in the receiving environment;
- provide an indication of variability in effluent quality and temporal or seasonal trends; and
- provide supporting environmental variables to help interpret results from the biological monitoring (fish and benthic invertebrate community survey) and the sublethal toxicity testing.

Effluent characterization is conducted by analyzing a sample of effluent to provide information on the concentrations of potential contaminants in the mine effluent.

Water quality monitoring is conducted by collecting and analyzing samples of water from the exposure area surrounding the point of entry of effluent into water from each final discharge point and from the related reference areas. In addition, samples of water are collected and analyzed from sampling areas in receiving environments where biological monitoring is completed (*Metal Mining Effluent Regulations* [MMER], Schedule 5, section [s.] 7).

5.2 Sampling Frequency

Effluent characterization and water quality monitoring shall be conducted 4 times per calendar year and not less than 1 month apart on the samples of effluent and water collected, while the mine is depositing effluent (MMER, Schedule 5, s. 7). It is recommended that, where possible, samples for effluent characterization and water quality monitoring be collected once in each calendar quarter. It is also recommended that samples for effluent and water be collected on the same day.

The following factors should be taken into consideration to decide when the aliquots of effluent samples are collected for effluent characterization:

- seasonal variability of effluent, based on composition and flow;

- the time of year when previous effluent samples have been collected;
- the time of year when sampling for water quality monitoring is being conducted; and
- the time of year when concentrations of the contaminants are expected to be highest in the exposure area.

For water quality monitoring, the following factors should be taken into consideration to decide when water samples are collected in the receiving environment:

- seasonal variability in water quality and flow in the exposure area;
- the time of year when concentrations in the exposure area of contaminants are expected to be highest;
- the time of year when previous water quality monitoring samples have been collected;
- the time of year when samples for effluent characterization are collected; and
- the time of year when the biological monitoring is conducted.

5.3 Variables Measured

Effluent characterization and water quality monitoring are conducted for the parameters listed in Table 5-1. If a mine does not use cyanide as a process reagent within the operations area, cyanide does not need to be recorded (MMER, Schedule 5, paragraph [p.] 7(1)(d). Also, if the concentration of total mercury is less than 0.10 µg/L in 12 consecutive samples, the recording of the concentration of total mercury may be discontinued (MMER, Schedule 5, subsection [ss.] 4(3)). It is recommended that a letter be sent to the appropriate authority in Environment Canada advising that the mine has fulfilled the above requirement. Table 5-1 also includes optional parameters recommended on a site-specific basis that the owner or operator of a mine may record as additional supporting information in order to conduct a more complete chemical characterization. In addition to the required parameters listed in Table 5-1, the measurement of some effluent parameters such as conductivity, sulphate or chloride concentrations may be useful as tracers to determine the extent of effluent mixing in the exposure area. In addition, the concentrations of calcium, magnesium, chloride, potassium, sodium, sulphate and dissolved organic carbon can be used to estimate the potential toxicity of some metals using the biotic ligand model approach (e.g., US EPA 2007; Reiley 2007). Appendix 5-1 includes the justification for the parameters for effluent characterization and water quality monitoring.

Table 5-1: Analytical parameters measured for effluent characterization and water quality monitoring

Effluent Quality Variables¹ (MMER, Schedule 5, s. 4)	Water Quality Variables¹ (MMER, Schedule 5, s. 7)	Site-specific Variables³ (not a regulatory requirement)
Aluminium	Aluminium	Fluoride
Cadmium	Cadmium	Manganese
Iron	Iron	Uranium
Mercury ⁴	Mercury ⁴	Total phosphorus
Molybdenum	Molybdenum	Calcium
Ammonia	Ammonia	Chloride
Nitrate	Nitrate	Magnesium
Hardness	Hardness ^{6,7}	Potassium
Alkalinity	Alkalinity ^{6,7}	Sodium
Selenium	Arsenic	Sulphate
Electrical conductivity ¹⁰	Copper	Thallium
	Lead	Total thiosalts
	Nickel	Water depth ²
	Zinc	Optical depth or transparency ²
	Radium 226 ⁹	Dissolved organic carbon
	Cyanide ⁵	Total organic carbon
	Total suspended solids	Water flow
	Dissolved oxygen concentration ²	
	Temperature ²	
	pH ^{6,7}	
	Salinity ^{7,8}	
	Selenium	
	Electrical conductivity ¹⁰	

¹ All concentrations are **total** values; dissolved concentrations may also be reported; effluent loading (MMER, s. 20) will also be calculated and reported.

² In situ measured parameters only for water quality monitoring (in receiving waters).

³ These other parameters are potential contaminants or supporting parameters; analysis is optional and may be added based on site-specific historical monitoring data or geochemistry data.

⁴ The recording of the concentration of total mercury in effluent may be discontinued if that concentration is less than 0.10 µg/L in 12 consecutive samples (MMER, Schedule 5, ss. 4(3)).

⁵ Cyanide does not need to be recorded if that substance is not used as a process reagent within the operations area (MMER, Schedule 5, s. 7(d)).

⁶ In the case of effluent that is deposited into freshwater, record the pH, hardness and alkalinity of the water samples.

⁷ In the case of effluent that is deposited into estuarine waters, record the pH, hardness, alkalinity and salinity of the water samples.

⁸ In the case of effluent that is deposited into marine waters, record the salinity of the water samples.

⁹ Radium 226 does not need to be recorded if the conditions of ss. 13(2) of the MMER are met.

¹⁰ Please refer to Environment Canada document: Guidance Document for the Sampling and Analysis of Metal Mining Effluent (EPS 2/MM/5) for methods. Temperature calibration, and compensation when measuring conductivity, should be done according to the manufacturer's specifications.

5.3.1 Calculation of Loadings

The MMER requires mines to record the monthly mass loadings of the MMER-prescribed deleterious substances (MMER, s. 20). As part of effluent characterization for environmental effects monitoring (EEM), it is also recommended that mines calculate effluent loadings of the other parameters monitored. Loading can be calculated by multiplying the mean effluent concentration of the parameter by the total volume of effluent discharge over the time period of interest (typically 1 year for effluent characterization).

5.4 Sampling Locations

Samples for effluent characterization shall be collected from each final discharge point, identified by the owner or operator of the mine and in accordance with the MMER (Schedule 5, ss. 4(2)).

Guidance for determining the sampling location(s) for effluent characterization is provided in the *Guidance Document for the Sampling and Analysis of Metal Mining Effluents: Final Report* (Fowlie et al. 2001). This document focuses primarily on methods for collection of effluent samples from point sources (end of the pipe). If samples are to be collected from nonpoint sources, proposed sample collection locations and methods should be discussed with the Authorization Officer.

Water quality monitoring is conducted by collecting samples of water from the exposure area surrounding the point of entry of effluent into water from each final discharge point and from the related reference areas (MMER, Schedule 5, ss. 7(1)). These sampling stations will not likely be the same sampling stations used for biological monitoring. In selecting sampling stations for the exposure area, the owner or operator of a mine should take into consideration the location where effluent concentrations are the highest.

In addition to the above, the owner or operator of a mine shall collect samples of water from the sampling areas selected for the fish population and fish tissue studies and the benthic invertebrate community studies. Therefore, water quality monitoring is conducted at the same time as the biological monitoring studies, should the mine be required to conduct these studies (MMER, Schedule 5, p. 7(1)a(ii)). The water samples are analyzed for the water quality monitoring variables outlined in Table 5.1.

It is recommended that at least 3 water samples be collected at each sampling station to provide an estimation of the variability and determine if concentrations of the contaminants are homogeneous within the sampling station. However, this may not be sufficiently robust to assess data statistically. More sampling stations within each area may help to better understand contaminant concentrations in the exposure area. At the minimum, a composite sample, consisting of few sub-samples spaced within the station, should be collected.

It is strongly recommended, where the benthic and/or fish sampling areas are not in close proximity to the sampling stations for water quality monitoring, that samples be collected concurrently at the sampling stations for the routine water quality monitoring. This will help to interpret the results of analyses of water samples collected in the benthic and/or fish sampling areas in comparison with the results of water samples collected under water quality monitoring.

5.5 Reporting

The results of effluent characterization and water quality monitoring shall be submitted to the Authorization Officer as part of an effluent and water quality monitoring report (MMER, Schedule 5, s. 8). As per Schedule 5, s. 8 of the MMER, a report on the effluent and water quality monitoring studies conducted during a calendar year shall be submitted to the Authorization Officer not later than March 31 of the following year. See Chapter 10 for information on electronic reporting of effluent and water quality monitoring data. The annual effluent and water quality monitoring report should include the following information.

(Note that in the list below, the regulatory requirements (as per the MMER, Schedule 5, s. 8) are written in *italics*; these requirements are followed by further recommendations and descriptions.

a) *The dates on which each sample was collected for effluent characterization, sublethal toxicity testing and water quality monitoring:*

- four dates for effluent characterization (*4 times per calendar year and not less than 1 month apart* (MMER, Schedule 5, ss. 7(2)), while the mine is depositing effluent;
- four dates for water quality monitoring (*4 times per calendar year and not less than 1 month apart* (MMER, Schedule 5, ss. 7(2)), while the mine is depositing effluent;
- dates for sublethal toxicity testing (*2 times each calendar year for 3 years and once each year after the third year, with the first testing to occur on an effluent sample collected not later than 6 months after the mine becomes subject to section 7 of these regulations* (MMER, Schedule 5, s. 6)). The sublethal toxicity testing date(s) should match the date(s) for effluent characterization, as the sublethal toxicity sample must be an aliquot of the effluent characterization sample; and
- if the required number of tests were not conducted, indicate the reason why (i.e., the number of days that the effluent was being discharged or the habitat conditions that prevented the collection of effluent characterization and/or water quality monitoring samples).

b) *The locations of the final discharge points from which samples were collected for effluent characterization, noting that effluent characterization is conducted at ALL identified final discharge points (FDPs).*

c) *The location of the final discharge point from which samples were collected for sublethal toxicity testing and the data on which the selection of the final discharge point was based, in compliance with the MMER, ss. 5(2):*

- Indicate from which FDP the effluent was collected for the sublethal toxicity testing.
- Indicate why that FDP was chosen, if there is more than one FDP at the mine site (e.g., effluent that discharges into a sensitive receiving environment, has the greatest mass loading).

d) *The latitude and longitude of sampling areas for water quality monitoring, in degrees, minutes and seconds, and a description that is sufficient to identify the location of the sampling areas:*

- If units other than latitude and longitude were taken (e.g., UTM) at the sampling station, there are Web-based tools provided by Natural Resources Canada that can be used to convert them.
- Provide a written description (possibly supplemented with maps) of the sampling station that is sufficient to identify the location within the sampling areas. For example, “water collected under first bridge.” This description should allow for ease of re-sampling at the same stations.

e) *The results of effluent characterization, sublethal toxicity testing and water quality monitoring:*

- Include the results from all analyses completed on effluent (chemical and physical parameters), sublethal toxicity testing and water quality monitoring.
- Include results from all required parameters, as well as any optional site-specific parameters that were measured (see Table 5.1).
- For sublethal toxicity testing, the laboratory reports should be included as an appendix in the annual report.

f) *The methodologies used to conduct effluent characterization and water quality monitoring, and the related method detection limits:*

- Some sampling methods are outlined in the *Guidance Document for the Sampling and Analysis of Metal Mining Effluent: Final Report* (Fowlie et al. 2001); available at <http://dsp-psd.pwgsc.gc.ca/Collection/En49-24-1-39E.pdf>.

- Indicate the methodology used (e.g., inductively coupled plasma combined with mass spectrometry [ICP-MS], graphite furnace atomic absorption spectrometry [GFAAS]) for effluent characterization and water quality monitoring.
- Indicate the method detection limits for the methodology used—for MMER deleterious substances, the method detection limits identified in Schedule 3 of the MMER should be met. See section 5.6.2 of these guidelines for the method detection limit for mercury. Note that the Canadian Council of Ministers of the Environment's *Canadian Environmental Quality Guidelines* (CCME 1999) (Chapter 4: Canadian Water Quality Guidelines for the Protection of Aquatic Life) or additional provincial water quality guidelines should also be considered.
- Indicate whether the Canadian Water Quality Guidelines for the Protection of Aquatic Life (CCME 1999 - Chapter 4) or additional provincial water quality guidelines are met.

g) *A description of quality assurance and quality control measures that were implemented and the data related to the implementation of those measures:*

- Provide a brief description of the quality assurance / quality control (QA/QC) measures that were taken and the results of such measures with respect to collection of the effluent and water sampling, shipping and storage.
- See sections 7.3–7.5 of Fowlie et al. (2001) and section 5.8 below for further information.

Since effluent samples for effluent characterization are aliquots of samples collected for effluent compliance monitoring, the measurements of pH and the concentrations of the deleterious substances (arsenic, copper, total cyanide, lead, nickel, zinc, radium 226 and total suspended solids) should be available as part of the effluent characterization and water quality monitoring reports from each mine.

5.6 Effluent Characterization

5.6.1 Sampling Methods and Laboratory Analysis

Since effluent samples for effluent characterization are aliquots of samples collected for effluent compliance monitoring as part of the MMER, the sampling and chemical analysis considerations and recommended procedures provided in the *Guidance Document for the Sampling and Analysis of Metal Mining Effluent: Final Report* (Fowlie et al. 2001) apply also to effluent characterization conducted as part of the EEM program. The volume of sample taken should be sufficient to allow for all required analyses and tests plus associated quality control samples (e.g., field duplicates, laboratory replicates and spiked sample).

5.6.2 Method Detection Limit Change for Mercury

The method detection limit for mercury in effluent has been changed to 0.01 µg/L (0.00001 mg/L) so that the concentration of 0.1 µg/L specified in Schedule 5, s. 9(c) of the MMER can be detected with confidence. Analytical methodologies suitable to achieve this level of detection include cold vapour atomic absorption spectrometry (CVAAS), cold vapour atomic fluorescence spectrometry (CVAFS) and inductively coupled plasma mass spectrometry (ICP-MS).

5.6.3 Methods for Determination of Thiosalts

Total thiosalts is an optional site-specific parameter that may be measured in mine effluent; however, information on sampling or analysis is not available in Fowlie et al. (2001). Thiosalts are soluble sulphur/oxygen ions that form as a result of the incomplete oxidation of sulphide minerals. They have the potential to be generated whenever sulphide minerals are in contact with oxygen, but in practice they tend to be formed during the processing of ores bearing sulphide minerals. If thiosalts occur in effluent, once the effluent is discharged the oxidation of the thiosalts is completed; this results in the generation of sulphuric acid and the lowering of pH in the exposure area. Such pH alterations in receiving waters could also be related to low levels of thiosalts and to thiosalt speciation, which cannot be addressed entirely with commonly used analytical techniques (Vigneault et al. 2002). At a concentration of 10 ppm, thiosalt degradation can still potentially drop the pH to 3.7 in unbuffered receiving water (Vigneault et al. 2002). Information related to thiosalt speciation may be required to predict pH depression, since individual thiosalt species can produce different amounts of acidity and are stable in markedly different conditions.

Despite the ability of thiosalts to alter pH in receiving water bodies, toxicity due to thiosalts in mine effluent has been limited to a few sites. This may be due to the low toxicity of thiosalts to animals. Thiosalts are not expected to be acutely lethal in mine effluents, with lethal concentrations for Rainbow Trout higher than 800 mg/L (Schwartz et al. 2006). Sublethal toxicity testing suggests further that the sensitivity of aquatic species to thiosalts and the toxicity of the different anions composing thiosalts vary by an order of magnitude. Schwartz et al. (2006) reported that *Ceriodaphnia dubia* was the most sensitive EEM test species, with a 25% inhibition concentration (IC₂₅) of 60 mg/L for thiosulphate. They further noted that tetrathionate was much less toxic than thiosulphate. Few mine sites in Canada have known thiosalt problems, but the potential for thiosalt generation may exist at many mine sites. As such, total thiosalt determination is optional for effluent characterization and water quality monitoring in the EEM program.

The total concentration of thiosalts is most commonly determined with a titration method having a detection limit around 10 ppm (expressed as thiosulphate) (Makhija and Hitchen 1979). Thiosulphate is stable at neutral pH and unstable at low pH, while the opposite is

true for polythionates. Ion chromatography can be used to determine the concentration of different thiosalt species in synthetic solutions in the ppb range, but it is difficult to apply to field samples because of the instability of thiosalt. In order to better predict the environmental impacts of thiosalts and thiosalt degradation, more information regarding in situ speciation and measurement methods with lower detection limits are required.

The main concern with this method is that the samples should be analyzed within 24 hours. Given that every available preservation method has limitations, there is in fact no substitute for immediate analysis (O'Reilly et al. 2001). As a result, total thiosalts analyses should ideally be done on-site. These analytical capabilities are likely restricted to sites with known thiosalt problems. Alternatively, samples may be frozen immediately after collection and be analyzed within 7 days. Longer storage time of frozen samples may affect thiosalt stability. Alternatively, an anion exchange resin can also be used to preconcentrate and preserve thiosalts (Drushel et al. 2003; Vigneault et al. 2002).

5.7 Water Quality Monitoring

5.7.1 Preparation for the Field

The reagents for cleaning, operating or calibrating equipment and collecting, preserving and/or processing samples should be handled by appropriately qualified personnel, and the appropriate data for health and safety (e.g., Material Safety Data Sheets) should be available.

Written protocols and standard operating procedures (including QA/QC requirements) should be readily accessible at all times, to ensure proper and safe operation of equipment. Data forms and logbooks should be prepared in advance so that field notes and data can be quickly and efficiently recorded. Extra forms should be available in the event of a mishap or loss. These forms and books should be waterproof and tear resistant. Under certain circumstances, audio or audio/video recordings may prove valuable.

All equipment used to collect and handle samples should be cleaned and all parts examined to ensure proper functioning (e.g., on-site assembly or operation) prior to going into the field. A repair kit should accompany each major piece of equipment in case of equipment failure or loss of removable parts. Backup equipment, batteries and sampling gear should be available. Sampling equipment used for field measurements of water quality parameters should be properly calibrated or standardized according to the manufacturer's recommendations.

All sample containers and required preservatives should be provided by the laboratory hired to conduct the analyses of the samples. Bottles should preferably be unused and purchased as certified clean. If bottles are reused, they should be cleaned by a documented cleaning procedure with a bottle lot number-control system, and cleanliness should be demonstrated by the use of blanks (Fowle et al. 2001).

Storage, transport and sample containers, including extra containers in the event of loss or breakage, should be pre-cleaned and labelled appropriately (i.e., with a waterproof adhesive label to which the appropriate data can be added with an indelible ink pen capable of writing on wet surfaces). The containers should have lids that are fastened securely and the appropriate container lids and lid liners should be used to prevent contamination (e.g., lid liners should be lined with an inert material like Teflon[®], not paper or cardboard). A sample-inventory log and a sample-tracking log should be prepared in advance of sampling. The responsibility for these logs should be assigned to one individual who will be required to monitor the samples from the time they are collected until they are analyzed and disposed of or archived.

5.7.2 Field Measurement of Water Quality Parameters

Standard in situ water quality parameters are dissolved oxygen, pH, hardness, alkalinity, conductivity, water temperature and salinity (marine and estuarine environments). Total water depth at the sampling area and water depth from where the water sample was collected should be recorded. Optical depth or transparency should also be measured in the field. Current flow should also be measured in riverine environments. Measurements of standard water quality parameters can be taken in the water directly, from a sample container in the boat or on shore immediately after collection of the sample, as long as the water is collected at the appropriate depth. If dissolved oxygen measurements are conducted on the shore, special care should be taken to ensure that air is not introduced into the sample.

In shallow water bodies ≤ 2 m deep, standard water quality parameters need only be measured at mid-depth. If the depth ranges from 2 to 4 m, standard water quality measurements should be taken at 2 depth intervals: approximately 25 cm above the bottom and 25 cm below the surface. In deeper bodies of water, measurement of standard water quality parameters should be taken throughout the water column. Information on bottom depth and water column profiles of conductivity, pH, hardness, alkalinity, salinity, temperature and dissolved oxygen should be obtained along intervals of 1 to 5 m (depending on total depth). For example, at a depth of 5 m, measurements should be taken every metre. At a depth of 25 m, measurements should be recorded at 5-m intervals.

For deep samples, a peristaltic sampler, with appropriate lengths of Teflon[®] tubing, should be used in preference to other types of pumps. If other types are used, they should be Teflon[®]-coated and non-metallic. Sampling should proceed from the least-contaminated to the most-contaminated station, with a weak nitric acid and distilled water rinse between stations. The solvent rinsate should be collected and returned to the laboratory for proper disposal. Laboratory blanks of the samplers should be run before and after use to demonstrate that no contamination is imparted to samples (Fowlie et al. 2001).

Profiles can be facilitated through the use of a data logger (or equivalent) equipped with a dissolved oxygen probe and associated stirrer, as well as pH, conductivity, depth and

temperature probes, which evaluate water column quality simultaneously. Such a unit is particularly useful for deeper evaluations (> 50 m). During profiling, the operator is able to visually review incoming data, noting particular areas of interest during descent and ascent of the unit (e.g., conductivity spikes, thermocline, unusual data records). This information is recorded either manually or directly stored in the data logger. To supplement computer records, parameter readings should be recorded manually onto field data sheets (every 2 or 5 m) depending on total depth profiled.

At shallow depths, hand-held meters are often the most convenient way to measure in situ water quality parameters. They are light, and several models are now available that can measure standard water quality parameters. The probes and the cables connecting them to the hand-held unit can range from 2 to 5 m, limiting the use of such a unit. These meters tend to require more regular maintenance and calibration, meaning extra care should be taken to make sure that the meters are in proper functioning order. Calibration and maintenance logs should be kept on file.

Water depth can be measured indirectly using a sonar-based fish finder, or directly using a calibrated tape, sounding cable or rod. Recommended accuracy is as follows:

Water depth less than 2 m:	recommended accuracy of ± 25 cm
Water depth of 2-10 m:	recommended accuracy of ± 50 cm
Water depth greater than 10 m:	recommended accuracy of ± 1 m

Optical depth is a measure of the transparency of water, and can be measured with a turbidity meter in the field or in the laboratory. Optical depth can also be measured using a Secchi disk. The disk is 20 cm in diameter, and is painted white in two opposite quarters and black in the other two. The disk is attached to a calibrated tape. To measure optical depth, the disk is lowered into the water in the shade until it has disappeared. It is then raised slowly, and the water depth at which it reappears is recorded. At least two measurements should be made at each station, and optical depth should be estimated based on the median value of the measurements. Measurements should be made at midday, and sunglasses should not be worn while measurements are made (Nielsen and Johnson 1983).

Water quality data should be screened on-site during sample collection to prevent the measurement and recording of false readings, as doing so will permit the use of alternative instrumentation or instrument checks in the event of equipment or sampling error. All sampling and monitoring equipment should be checked and calibrated daily, if necessary, to ensure good working condition.

It is recommended that additional field measurements and observations be recorded:

- sample number, replicate number, site identification (e.g., name);
- time and date of the collection of the sample;

- ambient weather conditions, including wind speed and direction, wave action, current, tide, vessel traffic, temperature of both the air and water, thickness of ice if present;
- sampling area location (e.g., positioning information) and location of any replicate samples;
- type of platform/vessel used for sampling (e.g., size, power, type of engine);
- name of personnel collecting the samples;
- details pertaining to unusual events that might have occurred during the operation of the sampler (e.g., possible sample contamination, equipment failure, unusual appearance, control of vertical descent of the sampler); and
- deviations from standard operating procedures.

5.7.3 Collection of Water Samples for Laboratory Analyses

Water samples collected in the field and sent to a laboratory for analysis make up the bulk of the water quality monitoring program and involve the analysis of metals, nutrients, major anions and cations, and several other general water quality variables.

Total metals analysis (total values) is required (MMER, Schedule 5, s. 4) during water quality monitoring, as studies often found no difference between measuring “total” and “dissolved” metals (ESG 1999). However, significant differences between total and dissolved metals can be found in some cases, and analyzing for metals in both the dissolved and total fractions could be relevant on a site-specific basis especially in the context of investigation of causes.

In general, samples should be collected at 2 depth intervals: the subsurface (epilimnion) and near bottom (hypolimnion) in order to obtain samples from both areas of the water column (above and below the thermocline). If the water depth is ≤ 2 m, it is sufficient to collect water samples only at mid-depth or at least 15 cm below the surface. Samples collected below the surface of the water can be collected by hand directly into the sample bottle.

Water collections at discrete depths should be facilitated through the use of appropriate samplers (e.g., Niskin sampler, non-metallic 2-16-L Van Dorn or 0.5-8-L Kemmerer samplers). For streams, depth-integrated samplers that are representative of the suspended sediment and related substances can be used. These samplers can be used from a boat, bridge or ice surface, and usually require two persons for safe operation. For very deep samples, a peristaltic sampler is preferred to other types. If other samplers are used, they should be Teflon[®]-coated.

The water sampler should be triple-rinsed with the water from the sampling station between each sample. In addition, it is recommended that sampling in the reference area be completed first to avoid any potential contamination of the sampler with water from the exposure area. The sampler should be double-rinsed with reagent-grade weak nitric acid between sampling areas, particularly if it is not possible to complete sampling in the reference area first. The

solvent residues should be collected and returned to the laboratory for proper disposal. Laboratory blanks of the samplers should be run before and after use to demonstrate that no contamination is imparted to the samples.

When collecting water samples, it is important to use as many of the following ultra-trace techniques and proper water sampling protocols as possible:

- Sampling should proceed from the least-contaminated to the most-contaminated station.
- Sample bottles and caps should be rinsed 3 times prior to water collection.
- No preservatives should be placed into the sampling bottles prior to sample collection.
- Samples should be collected with the bottle mouth facing up-current and away from the sample collector's hand.
- At no time should the inside of the sample container, the bottle mouth, or the inside of the container lid be touched by sample collectors, even while wearing disposable gloves.
- Sample collectors should wear unlined, powder-free latex or nitril gloves to avoid contamination of the sample.
- Label all samples immediately and clearly, and follow proper preservation techniques. Record all sampling data in the field notebook immediately.
- Caps of water containers should be held lid-down during sample collection.
- The sampling-point locations should be recorded.

5.7.4 Sample Handling, Storage and Analysis for Water Quality Monitoring

5.7.4.1 Sample Handling and Preservation

The *Guidance Document for the Sampling and Analysis of Metal Mining Effluents: Final Report* (Fowlie et al. 2001) contains information on sample handling recommendations regarding containers, preservatives and holding times for specific parameters. Where appropriate, preservatives should be added to the sample bottle immediately upon completion of the collection. The actual sample volumes required may vary depending on the needs of the laboratory.

Note that to reduce the number of samples collected, several analytes may be analyzed from one sample bottle. Prior to sample collection, the list of variables should be discussed with the laboratory to determine the number and type of sample bottles required.

When collecting samples, it is useful to have a checklist that lists the collection bottles, corresponding analytes, and whether or not a preservative is required. As a sample is

collected it should be checked off the list. In certain situations, a maximum holding time of 7–10 days (major cations and anions, nitrate/nitrite, dissolved organic carbon) may be problematic. If the shipping of a mine's water samples has been unavoidably delayed but the integrity of samples was retained, the Authorization Officer should be notified without delay.

5.7.4.2 Sample Shipping and Storage

It is recommended that samples be cooled to 4°C during collection and stored at the same temperature for shipping, to minimize degradation. Samples should also be refrigerated, and shipping coolers should be equipped with ice packs or bagged ice to ensure that samples are kept cold.

Samples should be transported to a laboratory as soon as possible after collection (within 24–48 hours maximum). Analyses should be completed within the accepted storage times, which will vary depending on the variable. Storage time is defined as the time interval between the end of the sample collection period and the initiation of analyses. All samples should be stored for as short a time interval as possible and under conditions that minimize sample degradation. Samples should be maintained at temperatures above their freezing point and under 10°C, with minimal exposure to light. Samples digested for metals analysis may be maintained in a sealed container and analyzed within 30 days. For additional information refer to Fowlie et al. (2001).

5.7.4.3 Laboratory Analyses of Samples

Laboratory analyses should be carried out in a qualified laboratory by trained personnel operating under quality-controlled conditions and using documented standard operating procedures. Laboratories contracted by the mining industry should be accredited under the International Organization for Standardization standard ISO/IEC 17025:2005 entitled "General requirements for the competence of testing and calibration laboratories," as amended from time to time. The analytical methods selected should be generally accepted and in common use in laboratories in Canada. The overall method principle should be peer-reviewed and published in a widely available publication so that it can be located easily for details.

The analytical methods selected should meet the criteria in this document plus any other objectives identified by the mine (or those acting on the mine's behalf) or Environment Canada. The project manager and the laboratory need to confirm what parameters of interest will be measured and that holding times can be met. The laboratory and analysis methods should be selected and discussed before the sample is collected, to ensure that the laboratory sample requirements are met.

The methods chosen should reliably measure the detection limits indicated for the deleterious substances identified in Schedule 3 of the MMER (i.e., any concentration above about one-tenth of the maximum authorized sample concentration (Fowlie et al.

2001). Normally accepted methods, method detection limits, and precision and accuracy objectives for metal mining effluent analysis are discussed in Fowlie et al. (2001). For the other required or site-specific recommended water quality parameters, for which detection limits are not specified, if there is a CCME Canadian environmental quality guideline (CCME 1999) for the variable measured, the chosen method's detection limits should be sufficiently low to determine if the parameters measured exceed these guidelines. CCME guidelines can be found at <http://ceqg-rcqe.ccme.ca>. Several provinces have also developed water quality guidelines, and in cases where both CCME and provincial water quality guidelines exist for a particular parameter, the provincial guidelines take precedence, although both should be reported.

5.7.5 Comparison of Water Quality Data in Exposure and Reference Areas

It is recommended that the Biological Interpretative Report include a comparison of water quality data in the exposure and reference areas. This comparison would examine all parameters included as part of water quality monitoring. In particular, this comparison should identify any parameters for which there are differences in measurements taken in the exposure and reference areas of more than a factor of 2. This comparison is intended to help with the interpretation of biological data in the interpretative report.

The factor of 2 for exceedance of concentrations from the reference area is intended to ensure that differences between exposed and reference area concentrations are real differences, and not just differences that may be attributed to such factors as low concentrations of target contaminants, analytical variability, small minimal sample size ($n = 4$) and seasonable variability. At sites where the reference area is on a different water body or watershed than the exposed area, the factor of 2 difference may not be applicable.

Determination of whether or not concentrations are different between the exposure and reference areas should be based on the median value of a minimum of 4 samples collected over a 12-month period from the same exposure and reference area locations. The median in a set of n measurements $y_1, y_2, y_3, \dots, y_n$ is defined to be the value of y that falls in the middle when the measurements are arranged in order of magnitude. If there is an even number of measurements, then the median is the value of y halfway between the two middle measurements. If larger data sets ($n \gg 4$) are available, determination of whether concentrations are elevated in the exposure area could be based on a statistical test such as mean or median greater than the 95% confidence interval or greater than 2 standard deviations. If there are adequate pre-mining water quality data in the exposure area, then pre-mining data may be used as a basis for comparison.

In cases where there are differences of more than a factor of 2, it is recommended that the mine estimate and report the geographical extent for which this condition exists,

based on expanded water quality monitoring or modelling. However, before completing such an estimate, there are a number of factors that should be considered:

- 1) *Site-specific water quality objectives*: If there is a site-specific water quality objective for a particular parameter, and that objective is exceeded in the exposure area, the extent of this exceedence should be determined, regardless of the concentration in the reference area.
- 2) *Water quality guidelines*: If there are water quality guidelines for a particular parameter (e.g., federal or provincial), and the concentrations of that parameter in the exposure area are greater than concentrations in the reference area by more than a factor of 2, and are greater than the water quality guideline, then the extent of this exceedence should be determined.

The CCME *Canadian Environmental Quality Guidelines* (CCME 1999 - Chapter 4: Canadian Water Quality Guidelines for the Protection of Aquatic Life) for water quality monitoring parameters are available at <http://ceqg-rcqe.ccme.ca/>. Several provinces have also developed water quality guidelines, and in cases where both CCME and provincial water quality guidelines exist for a particular parameter, the provincial guidelines take precedence, although both should be reported.

- 3) *Detection limits*: In cases where the generic water quality guideline for a parameter is close to the analytical method detection limit, and concentrations of the parameter in the study area are close to the guideline, a factor of 2 difference may not be meaningful (as a result of analytical uncertainty close to the detection limit). In such cases the Authorization Officer should be consulted. McQuaker (1999) provides a comparison of achievable detection limits with generic water quality guidelines and, for most parameters, method detection limits (MDL) significantly less than the water quality guideline (WQG) (at least a ratio of 1:10 MDL:WQG) are available. However, McQuaker concluded that there are some parameters (arsenic, cadmium, mercury, selenium, silver and cyanide) for which an MDL at least 10 times lower than the WQG is not currently achievable. As the MDL:WQG ratio decreases, the measurement uncertainty increases; beyond a ratio of 1:2, the results are not considered to be statistically significant.
- 4) *pH*: For pH, a difference of a factor of 2 may be particularly important, since the pH scale is logarithmic. If there is a site-specific objective for pH, and if the pH in the exposure area is outside the range specified in the site-specific objective, the geographical extent of pH values outside the range of the site-specific objective should be determined. If there is no site-specific objective for pH, and if the pH in the exposure area is more than 0.5 pH units different than the pH in the reference area and is also outside Canadian environmental guidelines (e.g., CCME 6.5 to 9.0), the geographical extent of the exposure area within which pH is more than 0.5 pH units different from the reference area should be determined. According to the *Canadian*

Environmental Quality Guidelines (CCME 1999), human activity should not alter the pH by more than 0.2 pH units in marine or estuarine environments.

- 5) *Location of the reference area:* At sites where the reference area is on a different water body or watershed than the exposure area, a difference of a factor of 2 may not be applicable. If it is felt that this is the case, the Authorization Officer should be consulted.

5.7.6 Estimation of Extent of Elevated Concentrations

Two methods may be used to estimate the extent of elevated concentrations:

- direct measurement
- modelling

1) Direct Measurement

Direct measurement requires an increase in the number of sampling stations within the exposure area to determine where within the exposure area concentrations of contaminant(s) of concern are no longer elevated. The number of additional stations needed would be determined on a site-specific basis, but generally, a minimum of 3 stations would be needed:

- the exposure area station used for water quality monitoring, and
- two or more further stations that bracket the location within the exposure area where the concentrations of parameters of concern are no longer expected to be elevated.

2) Modelling

If the seasonal variations of concentrations of the key parameter of concern in effluent and the exposure area are well understood, and if seasonal variations in effluent and receiving environment flow are well understood, it may be possible to predict the location within the exposure area where the concentrations of the parameter(s) of concern are no longer expected to be elevated.

5.8 Quality Assurance and Quality Control for Water Quality Monitoring

General aspects of quality assurance and quality control are discussed in Chapter 2.

5.8.1 General Aspects of Quality Control in the Field

General QC aspects of a field sampling program are as follows:

- All personnel involved in field procedures should have appropriate education and training.
- Sampling methods should be consistently applied among sites throughout the study.
- Samples should be collected according to standard operating procedures that should be available to personnel at all times during the field study.
- Sampling equipment should be appropriate for the habitat being studied, properly cleaned, and accompanied by the appropriate documentation (i.e., manual, calibration and maintenance schedule).
- All samples should be properly labelled with date, location, type, number and collector's name.
- Samples should be in the proper container with the appropriate preservative or fixative if necessary.
- Field technicians should maintain detailed field notes using indelible ink and waterproof notebooks.
- Personnel should use chain-of-custody / sample submission forms and custody seals for contaminant samples.
- Personnel should follow appropriate shipping and storage methods.
- Standardized field collection forms should be used during the field program.

5.8.2 Field Aspects of Quality Assurance

Field QA for water quality monitoring should be achieved through several methodologies, including duplicate readings, comparison of readings with known standards, collection of profile samples for analytical evaluation, and parameter evaluation using alternate equipment (e.g., Hanna CTD meter, thermometer).

Some of the most common quality problems are the result of mislabelling or switching bottles, failure to add proper preservatives, improper storage conditions, sample contamination from sampling equipment, and exceeding the holding time. Each sample should be clearly labelled in a manner that identifies the sample and distinguishes it from all other samples. Labels should be filled out in indelible ink and fixed to the sample container such that they will not fall off when wet or during transport.

The field logbook is an integral part of the sampling program and forms the basis of the sampling report. Items documented in the logbook are often highly relevant to the interpretation of the laboratory data. Any deviations from the sampling plan or any other observation about the sample or the sampling locations should also be noted in the logbook. Some common deficiencies in field logbooks include the failure to make planning notes, make notes at the time events occur, sign and date entries, and write legibly.

5.8.3 Quality Assurance during Sample Handling, Shipping and Storage

The Canadian Association for Environmental Analytical Laboratories (CAEAL) (currently the Canadian Association for Laboratory Accreditation [CALA 1991]) recommends the following with respect to QA during sample handling, shipping and storage:

- a) **Chain of Custody:** Chain-of-custody forms should be used in the transportation of samples, especially in cases where several contracted parties are involved in the sampling, shipping and analysis of the samples.
- b) **Sample Inspection:** The condition of each sample should be noted upon receipt. Discrepancies between required sample conditions and the observed conditions should be recorded in a logbook or on a computer file. It is preferable to preserve samples in the field immediately. However, the samples should be preserved immediately if submitted unpreserved, and a record made of the preservation methodology.
- c) **Sample Tracking:** Samples should be assigned a unique number or code to identify the sample in a tracking system. The sample tracking system should identify the sample, the source, the date of receipt, analyses, due date, and any other pertinent information. A computerized laboratory information management system (LIMS) is recommended for tracking samples in laboratories processing large numbers of samples for a variety of clients.
- d) **Sample Storage:** Samples should be stored in an assigned location in a refrigerator or sample storage area accessible only to authorized personnel. Samples should be refrigerated at 4°C, where applicable, and removed only for inspection, logging and analysis. The temperature of the refrigerator should be measured and recorded daily.

5.8.4 Use of Blanks and Duplicate Samples

The use of blanks and duplicate samples in the field and laboratory is an important component in a QC program.

Field blanks and field duplicates are essential throughout the execution of a field program involving the collection of water. Field QC samples are used to establish whether any errors are being introduced during the sampling process so that corrective action can be taken if necessary. Field QC samples are distinct from laboratory QC samples in that they measure sampling effects rather than laboratory effects.

Field blanks are used to check contamination from all potential sources of contamination of the sample. These include possible contamination of sample bottles, caps, preservatives, equipment, filter paper (if samples are to be filtered), atmospheric contamination, sampling techniques, and analysis. Field blanks are collected by obtaining blank water (i.e., deionized water) from the laboratory conducting the analyses,

transporting the water to the field, and taking it through all sample collection, handling and processing steps that the test samples undergo (e.g., transfer to a sample container, preservation, and exposure to the environment). Field blanks are transported, stored and analyzed in the same manner as test samples (McQuaker 1999).

Duplicate samples should be taken to verify analytical results and equipment reliance. Field duplicates are used to evaluate homogeneity of the sample site and the ability of the sampling system to take the sample the same way every time. A field duplicate is a completely separate sample, not a split of a single sample into two bottles. Duplicate samples should be treated as blind samples, and are not identified to the laboratory.

The last type of QC sample is the trip blank, also referred to as travel or transport blanks. Trip blanks are used to check contamination from sample bottles, caps and preservatives during transport, storage and analysis. A sample bottle is filled in the laboratory with blank water (i.e., deionized water) and preserved in the same manner as the test samples (Fowle et al. 2001). Trip blanks are transported to the field with regular sample bottles and submitted to the laboratory unopened, together with the test samples. They are opened at the time of analysis, and analyzed in the same manner as the samples (McQuaker 1999).

Field and trip blanks as well as duplicate field samples should be collected at a frequency of 5-10% of the total number of samples. Therefore, if a total of 10 water quality areas were being sampled, only one of each of the QC samples would be needed from each station. This proportion can be increased if necessary, to monitor errors due to sampling and matrix homogeneity. If field and trip QC samples are not used, any inaccuracy introduced due to sampling will go undetected or be inappropriately attributed to the analytical laboratory. The use of blanks and duplicate samples in the laboratory is further discussed in section 5.8.5. Table 5-2 summarizes recommended use of blanks and duplicate samples in the field and the laboratory, for larger sampling programs. For routine sampling, with one station from the exposure area and one from the reference area, it is recommended that a single field blank be submitted together with the test samples. In such cases, these samples will be analyzed by the laboratory as a batch, together with samples from other clients. The laboratory will achieve necessary internal QC using the complete batch.

Table 5-2: Summary of recommended use of blanks and duplicate samples in the field and laboratory. Numbers are based on a batch of 20 samples or less.

Parameter	Number of Samples	Internal or Field QC	Control Limits	Description
Field blank	1	Field		Checks contamination as a result of sample handling. One per day per matrix.
Trip blank	1	Field		Tests validity of sample preservation and storage conditions. One per day per matrix.
Field duplicate	1	Field		Used to evaluate homogeneity of the sample site and the ability of the sampling system to take the sample the same way every time.
Method blank	1	Internal	< detection limit (D.L.) or < 0.1 of sample concentration	Checks contamination from reagents and procedures ^a
Laboratory duplicate sample	1	Internal		Checks precision of sampling process. One per day per matrix type. ^a
Glassware proof	1	Internal	< D.L. or < 0.1 of sample concentration	Checks contamination of lab glassware used during processing ^a
Standard reference material (SRM)	1	Internal		Checks accuracy of method ^a
Matrix spike	1	Internal	75-125%	Used interchangeably with SRM ^b
Calibration control:				
Within-run (blank and mid-range standard)	1	Internal	10% drift max.	Statistical control over calibration can be confirmed between runs by means of two control standards, A and B, and within-run by means of blanks and mid-range standards (King 1976).
Between runs (20% and 80% of full scale)	2 per run	Internal	± 5% of target value	

^a Intrinsic to every batch of 20 samples

^b Used interchangeably with SRM if SRM is not available

5.8.5 Quality Control in the Laboratory

The following are general QC aspects of laboratory analyses performed:

- Data should be verified and validated through transcription checks; chemical data will be verified by reference to the analytical laboratory QA reports accompanying the data.
- Data analyses will be repeatable and robust and will be cross-checked with data quality objectives.
- Data analyses will be rigorous and defensible and should include the rationale for all statistical analyses and data transformations.

5.8.5.1 Details on Quality Control Aspects of Laboratory Analyses

Analytical QC procedures are designed to demonstrate statistical control over calibration, precision, accuracy/bias, and recovery (CALA 1991).

Statistical control over these parameters can be demonstrated by running specific QC samples during each analytical run. The results of these QC samples are compared statistically with confidence intervals calculated from historical data. These confidence intervals or control limits are normally calculated at 3 standard deviations (SDs) of the mean of the controlled variable. Warning limits are frequently set at 2 SDs. Indicators of a run considered out of control include the following:

- two successive results for method blanks, laboratory duplicates, standard reference materials, spiked blanks, calibration control samples, or organic surrogate recoveries;
- one of these results outside of the control limits.

QC data can be plotted on appropriate control charts. Control charts are graphic presentations of the QC data as a function of time or consecutive run number. Control charts demonstrate trends in time and provide graphic evidence of long-term statistical control of the analysis. Control limits and control charts are described in detail in ASTM (1986).

5.8.5.2 Good Laboratory Practices

Well-established good laboratory practices (GLPs) should be followed. The following is a brief listing of recommended laboratory practices (a description of GLPs can be found in greater detail in ELAP 1988):

- Records on reagent preparation should be maintained in a logbook. Prepared

reagent containers should be labelled with the reagent, its date of preparation, the expiry date, and the person responsible.

- Instruments should be maintained or serviced on a regular basis. Maintenance records should be kept in a logbook.
- Written instructions should be available for all instruments.
- Standard procedures for cleaning glassware and containers should be followed.
- Routine checks of the purity of the distilled water should be conducted and documented. Distilled/deionized water should be checked on a conductivity meter at least daily.
- Chemical reagents should meet the purity requirements of each analytical method.
- Reagents and solvents should be stored according to the manufacturer's directions.
- Working standards and stock solutions should be checked to determine changes in concentration.
- Reagents should be prepared and standardized against primary reference standards.
- The temperatures of all refrigerators and incubators should be checked daily and temperature excursions should be recorded.
- Each oven should have a dedicated thermometer and the temperature should be checked prior to and following each use.
- Proper volumetric glassware should be used.
- Glassware should be cleaned according to specifications of the method.
- Gas cylinders should be replaced at 700–1400 kilopascals (kPa).
- Laboratory personnel should have appropriate training in analytical laboratory procedures, and in the particular analysis for which they are responsible.

5.8.5.3 Calibration Control

Statistical control over calibration can be confirmed between runs by means of two control standards, A and B, and within-run by means of blanks and mid-range standards.

- a) **Between-run Calibration Control:** Two control standards, A and B, can be used to analyze and control between-run changes in calibration, once at the beginning of each analytical run. These standards are made up and maintained independently of the calibration standards and are normally chosen to be about 80% and 20% of full scale, respectively. Results are accumulated over many runs and the sums ($A + B$) and differences ($A - B$) are plotted on control charts. During a specific run, a significant change in the sum ($A + B$) from the historical mean implies that a significant change in intercept has occurred, other factors remaining constant. A significant change in the difference ($A - B$) implies a significant change of slope, other factors remaining constant. Control and warning limits for $A - B$ are calculated for the mean and the SD of the population of differences:

$$\text{Upper and lower warning limits (UWL, LWL)} = X_{A-B} \pm 2 \text{ SD}_{A-B}$$

Upper and lower control limits (UCL, LCL) = $X_{A-B} \pm 3 SD_{A-B}$

Control and warning limits for A + B are similarly calculated using the same SD:

$$UWL / LWL = X_{A+B} \pm 2 SD_{A+B}$$

$$UCL / LCL = X_{A+B} \pm 3 SD_{A+B}$$

The run should not proceed until it is shown that A + B and A - B are within control limits. Control limits should not exceed $\pm 5\%$ of the average value for A + B and A - B.

- b) **Within-run Calibration Control (Inorganic Analyses):** Within-run changes in calibration attributable to slope and baseline drift should be checked at regular intervals. This can be accomplished by use of a mid-range standard and reagent blank run after every 20 samples. Control limits should be established by each laboratory for each procedure. The drift should not exceed 10%. If a greater drift is detected, the analysis should be stopped, the instrument recalibrated, and samples run after the last acceptable check sample and blank are reanalyzed.
- c) **Within-run Calibration Control (Organic Analyses):** In organic analyses by gas chromatography (GC), within-run changes in calibration should be checked by injection of a mid-level check standard at a frequency of 5% or every 12 hours. This injection is compared to the initial calibration by calculating the percent deviation in the response factor of each analyte in the check standard to the average response factor determined during the initial calibration. If the relative percent difference is greater than 25%, the calibration check should be repeated. If the repeated check standard still has a relative percent deviation greater than 25%, corrective action is recommended.

5.8.5.4 Precision

Precision is the degree of variation among individual measurements of the same variable using a specific analytical method, and is usually expressed as the SD of replicates (US EPA 1990). Statistical control of analytical precision is maintained by analyzing within-run duplicates at a frequency of at least 10%. Laboratory duplicates are separate aliquots split in the laboratory from a single sample.

The absolute difference between within-run duplicates is compared to a control limit determined from historical data. To obtain these control limits, the results of duplicate analyses are accumulated over many runs and sorted according to concentration ranges.

Convenient concentration ranges are 0-20%, 20-50%, and 50-100% of full scale (King 1976). Within each concentration range, control limits for the absolute difference between within-run duplicates is determined from the formula:

$$UCL = D_4 \times R$$

where D_4 (3.267) is a statistical factor and R is the mean difference between duplicates (ASTM 1986; Taylor 1987).

If the difference between laboratory duplicate analyses exceeds the upper control limit, the situation should be evaluated to determine the most appropriate corrective action.

5.8.5.5 Accuracy and Bias

Accuracy is the degree of agreement between an observed value and the true value as determined by analysis of an accepted reference material (US EPA 1990). The converse of accuracy is the degree of systematic error in the analysis, i.e., the bias. Accuracy is controlled by means of method blanks and certified reference materials. Information on recommended quality control for inorganic analyses can be found in CALA (1991).

- a) **Method Blanks:** A method blank is an aliquot of reagent water equivalent in volume to the samples being processed and run in exactly the same manner as the samples. The method blank quantifies the level of contamination introduced to the samples during sample processing and analysis. Method blanks should be analyzed at a frequency of 10% or 1/run, charted, and controlled at ± 2 SD (warning limits) and ± 3 SD (control limits). If a method blank is judged out of control and contaminated, those samples processed with the blanks and greater than the detection limit should be repeated for the variable(s) affected. In general, a method blank is considered free of contamination if the analysis yields results less than the detection limit or less than 0.1 times the level found in all associated samples (CALA 1991).
- b) **Standard Reference Materials:** SRMs are samples available in different matrices that have been extensively analyzed by several laboratories and have concentrations certified by standard-setting organizations such as the National Institute of Science and Technology, the U.S. EPA, the National Water Research Institute of Environment Canada and the National Research Council. When available, an SRM should be analyzed at a frequency of 5% or 1/run (CALA 1991; King 1976). The matrix and concentration of the SRM should be as close as possible to the samples being analyzed. The results of SRMs should be accumulated, and control and warning limits determined as ± 3 SD and ± 2 SD, respectively.

5.8.5.6 Recovery

Recovery of the analyte over the entire analytical process is determined from matrix spikes, spiked blanks, and surrogate spikes.

- a) **Matrix Spike:** A matrix spike is a separate aliquot of a randomly chosen sample to

which is added all the analytes of interest before processing of the sample. Analysis of a matrix spike gives an indication of the recovery efficiency obtained for the matrix particular to that sample. The sample should be spiked with all the analytes of interest at a concentration as close as possible to that concentration, giving a response equal to the mid-level calibration standard. The spiking solution should be prepared from a stock source separate from that used for calibration. The recommended distribution of matrix spikes is 10% or 1/run. One method to calculate recovery is:

$$\% \text{ Recovery} = \frac{\text{Measured Conc.} - \text{Unspiked Conc.}}{\text{Spike Amount}} \times 100$$

The results of matrix spikes should be plotted on separate control charts for each matrix. In-house limits should be set on the basis of ± 3 SD on a minimum of 10 data points. In multi-parameter analyses, at least 90% of the analytes should have recoveries within the specified limits. Recoveries for inorganic analytes should fall within 75-125%. Recoveries for organic variables should fall within the limits specified in Table 4 of CALA (1991). If a matrix spike does not meet these criteria, the spike should be repeated. If the recoveries do not meet the criteria in the repeat analysis and there are no indications of other problems with the analysis, a matrix effect should be noted and reported.

- b) **Spiked Method Blank:** The spiked method blank is a separate aliquot of the same reagent water used for the method blank that is spiked with the compound of interest at a concentration as close as possible to the concentration of the mid-level calibration standard. The spiked method blank gives an indication of the reliability of a method without the matrix effects of real samples. The spiked method blank should be processed with and in the same manner as the samples. As with the matrix spike, the spiking solution should be prepared from stocks separate from those used for calibration.

In-house recovery limits should be calculated for the spiked method blank based on ± 3 SD and a minimum of 10 data points. Recoveries for inorganic analyses should fall within 75-125%. Recoveries for organic variables should fall within 70-120%. If a spiked blank recovery does not meet the criteria established, the spike should be repeated. If the spike still does not recover, the samples related to the spike should be repeated. If insufficient sample remains for a repeat analysis, the results should be reported and flagged as suspect with an explanation.

- c) **Internal Standards (Organic Analyses):** All analyses using GC should be performed using internal standards, or properly validated methods using external standards. An internal standard is a compound that behaves similarly in an analytical system as the compound of interest, but is unlikely to be found in the sample. Internal standards are added at the same level to all samples, standards, and control samples prior to measurement but after sample preparation. All analyte responses should be normalized for the internal standard response to correct for instrument variability in response to such

factors as varying injection volumes, temperature fluctuations, and final extract volume. The response of the internal standard in the sample measurement should be within 20% of the internal response of a calibration standard analyzed within the same 12-hour period. If this criterion is not met, the sample should be repeated. If upon reanalysis the criterion is still not met, the sample results should not be corrected for internal standard response and should be flagged with an explanation.

- d) **Surrogate Spikes (Organic Analyses):** A surrogate standard is a compound not expected to be found in the sample that behaves similarly to the analytes of interest during sample preparation and analysis. Where applicable, surrogates should be added to all samples (including QC samples) before sample preparation to indicate method performance and sample matrix effects. Analyses run by gas chromatography / mass spectrometry (GC/MS) should have at least two surrogates, while those run by GC should have at least one surrogate. The amount of surrogate added to all samples should be the same as that added to the calibration solutions. In-house control limits for surrogate recoveries are based on ± 3 SD on a minimum of 10 data points. In-house control limits for surrogate recoveries should be within 60-120%. If any surrogate is outside the expected recovery range, the sample should be reanalyzed. If, upon reanalysis, the surrogate recovery is still outside the permissible range, the results should be reported with a flag and an explanation.

5.8.5.7 Detection Limits

Detection limits should be reported as the method detection limit (MDL) as described by the U.S. EPA (1984). The MDL is defined as the minimum quantity of an analyte that should be observed to justify the claim to have detected the analyte with a specified risk (normally 5% or 1%) of making a false detection.

One method to calculate the MDL is from the SD of the analysis at the lowest concentration range:

$$\text{MDL} = t_{0.05, n-1} \times S$$

where: $t_{0.05, n-1}$ is the one tailed value of Student's t for a 5% risk of false detection, $n-1$ degrees of freedom, and S is the SD.

Ideally the SD is calculated from low-level replicate analysis on real samples having the same or similar sample matrix as the samples under consideration. This SD can be calculated from a minimum of seven replicates in the same run using the standard statistical formula (US EPA 1984). However, it is preferable to calculate S from between-within-run replicate pairs accumulated over many runs.

The SD of low-level replicate pairs accumulated over a large number of analytical runs is:

$$S = \frac{\sum S D^2}{2n}$$

where D is the individual replicate difference and n is the number of replicate pairs. A minimum of 40 replicate pairs is recommended (OMOE 1988). The value of either SD is then entered in the equation to calculate the MDL.

Values below the detection limit should be reported as < MDL, with the applicable MDL for that sample (Fowlie et al. 2001). There are three common approaches to deal with values that are < MDL when analyzing data: set the value at the MDL, half the MDL, or 0. For the purposes of the EEM program, half the MDL is currently used for all data analysis and interpretation. For additional information on how to interpret non-detectable data, refer to Helsel (2005a, 2005b) and Shumway et al. (2002).

5.8.5.8 Data Reporting Conventions

Established protocols for rounding off analytical results should be followed. If too many figures are rounded off before reporting, information is lost and real differences in the concentrations of samples from different locations or occasions may be concealed. QC may be on a coarser basis than is desirable, or necessary, with the result that values of the mean, SD or other statistics of a set of results may be biased. Conversely, when too many significant figures are reported, relatively small, statistically insignificant differences may appear falsely large (Hunt and Wilson 1986).

The SD of the analysis is the preferred criterion for deciding the number of significant figures (King 1989). The process of rounding off should ensure retention of the digit that is in the same decimal position as the most significant digit in the calculated SD. For example, if the analysis provides a value such as 12.345 and the calculated SD based on within-run replicate analysis at this concentration level is 0.32, the result should be truncated to 12.3.

5.8.5.9 Analytical Precision and Accuracy

Precision is the degree of agreement among replicate analysis of a sample, usually expressed as the SD. Reproducibility is the closeness of agreement between the results of measurement of the same parameter carried out under changed conditions of measurement. Reproducibility is the SD obtained measuring the same sample in different analytical runs and is called between-run precision. Between-run precision includes variability due to calibration on different days, instrument drift and many other factors.

Precision is affected by random errors and is a measurable and controllable parameter. Precision should be estimated for all analyses by processing separate sample aliquots through the entire analytical method. A laboratory should monitor their precision and be able to report precision using several days of data. For most parameters, the precision

should be within 10%. For total suspended solids, the precision should be within 15% at concentrations greater than 10 times the MDL. For pH, precision should be within ± 0.1 pH unit (MMER, Schedule 3).

Accuracy is the combination of bias and precision of an analytical method, which reflects the closeness of a measured value to the true value of a sample. Bias is a systematic error caused by something in the measuring system resulting in the data being high or low. Bias can be caused by a number of factors including contamination, mechanical losses, blanks, spectral interference, calibration errors or the influence of different operators. Accuracy is measured as percent recovery of known concentrations such as certified reference materials, spiked samples or reference samples prepared by the laboratory and analyzed as samples.

Whether data are considered accurate or inaccurate is relative to the final use of the data. A laboratory should monitor their accuracy and be able to report this using several days of data. For metals and most other parameters, accuracy should be within 10%. For total suspended solids, accuracy should be within 15% at concentrations greater than 10 times the MDL. For pH, accuracy should be within ± 0.1 pH unit (MMER, Schedule 3).

5.8.6 Quality Assurance in the Laboratory

QA encompasses a wide range of internal and external management and technical practices designed to ensure that data of known quality are commensurate with the intended use of the data.

External QA activities include participation in relevant inter-laboratory comparisons and audits by outside agencies. Outside audits may be based on performance in analysis of standard reference materials, or on general review of practices as indicated by documentation of sampling, analytical and QA/QC procedures, test results, and supporting data.

5.8.7 Recording and Reporting of QA/QC Information

5.8.7.1 Documentation

Documentation of all aspects of the analysis is recommended to confirm the quality and reliability of the analytical results. The owner or operator of a mine shall keep all records, books of account or other documents required by the *Metal Mining Effluent Regulations* at the mine's location for a period of no less than five years (MMER, s. 27). For each sample or batch of samples, information on the following is recommended:

- a) **Method Detection Limits:** If MDLs are different from the laboratory-determined MDLs (due to interference, dilutions, etc.), this should be recorded.
- b) **Sample Storage Times:** Records should be kept on the sampling date, date of receipt,

date of sample preparation, and date of analysis. This information is normally handled as part of the sample-tracking process.

- c) **Instrument Performance and Maintenance:** A log should be kept of instrument performance, including records of tuning and instrument response. Maintenance or service records should be kept for each instrument.
- d) **Quality Control Samples:** Records of duplicate analyses, blanks, spiked blank recoveries, surrogate recoveries, matrix spike recoveries and results from certified reference materials, and records of calibration and calibration checks should be maintained.
- e) **Sample Reception, Preparation and Analysis:** All anomalies in delivery, storage, condition, preparation and analysis of samples should be recorded. These include any deviations from standard operating procedures.

5.8.7.2 Reporting of QA/QC Information

Analytical results are reported as a test or analysis report and should include all relevant data needed to assess the validity of the data, including QA/QC components. The report should be accurate, clear, unambiguous and objective. Items that should appear in the report include:

- a title (Test Report, Report of Analysis, Quality Report);
- name, address and location of laboratory, and location where tested;
- unique identification of the report so it can be traced easily (serial number, group number);
- name and address of client;
- identification or description of the sample tested;
- condition of the test item (unpreserved, leaking bottle—where relevant);
- date of sample receipt, date of report;
- identification of the analysis method and description of any non-standard tests;
- reference to sample date and sampling method (grab sample, time-proportioned composite sample, etc.);
- deviations from the usual test method (filtering, pH, adjustment, standard addition, etc.);
- the analytical results with units clearly identified;
- statement indicating whether the results were corrected for blanks;
- QC data;
- identify if result is qualified (did not pass QC tests, sample size too small, etc.);
- signature of accountable person and date authorization;
- name of technician who completed the test;
- subcontractors clearly identified;
- updates or corrections to reports clearly identified; and

- the laboratory should notify clients if new information invalidates reports already issued.

Data below the analytical detection limit should be clearly reported as such along with the applicable MDL for that sample.

5.9 References

AQUAMIN. 1996. Assessment of the Aquatic Effects of Mining in Canada: Final report. Prepared by AQUAMIN Working Groups 7 and 8 for the AQUAMIN Steering Group.

[ASTM] American Society for Testing and Materials. 1986. Manual on presentation of data and control chart analysis. Committee E-11 on Statistical Methods. ASTM Special Technical Publication 15D. Philadelphia (PA): American Society for Testing and Materials.

[CALA] Canadian Association of Laboratory Accreditation. 1991. Code of practice and QA manual for laboratory analysis of sewage treatment effluent in support of the MISA program. Draft report prepared for CAEAL and the Ontario Ministry of the Environment by Zenon Environmental Laboratories.

[CCME] Canadian Council of Ministers of the Environment. 1999. Canadian environmental quality guidelines. Chapter 4: Canadian water quality guidelines for the protection of aquatic life. Hull (QC): Canadian Council of Ministers of the Environment. Available from: <http://ceqg-rcqe.ccme.ca/>

Druschel GK, Schoonen MAA, Nordstrom DK, Ball JW, Xu Y, Cohn CA. 2003. Sulfur geochemistry of hydrothermal waters in Yellowstone National Park, Wyoming, USA. III. An anion-exchange resin technique for sampling and preservation of sulfoxyanions in natural waters. *Geochem Trans* 4:12-19.

[ELAP] Environmental Laboratory Approval Program. 1988. Environmental Laboratory Approval Program certification manual. New York State Department of Health.

[ESG] Ecological Services Group. 1999. AETE synthesis report of selected technologies for cost-effective environmental monitoring of mine effluent impacts in Canada (AETE Project No. 4.1.4). Report for AETE program. Ottawa (ON): CANMET, Natural Resources Canada.

Fowlie P, Hart DR, Turle R. 2001. Guidance Document for the Sampling and Analysis of Mine Effluents: Final Report. Ottawa (ON): Environment Canada, Environmental Protection Service, Minerals and Metals Division. Report EPS2/MM/5. Available from: <http://dsp-psd.pwgsc.gc.ca/Collection/En49-24-1-39E.pdf>

Helsel DR. 2005a. Nondetects and data analysis: statistics for censored environmental data. Hoboken (NJ): Wiley-Interscience. 250 p.

Helsel DR. 2005b. More than obvious: better methods for interpreting nondetect data. *Environ Sci Technol* 39(20):419A-423A.

Hunt DTE, Wilson AL. 1986. The chemical analysis of water; general principles and techniques. 2nd edition. London (UK): The Royal Society of Chemistry.

[ISO/IEC] International Organization for Standardization. 2005. ISO/IEC 17025: 2005. General requirements for the competence of testing and calibration laboratories. Geneva (CH): ISO/IEC.

King DE. 1976. Quality control and data evaluation procedures. Section I. Analytical responsibility. Special report to Laboratory Services Branch, Ontario Ministry of the Environment.

King DE. 1989. Code of practice for environmental laboratories. Special report to the Ontario Ministry of the Environment. ISBN 0-7729-5874-2.

Makhija R, Hitchen A. 1979. The titrimetric determination of sulphate, thiosulphate and polythionates in mining effluents. *Anal Chim Acta* 105(1):375-382.

McQuaker NR. 1999. Technical evaluation on water quality design and analysis (AETE Project No. 3.1.1). Draft report for AETE program. Ottawa (ON): CANMET, Natural Resources Canada.

Nielson LA, Johnson DL. 1983. Fisheries techniques. Bethesda (MD): American Fisheries Society. 468 p.

[OMOE] Ontario Ministry of the Environment. 1988. Estimation of analytical detection limits (MDL). Report by the Ontario Ministry of the Environment. ISBN-0-7729-4117-3.

Ontario Ministry of the Environment and Energy. 1993. MISA draft development document for the *Effluent Limits Regulation for the Metal Mining Sector*. Toronto (ON): Queen's Printer for Ontario.

O'Reilly JW, Dicinoski GW, Shaw MJ, Haddad PR. 2001. Chromatographic and electrophoretic separation of inorganic sulfur and sulfur-oxygen species. *Anal Chim Acta* 432(2):165-192.

Reiley M. 2007. Science, policy, and trends of metals risk assessment at EPA: How understanding metals bioavailability has changed metals risk assessment at US EPA. *Aquat Toxicol* 84(2):292-298.

Schwartz M, Vigneault B, McGeer J. 2006. Assessing the potential toxicity of thiosalts in the context of the *Metal Mining Effluent Regulation*. Presentation made at the 33rd Aquatic Toxicity Workshop, Jasper, AB.

Shumway RH, Azari RS, Kayhanian M. 2002. Statistical approaches to estimating mean water quality concentrations with detection limits. *Environ Sci Technol* 36(15):3345-3353.

Taylor JK. 1987. *Quality assurance of chemical measurements*. Chelsea (MI): Lewis Publishers Inc.

[US EPA] United States Environmental Protection Agency. 1984. Definition and procedure for the determination of the method detection limit – Revision 1.11. Appendix B to Part 136. *Federal Register*. Vol. 49, no. 209. Oct. 26, 1984, Part VI, 40 CFR Part 136.

[US EPA] United States Environmental Protection Agency. 1990. Proposed glossary of quality assurance related terms. QAMS RD-680. Draft report.

[US EPA] United States Environmental Protection Agency. 2007. Aquatic life ambient freshwater quality criteria—Copper 2007 revision. EPA-822-F-07-001.

Vigneault B, Holdner J, Bélanger J. 2002. Validation of an anion exchange method for the preservation and analysis of thiosalt speciation in mining waste waters. Ottawa (ON): CANMET Mining and Mineral Science Laboratories. Division Report MMSL 03-002(TR).

Appendix 5-1: Justifications for Parameters for Effluent Characterization and Water Quality Monitoring

Deleterious Substances and pH as per Schedule 3 of the MMER	
Arsenic	• AQUAMIN Working Groups 7 and 8 (1996) recommended that arsenic be measured in effluent characterization. • Arsenic can occur in effluent from a wide range of mine types, including gold, base metal and uranium, and can be expected to occur across Canada. • The MISA draft development document ¹ (Ontario Ministry of the Environment and Energy 1993) stated that arsenic was found in 26% of the metal mine effluents sampled, with an average concentration of 0.036 mg/L. • Arsenic can bioaccumulate in fish and is known to be toxic to aquatic organisms. • The Canadian environmental quality guideline (CEQG) ² for arsenic for the protection of freshwater aquatic life is 0.005 mg/L (0.0125 mg/L for marine environments).
Copper	• AQUAMIN Working Groups 7 and 8 (1996) recommended that copper be measured in effluent characterization. • Copper can occur in effluent from a wide range of mine types, particularly gold and base metal, and can be expected to occur across Canada. • The MISA draft development document (Ontario Ministry of the Environment and Energy 1993) reported that for 39 Ontario effluent streams sampled for 12 months, the average copper concentration was 0.07 mg/L. • Copper is known to be toxic to aquatic organisms. • The CEQG for copper for the protection of freshwater aquatic life ranges from 0.002 to 0.004 mg/L, depending on the water hardness.
Lead	• AQUAMIN Working Groups 7 and 8 (1996) recommended that lead be measured in effluent characterization. • Lead can occur in effluent from a wide range of mine types, particularly base metal mines, and can be expected to occur across Canada. • The MISA draft development document (Ontario Ministry of the Environment and Energy 1993) stated that lead was found in 20% of the metal mining effluents sampled. The average lead concentration in the sampled effluent was 0.02 mg/L. • Lead is known to be toxic to aquatic organisms. • The CEQG for lead for the protection of freshwater aquatic life ranges from 0.001 to 0.007 mg/L, depending on the water hardness.
Nickel	• AQUAMIN Working Groups 7 and 8 (1996) recommended that nickel be measured in effluent characterization. • Nickel can occur in effluent from a wide range of mine types, particularly base metal and uranium mines, and can be expected to occur across Canada. • The MISA draft development document (Ontario Ministry of the Environment and Energy 1993) stated that nickel was found in 68% of the metal mine effluents sampled, with an average concentration of 0.14 mg/L. • Nickel is known to be toxic to aquatic organisms. • The CEQG for nickel for the protection of freshwater aquatic life ranges from 0.025 to 0.150 mg/L, depending on the water hardness.

¹ Note: the MISA draft development document is referred to throughout this document. The MISA document summarizes data from one year of comprehensive effluent characterization at mines across Ontario in the early 1990s. These data are not representative of the current state of effluent quality from mines across Canada. Reference: Ontario Ministry of the Environment and Energy. 1993. MISA Draft Development Document for the *Effluent Limits Regulation for the Metal Mining Sector*. Queen's Printer for Ontario, Toronto, Canada.

² Note: throughout this document, the *Canadian Environmental Quality Guidelines* (CEQG) are referred to in order to provide a preliminary assessment of concentrations in the effluent for their ecological significance and their potential effects on the receiving environment. Reference: Canadian Council of Ministers of the Environment. 1999. Updated 2001. Canadian Water Quality Guidelines for the Protection of Aquatic Life, available from: <http://ceqg-rcqe.ccme.ca/>.

pH	• AQUAMIN Working Groups 7 and 8 (1996) recommended that pH be measured in effluent characterization. • pH extremes can occur in effluent from a wide range of mine types and can be expected to occur across Canada. • pH often determines the solubility of metal species and, therefore, is linked to the toxicity of the effluent. • Extremes of pH are known to be toxic to aquatic organisms. • The CEQG for pH for the protection of freshwater aquatic life is 6.5 to 9.0 (7.0 to 8.7 for marine and estuarine environments).
Radium 226	• AQUAMIN Working Groups 7 and 8 (1996) recommended that radium 226 be measured in effluent characterization. • Radium 226 occurs primarily in effluent from uranium mines. However, it does not occur across Canada. • There is no CEQG for radium 226.
Total cyanide	• Cyanide is used as a process reagent at most gold mines and some base metal mines. • AQUAMIN Working Groups 7 and 8 (1996) recommended that cyanide be measured in effluent characterization, for mines that use cyanide as a process reagent. • The MISA draft development document (Ontario Ministry of the Environment and Energy 1993) stated that cyanide was found in 54% of the metal mine effluents sampled, with an average concentration of 0.084 mg/L in gold mining effluent and 0.006 mg/L in base metal mining effluent. • Cyanide is known to be toxic to aquatic organisms. • The CEQG for free cyanide for the protection of freshwater aquatic life is 0.005 mg/L.
Total suspended solids	• AQUAMIN Working Groups 7 and 8 (1996) recommended that total suspended solids be measured in effluent characterization. • Suspended solids can occur in effluents from all mine types, and occur across Canada. • The MISA draft development document (Ontario Ministry of the Environment and Energy 1993) stated that suspended matter was found in 80% of the metal mine effluents sampled, with an average concentration of 7 mg/L. • Suspended solids can kill fish by clogging their gills, and can affect fish habitat by smothering fish habitat, contaminating sediments, or reducing light penetration in receiving waters.
Zinc	• AQUAMIN Working Groups 7 and 8 (1996) recommended that zinc be measured in effluent characterization. • Zinc can occur in effluent from a wide range of mine types, particularly base metal mines, and can be expected to occur across Canada. • The MISA draft development document (Ontario Ministry of the Environment and Energy 1993) stated that zinc was found in 76% of the metal mine effluents sampled, with an average concentration of 0.07 mg/L. • Zinc is known to be toxic to aquatic organisms. • The CEQG for zinc for the protection of freshwater aquatic life is 0.030 mg/L.
Parameters Required in Schedule 5 of the MMER for Effluent Characterization and Water Quality Monitoring	
Alkalinity	• Alkalinity is a measure of the buffering capacity of water, and gives an indication of how sensitive water is to changes in pH. • Alkalinity is a factor affecting the fate and bioavailability of metals.
Aluminium	• AQUAMIN Working Groups 7 and 8 (1996) recommended that aluminium be measured in effluent characterization. • Aluminium occurs in a number of important rock-forming minerals, and tailings pond effluents from a range of mine types may contain dissolved aluminium ions as well as chemically bound aluminium in the form of clays and other alumino-silicate mineral particles. • The MISA draft development document (Ontario Ministry of the Environment and Energy 1993) stated that aluminium was present in 70% of the Ontario metal mining effluents sampled, with an average concentration of 0.20 mg/L. • The draft development document also stated that dissolved aluminium is not a significant component of most metal mining effluents. • The CEQG for aluminium for the protection of freshwater aquatic life ranges from 0.005 to 0.100 mg/L, depending on the water hardness. • Aluminium is toxic to aquatic organisms and its toxicity varies with pH. • Aluminium data may assist in the interpretation of the potential impacts of metals and other parameters.
Cadmium	• AQUAMIN Working Groups 7 and 8 (1996) recommended that cadmium be measured in effluent characterization.

	• Cadmium occurs in a relatively small range of ore types, but can be expected to occur at mines across Canada. • The MISA draft development document (Ontario Ministry of the Environment and Energy 1993) states that cadmium was found in 12% of the metal mine effluents sampled, with an average concentration of 0.003 mg/L. • Cadmium is known to be toxic to aquatic organisms and is bioaccumulative. • The CEQG for cadmium for the protection of freshwater aquatic life is 0.000017 mg/L. Note that a formula adjusting the guideline value based on hardness is given, i.e., cadmium guideline = $10^{(0.86[\log(\text{hardness})]-3.2)}$ (0.12 µg/L for marine environments).
Iron	• AQUAMIN Working Groups 7 and 8 (1996) recommended that iron be measured in effluent characterization. • Iron occurs in virtually all ore types, and occurs in mines across Canada. • The MISA draft development document (Ontario Ministry of the Environment and Energy 1993) stated that iron was found in 100% of the metal mine effluents sampled, with an average concentration of 0.45 mg/L. • Iron is toxic to aquatic organisms, and iron hydroxide precipitates can affect fish habitat. • The CEQG for iron for the protection of freshwater aquatic life is 0.30 mg/L. • Iron can also have an important influence on the fate of other contaminants. • Iron data may assist in the interpretation of the potential impacts of metals and other parameters.
Nitrogen compounds (ammonia and nitrate)	• AQUAMIN Working Groups 7 and 8 (1996) recommended that total ammonia be measured in effluent characterization. • Nitrogen compounds are used in explosives in most mines, and residue from these explosives can result in nitrogen compounds occurring in effluent. Nitrogen compounds can also occur in effluent as a result of the breakdown of cyanide. • During MISA development in Ontario, ammonia plus ammonium, total Kjeldahl nitrogen, and nitrate plus nitrite were measured in effluents. • “Ammonia plus ammonium” was found in all 50 (100%) of the metal mining effluents sampled. The average concentration of “ammonia plus ammonium” measured in the metal mining sector was 1.4 mg/L in base metal mining effluents and 6.3 mg/L in gold mining effluents. • “Total Kjeldahl nitrogen” was found in 96% of the metal mining effluents sampled. The average concentration of “total Kjeldahl nitrogen” measured in the metal mining sector was 8 mg/L. • Nitrate plus nitrite” was found in 90% of the metal mining effluents sampled. The average concentration of “nitrate plus nitrite” measured in the metal mining sector was 8.8 mg/L; • Nitrogen compounds can be toxic to aquatic organisms. In addition, nitrogen compounds are nutrients, and can lead to excessive plant growth. Excessive plant growth can lead to oxygen depletion, resulting in fish kills. • The CEQG for total ammonia for the protection of freshwater aquatic life ranges from 1.370 to 2.200 mg/L, depending in temperature and pH. • The proposed interim CEQG for nitrate is 13 mg/L in freshwater (16 mg/L for marine environments). • The CEQG for nitrite for the protection of freshwater aquatic life is 0.060 mg/L.
Mercury	• AQUAMIN Working Groups 7 and 8 (1996) recommended that mercury be measured in effluent characterization. • Mercury occurs in a range of rock types. It can occur at gold and silver mines, and less commonly at base metal mines, and is expected to occur across Canada. • Mercury can come from a range of sources, including airborne transport, natural sources and mine effluent. As a result, AQUAMIN Working Groups concluded that it is often difficult to determine the source of mercury contamination within an aquatic environment. This was the basis for recommending that mercury be included in effluent characterization. • The MISA draft development document (Ontario Ministry of the Environment and Energy 1993) stated that mercury was found in 14% of the metal mine effluents sampled, with an average concentration of 0.0002 mg/L. • Mercury is toxic to aquatic organisms and is biomagnified within food chains. • The CEQG for mercury for the protection of freshwater aquatic life is as follows: inorganic mercury = 0.026 µg/L; methylmercury = 0.004 µg/L; inorganic mercury for marine environments = 0.016 µg/L.
Molybdenum	• AQUAMIN Working Groups 7 and 8 (1996) recommended that molybdenum be measured in effluent characterization.

	• Molybdenum occurs primarily in uranium ores, but may also occur in base metal ores and a small number of gold ores, and is not expected to occur across Canada. • Molybdenum can be toxic to aquatic organisms, but its toxicity is not well understood. • Molybdenum can also affect drinking water quality and cause molybdenosis in cattle. • The CEQG for molybdenum for the protection of freshwater aquatic life is 0.073 mg/L.
Hardness	• Water hardness is a measure of cations, predominantly divalent cations, dissolved in water. • Calcium and magnesium are the major contributors to hardness, and hardness can be calculated based on concentrations of these ions. • Hardness is an important factor affecting the fate, bioavailability and toxicity of metals.
Selenium	• AQUAMIN Working Groups 7 and 8 (1996) did not consider selenium for inclusion in effluent characterization. • The MISA draft development document (Ontario Ministry of the Environment and Energy 1993) stated that selenium was found in 10% of the metal mine effluents sampled, with an average concentration of 0.007 mg/L. • Selenium occurs most commonly in association with sulphur, and is not common in mine effluents. • Selenium is toxic to aquatic organisms. • The CEQG for selenium for the protection of freshwater aquatic life is 0.001 mg/L.
Electrical conductivity	• Conductivity is a measure of water's ability to conduct an electrical current. • Conductivity can be measured in the field or in the lab. • Conductivity gives an approximate measure of total dissolved solids, and can be used to identify the location of an effluent plume in freshwater environments.
Required Parameters for Water Quality Monitoring Only	
Dissolved oxygen	• Dissolved oxygen can be measured in the field or in the lab. • Dissolved oxygen is a factor affecting the fate and bioavailability of metals.
Temperature	• Temperature changes can affect limnological properties of lakes, and can also affect aquatic organisms. • The CEQG for temperature for the protection of freshwater aquatic life states that thermal additions should not alter thermal stratification or turnover dates, exceed maximum weekly temperature averages, or exceed maximum short-term temperatures.
Additional Site-specific Recommended Parameters for Effluent Characterization and Water Quality Monitoring	
Calcium	• AQUAMIN Working Groups 7 and 8 (1996) recommended that calcium be measured in effluent characterization. • Calcium is an important cation in aquatic ecosystems, and may also occur in mine effluent as a result of acid neutralization using lime. • Calcium discharges from mines may have effects on fish habitat. • Calcium concentrations are essential to calculating hardness. • Calcium is known to affect the toxicity of metals and/or other mine effluent parameters;. • Calcium data may assist in the interpretation of the potential impacts of metals and other parameters.
Chloride	• The MISA draft development document (Ontario Ministry of the Environment and Energy 1993) stated that chloride was found in 80% of the metal mine effluents sampled, with an average concentration of 83 mg/L. • Chloride can affect the toxicity of metals and/or other mine effluent parameters. • Chloride data may assist in the interpretation of the potential impacts of metals and other parameters.
Fluoride	• AQUAMIN Working Groups 7 and 8 (1996) recommended that fluoride be measured in effluent characterization for operations where it is likely to be present. • Fluoride occurs in a limited number of ore types, and is not expected to occur across Canada. • Fluoride has been shown to bioaccumulate in fish bones, but its effects on aquatic organisms are not well understood. • Fluoride is lethal to fish at concentrations ranging from 10 to 200 mg/L. • Fluoride was rated "toxic" following Priority Substance List 1 assessment for "having an immediate or long-term harmful effect on the environment."

Magnesium	• Magnesium is an important cation in aquatic ecosystems, and magnesium concentrations are essential to calculating hardness. • Magnesium is known to affect the toxicity of metals and/or other mine effluent parameters. • Magnesium data may assist in the interpretation of the potential impacts of metals and other parameters.
Manganese	• AQUAMIN Working Groups 7 and 8 (1996) did not consider manganese for inclusion in effluent characterization. • Manganese occurs in many ore types, and is expected to occur at mines across Canada. Manganese makes up 0.1% of the Earth's crust. • There is no CEQG for manganese for the protection of freshwater aquatic life. • Manganese can have an important influence on the fate of other contaminants, specifically on the availability of other metals. • Manganese is toxic to aquatic life but the factors that affect its toxicity are not well understood.
Potassium	• Potassium is known to affect the toxicity of metals and/or other mine effluent parameters. • Potassium data may assist in the interpretation of the potential impacts of metals and other parameters.
Sodium	• Sodium is known to affect the toxicity of metals and/or other mine effluent parameters. • Sodium data may assist in the interpretation of the potential impacts of metals and other parameters.
Sulphate	• Sulphate is an important anion in water. • Mine effluents from mines with sulphide-bearing ore can be important sources of sulphate. • Sulphate data may assist in the interpretation of the potential impacts of metals and other parameters. • The MISA draft development document stated that sulphate was found in 86% of the metal mine effluents sampled, with an average concentration of 644 mg/L.
Total phosphorus	• AQUAMIN Working Groups 7 and 8 (1996) did not consider total phosphorus for inclusion in effluent characterization. • The MISA draft development document (Ontario Ministry of the Environment and Energy 1993) stated that total phosphorus was found in 24% of the metal mine effluents sampled, with an average concentration of 0.1 mg/L. • Total phosphorus may occur in mine effluent in particulate or dissolved form. • Total phosphorus is a nutrient, and may lead to excessive plant growth.
Uranium	• AQUAMIN Working Groups 7 and 8 (1996) did not consider uranium for inclusion in effluent characterization. • Small amounts of uranium occur in many rock types in Canada, but uranium is expected to occur primarily in effluents from uranium mines. • The MISA draft development document (Ontario Ministry of the Environment and Energy 1993) stated that uranium was found in 22% of the metal mine effluents sampled, with an average concentration of 0.06 mg/L. • Uranium has been shown to bioaccumulate in fish.
Parameters Recommended for Effluent Characterization Only	
Total thiosalts	• AQUAMIN Working Groups 7 and 8 (1996) recommended that thiosalts be measured in effluent characterization (listed in Table 4.1 of the AQUAMIN report, but not explicit in text) and also recommended that thiosalt impacts be monitored as part of EEM. • Thiosalts are a group of metastable oxysulphur anions formed by partial oxidation of sulphide minerals during processing. • Thiosalts have relatively low toxicity, but in the aquatic environment the oxidation of thiosalts can lead to significant reductions in pH. • Reductions in fish and benthic communities have been associated with thiosalts. • Thiosalts have the potential to occur at any mine that uses flotation to separate sulphide minerals, but only a few mines in Canada have known problems associated with thiosalts.
Site-specific Parameters Recommended for Water Quality Monitoring Only	
Dissolved and total organic carbon	• Dissolved and total organic carbon are important factors affecting the fate and bioavailability of metals.

Salinity	• Salinity is an important parameter in marine environments, and may be a contaminant at some uranium mines.
Optical depth or Transparency	• Optical depth is a field measurement of transmission of light through water as affected by water colour (dissolved constituents) and turbidity (particulate constituents). • Optical depth is measured in the field using a Secchi disk or a turbidimeter. • A low level of light transmission can reduce primary productivity in water and reduce the ability of predators to find prey.
Water depth	• Water depth can affect temperature, dissolved oxygen, and the degree of effluent dilution, all of which are modifying factors in effluent toxicity.
Thallium	• Metal mining can be a source of thallium to aquatic environments. • The CEQG for the protection of aquatic life is 0.0008 mg/L.

Table of Contents

6. Sublethal Toxicity Testing.....	6-1
6.1 Overview.....	6-1
6.2 Collection of Samples	6-2
6.3 Sampling Locations	6-3
6.4 Frequency and Reporting.....	6-3
6.5 Tabulation of Sublethal Toxicity Endpoints and Validation of Test Results ..	6-4
6.5.1 Validation of Test Results.....	6-5
6.5.2 Tabulation of Sublethal Toxicity Endpoints.....	6-5
6.6 Data Interpretation in Relation to the Toxicity Objectives	6-6
6.6.1 Changes in Effluent Quality.....	6-6
6.6.2 Understanding Multiple Discharge Situations.....	6-6
6.6.3 Contributions to the Weight-of-Evidence Approach	6-6
6.6.4 Considerations for Integration of Toxicity Test Results.....	6-7
6.7 Description of Freshwater and Marine Sublethal Toxicity Tests	6-7
6.8 Dilution Water in Freshwater Sublethal Toxicity Testing.....	6-12
6.8.1 Dilution Water Selection.....	6-12
6.9 Collection, Shipment and Storage of Samples for Sublethal Toxicity Testing.....	6-17
6.9.1 Test Organism Acclimation	6-17
6.9.2 Screening Tests	6-20
6.9.2.1 Screening Test Impairment.....	6-20
6.9.3 Effluent or Effluent-Exposed Surface Water Toxicity Tests.....	6-21
6.10 Use of Sublethal Toxicity Testing in Resolving Confounding Influences	6-21
6.11 References.....	6-24

List of Tables

Table 6-1: Methodologies for effluent sublethal toxicity tests	6-2
Table 6-2: Descriptions of the freshwater and marine sublethal toxicity tests included in the metal mining EEM program	6-9
Table 6-3: Specifications of the Environment Canada test methods and recommendations for collection, storage and use of site-collected dilution waters.....	6-15
Table 6-4: Dilution/control water and corresponding effluent volumes for sublethal toxicity tests	6-17
Table 6-5: Preparation of different hardness/alkalinities.....	6-19

6. Sublethal Toxicity Testing

6.1 Overview

There are two main uses for sublethal toxicity tests in the environmental effects monitoring (EEM) program: to compare processes and measure changes in effluent quality; and to contribute to the understanding of the relative contributions of the mine in multiple-discharge situations.

The purpose of sublethal toxicity testing in the metal mining EEM program is to provide an estimate of the potential effects on biological components (phytoplankton, zooplankton, benthic invertebrates, fish, macrophytes) in the exposure area, whether or not these components are being directly measured in the field.

To estimate the potential effects on biological components, sublethal toxicity testing shall be conducted by following the applicable methods referred to in the *Metal Mining Effluent Regulations* (MMER), section (s.) 5, subsections (ss.) 3 and 4 (as amended from time to time). Four freshwater sublethal toxicity tests (fish, invertebrate, algal and plant species) or three marine or estuarine sublethal toxicity tests (fish, invertebrate and algal species) shall be conducted, depending on the receiving environment type (MMER, Schedule 5, s. 5), and the results shall be recorded. The test chosen should primarily be based on the relevance of the species to the local receiving environment, and secondarily on the seasonal availability of test organisms.

Acceptable sublethal toxicity methods are outlined in Table 6-1. The following website contains all the biological test method documents published by Environment Canada's Biological Methods Section: <http://www.ec.gc.ca/faunescience-wildlifescience/default.asp?lang=En&n=0BB80E7B-1>. Test report checklists have been developed for assessing the validity of test results for each test option, which are available on the EEM website (www.ec.gc.ca/esee-em/default.asp?lang=En&n=D450E00E-1). Information on the relative sensitivity of the different sublethal toxicity tests can be found in ESG (1999). For additional information on assessing changes in effluent over time, and other data-interpretation situations, see sections 6.6 and 6.10.

Table 6-1: Methodologies for effluent sublethal toxicity tests

Test Description	Receiving Environment	Test Species	Methods
Fish early life stage development tests	Marine	Inland Silverside (<i>Menidia beryllina</i>) or Topsmelt (<i>Atherinops affinis</i>)	US EPA (2002)
	Freshwater	Fathead Minnow (<i>Pimephales promelas</i>) ¹ or Rainbow Trout (<i>Oncorhynchus mykiss</i>)	Environment Canada (1992a) or Environment Canada (1998)
Invertebrate reproduction tests	Marine	Echinoids (sea urchins or sand dollars)	Environment Canada (1992b)
	Freshwater	Water Flea (<i>Ceriodaphnia dubia</i>)	Environment Canada (2007a)
Plant and algae toxicity tests	Marine - algae	Barrel Weed (<i>Champia parvula</i>)	US EPA (2002)
	Freshwater - algae	Green Algae (<i>Pseudokirchneriella subcapitata</i>)	Environment Canada (2007b) or MDDEP (2007) ²
	Freshwater - plant	Lesser Duckweed or Common Duckweed (<i>Lemna minor</i>)	Environment Canada (2007c)

1. Where Fathead Minnows are not an indigenous species, Rainbow Trout will be used according to Environment Canada (1998).

2. In some jurisdictions, both Environment Canada (2007b) and the MDDEP (2007) testing requirements for *Pseudokirchneriella subcapitata* are acceptable for the EEM program.

Note: For all marine toxicity test procedures, it is recommended that the effluent salinity adjustment procedure by Environment Canada (2001) be followed. For all sublethal tests where the test organisms are purchased for sublethal toxicity testing, it is recommended that the test organism importation of Environment Canada (1999) be followed.

6.2 Collection of Samples

Sublethal toxicity testing shall be conducted on the aliquots of effluent samples collected in accordance with ss. 4(2) of the MMER (Effluent Characterization) (MMER, Schedule 5, ss. 5(2)).

In choosing when effluent for toxicity tests should be collected, two aspects should be considered:

- 1) when effluent poses the greatest potential for adverse environmental impact on the environment; and
- 2) when biological monitoring is conducted to look at potential linkages with effects in the exposure area.

6.3 Sampling Locations

To determine which outfall structure has potentially the most adverse environmental impact, the following should be taken into account:

- the mass loading of deleterious substances;
- the manner in which the effluent mixes in the exposure area; and
- historical characterization or sublethal toxicity data.

In cases where it is not clear which discharge source has the greatest potential to affect the environment, mines may wish to use a series of single-concentration sublethal toxicity tests from each final discharge location to determine the source with the greatest sublethal response.

To estimate the potency of the response from each discharge source, the “time to response” can be observed and calculated as the sublethal toxicity test endpoint (e.g., *Ceriodaphnia dubia* adults are exposed to undiluted effluent samples from each different effluent discharge, and observations are made as to how long it takes to find a 25% or 50% response). The sublethal toxicity test endpoint would be an LT₂₅ (time to 25% mortality) or LT₅₀ (time to 50% mortality) if survival was the key observation. The single-concentration test would be the more cost-effective approach to screening effluent sources in order to determine the discharge point with the greatest potential to affect the receiving environment.

6.4 Frequency and Reporting

Sublethal toxicity tests shall be conducted two times each calendar year for the first three years and once each year after the third year (MMER, Schedule 5, ss. 6(2)). The first effluent sample shall be collected, and sublethal toxicity testing conducted, not later than six months after the mine is subject to section 7 of the MMER (Schedule 5, ss. 6(1)).

The results of sublethal toxicity testing shall be submitted to the Authorization Officer as part of the Effluent and Water Quality Monitoring Report. Additional information on the content of the Effluent and Water Quality Monitoring report is discussed in Chapter 5. The reporting requirements of the sublethal toxicity results are described in the MMER (s. 23; and Schedule 5, s. 8). See Chapter 10 for information on electronic reporting of sublethal toxicity data.

The test methods in Table 6-1 can be referred to for reporting specifications for each test method.

The report should include the following:

- dates when the samples were collected for sublethal toxicity testing;
- the location of the final effluent discharge point from which samples were collected for sublethal toxicity testing, and data on how this point was chosen;

- the results of sublethal toxicity testing, including the median lethal concentration (LC_{50}), 25% inhibition concentration (IC_{25} ¹) and 25% effect concentration (EC_{25}) where applicable, 95% confidence limits, and indication of quantitative statistics employed;
- a description of the quality assurance / quality control (QA/QC) measures that were implemented, and the data related to the implementation of those measures; and
- minimum reporting outlined in the test methods and sublethal toxicity checklists.

In determining whether or not to use historical sublethal toxicity data as part of the EEM program, the mine should take the following factors into consideration:

- laboratory QA using the methods listed in Table 6-1;
- no fewer than three species tested (adequate to answer the question);
- age of the data (testing conducted after December 31, 1997);
- the nature of mine operating conditions (e.g., are mine operations similar to the operations in place at the time the sublethal toxicity tests were conducted?); and
- whether any of the sublethal toxicity testing was conducted on an effluent sample taken concurrent with fish and benthic invertebrate field monitoring.

Toxicity data submitted as part of the EEM program for the metal mining industry should be accompanied by a description of the materials and methods, and calculations for each test. Minimum reporting requirements are detailed in section 8 or 9 of the toxicity test method documents of Environment Canada. Test report checklists have been developed for assessing the validity of test results for each test option, which are available on the EEM website (www.ec.gc.ca/esee-eem/default.asp?lang=En&n=D450E00E-1). The minimum reporting requirements for methods of the U.S. Environmental Protection Agency (EPA) (US EPA 2002) have been prescribed for the purpose of subsequent EEM phases. The U.S. EPA requirements generally conform to Environment Canada specifications.

6.5 Tabulation of Sublethal Toxicity Endpoints and Validation of Test Results

Sublethal toxicity endpoints reported vary depending on the test being conducted (refer to test methods in Table 6-1). However, the IC_{25} will be discussed below for illustrative purposes. The geometric mean² of all IC_{25} s (GM- IC_{25}) for a given species should be calculated for each phase.

¹ IC_{25} is defined as the effluent concentration where a 25% inhibition is observed in the exposed test organisms.

² The geometric mean may be calculated as the n^{th} root of n numbers multiplied together. Alternatively, the logarithms of the n IC_{25} s (EC_{25} s or LC_{50} s) may be added together, the sum divided by n , and the antilog of the result is the geometric mean.

6.5.1 Validation of Test Results

Procedures for QA/QC will be followed by both the field crews collecting environmental samples and the laboratory carrying out the toxicity testing, as discussed in the required toxicity test method documents.

Therefore, a first step in the interpretation of toxicity data for EEM should be the resolution of any problems with QA/QC. In addition to the QA outlined in the individual sublethal toxicity test methods, further requirements and recommendations are as follows:

- reference toxicant test conducted in the same manner as the effluent or effluent-exposed surface water test;
- reference toxicant test conducted within ~ 30 days of the effluent or effluent-exposed surface water test;
- test-specific validity criteria met in all effluent sublethal testing conducted;
- sublethal toxicity testing initiated within 3 days of sample collection;
- quantitative sublethal toxicity endpoints provided for all sublethal toxicity tests conducted on effluent or effluent-exposed samples;
- sublethal toxicity test endpoint between 0.1 and 100% bracketed by at least one test concentration;
- sublethal toxicity tests that fail to meet test method validity criteria repeated on a new sample; and
- reporting of “less than” values as a sublethal toxicity test endpoint will no longer be acceptable.

Data could be declared rejected if one or more essential elements of the test method were not followed (e.g., failure to meet organism health criteria, inappropriate manipulations of the sample, failure to conduct the minimum in-test monitoring, incorrect statistic used for sublethal toxicity endpoint calculation).

Laboratories contracted by the metal mining industry to conduct sublethal toxicity testing should be accredited under the International Organization for Standardization standard ISO/IEC 17025:2005 entitled “General requirements for the competence of testing and calibration laboratories,” as amended from time to time.

6.5.2 Tabulation of Sublethal Toxicity Endpoints

If the effluent does not cause a 25% sublethal inhibition or effect for any of the freshwater sublethal tests, an IC_{25}/EC_{25} cannot be calculated and it is reported as $> 100\%$.

Mortality in some of the concentrations might also prevent calculation of the IC_{25}/EC_{25} . For example, there might be no measured effects (mortality, growth or reproduction) in 32% concentration, but appreciable mortality in 56%. It would then be impossible to obtain a good estimate of the inhibition of growth or reproduction for the 56% concentration, and hence impossible to determine IC_{25}/EC_{25} . In such a situation, the

IC₂₅/EC₂₅ should be assumed to be equal to the higher concentration, which in this case is 56%.

There should be IC₂₅s reported for each of the test species. This information is summarized by calculating the geometric mean for each set of IC₂₅s. For example, testing of a mine effluent on six occasions resulting in measured IC₂₅ values of 10, 15, 17, 23, 25 and 30% would lead to a geometric mean of 19%.

6.6 Data Interpretation in Relation to the Toxicity Objectives

6.6.1 Changes in Effluent Quality

The geometric mean of the IC₂₅ (GM-IC₂₅) for each species can be compared between phases to assess changes in the quality of effluent over time at each mine or between mines and mine types. Improvements are expected as a result of changes in process or effluent treatment.

6.6.2 Understanding Multiple Discharge Situations

Comparison of mine effluent toxicity data from other nearby industrial and/or municipal discharges can help in understanding the relative contribution to the potential impact of each effluent source on the environment. Provided that relevant toxicity, flow and dispersion information is available for the other discharges, the inter-relation or overlap of the effluent fields may be better understood. Another method of comparing relative loading (toxic contribution) from each source is to calculate the TER (toxicity emission rate; = (100/GM-IC₂₅) x flow). However, this calculation does not relate to the receiving environment, because the effluent dispersion and dilution are not taken into consideration. See section 6.10 for more information on confounding influences.

6.6.3 Contributions to the Weight-of-Evidence Approach

Where sublethal toxicity data have an IC₂₅ of less than 30%, it is recommended that mines calculate the geographic extent of the response in the exposure area and identify the zone where the concentration of effluent is comparable to the IC₂₅. Data on effluent and receiving flow for the appropriate month are needed to complete this estimate. The estimation of the potential geographic extent can be effectively reported in map form and reported in the Interpretative Report.

A potential-effects zone may be interpreted as a rough indication of the extent of 25% inhibition by effluent in the environment. If, for example, a mine's GM-IC₂₅ for a particular test was 1% volume/volume (v/v), this would match the extent of the 1% zone. Invertebrates and plant IC₂₅s are not expected to be similar, due to differing species sensitivities and test method sublethal toxicity endpoints, and therefore may have dissimilar potential-effects zones.

6.6.4 Considerations for Integration of Toxicity Test Results

The following points should be considered when toxicity data are used to estimate a potential-effects zone.

(1) *Laboratory results*: Sublethal laboratory tests provide estimates of toxicity under strictly controlled laboratory conditions for each test species. These conditions do not replicate environmental conditions at the site under study. Chapman (2000) describes various abiotic and biotic modifying factors present in the uncontrolled receiving environment, which may affect an organism's response to a toxicant.

(2) *Species differences*: Species differences in sensitivity to metal mining effluents will be taken into account when extrapolating results from laboratory sublethal toxicity tests to effects on indigenous biota.

(3) *Background toxicity*: The description above assumed that there were no other upstream contributions of toxicity. That assumption would be erroneous if there were overlapping plumes.

(4) *Type of receiving water*: Receiving-water pH, hardness, dissolved organic carbon (DOC) and other modifying factors could potentially increase or decrease toxicity of the effluent compared to tests with laboratory water.

(5) *Plume uncertainties*: Calculations of dilution might be difficult or inaccurate, or the position of the mixing zone might be variable. Where this uncertainty exists, estimating a zone of potential effect would have an equal level of uncertainty.

6.7 Description of Freshwater and Marine Sublethal Toxicity Tests

Table 6-2 provides short descriptions for the freshwater and marine sublethal toxicity tests included in the metal mining EEM program. Information on the relative sensitivity of the different sublethal toxicity tests can be found in ESG (1999).

For freshwater tests, laboratory or site water can be used as dilution/control water. For marine or estuarine environments, the mine has a choice of using uncontaminated sea water or artificial sea water produced from hyper-saline brine (HSB). The recommended procedures for adjusting the salinity of the effluent and dilution water and preparing the HSB are described in Environment Canada (2001).

Where applicable, the test organism importation methodology (Environment Canada 1999) should be referred to, where test organisms are purchased for immediate use in sublethal toxicity tests.

For QA/QC, detailed records of all aspects of the samples, test organisms, culture maintenance, test conditions, equipment and test results are validated and kept by the

laboratory. A reference toxicant test is used to establish the validity of effluent toxicity data. Successive reference toxicant data are plotted on a control chart. If results are within expected limits, the performance of the batch of test organisms is ensured. The minimum level of reporting is outlined in each test method.

Technical personnel should be skilled in algae, macrophyte, invertebrate and fish culture, and in conducting toxicity tests following aseptic techniques.

For more detailed descriptions, please refer to the specific test method documents.

Table 6-2: Descriptions of the freshwater and marine sublethal toxicity tests included in the metal mining EEM program

Test	Purpose and Results	Description	Biological Test Method and Cost
Larval growth and survival assay using Inland Silverside (<i>Menidia beryllina</i>)	Evaluate effects of effluent exposure on fish larvae. Result is expressed as the concentration at which larval growth is reduced by 25% (IC ₂₅). If mortality is significant, it may be possible to calculate the lethal concentration for 50% of the test population (statistical endpoint is a LC ₅₀).	The Inland Silverside is a small fish that populates a variety of habitats and tolerates a wide range of temperature (2.9 to 32.5°C) and salinity (0 to 58 g/kg). The Inland Silverside is a multiple spawner, and spawning can be induced by diurnal interruption in the circulation of water in the culture tanks. The eggs adhere to vegetation in the wild or to filter floss in laboratory culture tanks. The larvae hatch in 6 to 7 days when incubated at 25°C and maintained in seawater with salinity ranging from 5 to 30 g/kg. Seven- to 11-day-old larvae are exposed to a minimum of 5 concentrations of the effluent sample and a control for 7 days in a static renewal system at 25°C. During the 7 days, the larvae are fed brine shrimp once or twice per day and the solutions are replaced once each day. At the end of the 7-day exposure period, the surviving larvae are counted and individual replicate weights are measured to calculate the growth changes, which are compared statistically between exposure concentrations and the controls. For a valid test, average dry weight of control larvae will be ≥ 0.50 mg and control survival will be ≥ 80%. The test requires approximately 40 L of effluent.	Short Term Methods for Estimating Chronic Toxicity of Effluent and Receiving Waters to Marine and Estuarine Organisms (3rd Edition) (Reference Method EPA-821-R-02-014), October 2002, published by the U.S. EPA.* For the U.S. EPA method, the minimum reporting outlined in Environment Canada test methodologies should be followed.
Larval growth and survival assay using Topsmelt (<i>Atherinops affinis</i>)	Evaluate effects of effluent exposure on fish larvae. Result is expressed as the concentration at which larval growth is reduced by 25% (IC ₂₅). If mortality is significant, it may be possible to calculate the lethal concentration for 50% of the test population (statistical endpoint is a LC ₅₀).	Topsmelt are small fish which occur from the Gulf of California to Vancouver Island. Topsmelt reproduce from May through August, depositing eggs on benthic algae in the upper ends of estuaries and bays. Off-season spawning has been successful in a laboratory-held population. Spawning is induced by a combination of three environmental cues: lighting, tidal cycle and temperature. Nine- to 15-day-old larvae are exposed to a minimum of 5 concentrations of the effluent sample and a control for 7 days in a static renewal system at 20°C. During the seven days, the larvae are fed brine shrimp twice per day and the solutions are replaced once each day. At the end of the seven-day exposure period, the surviving larvae are counted and individual weights are measured to calculate the growth changes, which are compared statistically between exposure concentrations and the controls. For a valid test, average dry weight of control larvae will be ≥ 0.85 mg and control survival will be ≥ 80%. The test requires approximately 40 L of effluent.	Short Term Methods for Estimating Chronic Toxicity of Effluent and Receiving Waters to West Coast Marine and Estuarine Organisms (1st Edition) (Reference Method EPA/600/R-95-136), August 1995, published by the U.S. EPA. * For the U.S. EPA method, the minimum reporting outlined in Environment Canada test methodologies should be followed.
Fathead Minnow (<i>Pimephales promelas</i>) Growth and survival test	Evaluate effects of effluent exposure to an early life stage of fish. Results are expressed as the concentration at which larval growth/survival is reduced by 25% (IC ₂₅). If mortality is significant, it may be possible to calculate the lethal concentration for 50%	Fathead Minnows are small, warm-water fish found across North America in ponds and slow-moving water areas of rivers. The female lays her eggs on the underside of hard surfaces, where the male cares for the eggs until hatching. Fathead Minnow larvae, less than 24 hours old, are exposed to a minimum of 7 concentrations of the effluent sample and a control for seven days at 25°C. During the seven days, the larvae are fed brine shrimp 2 or 3 times per day and the test/control solutions are replaced once each day. At the end of the seven-day exposure period, the surviving larvae are counted and individual replicate weights are measured to calculate the growth changes, which are compared statistically between exposure concentrations	Test of Larval Growth and Survival Using the Fathead Minnows (Reference Method EPS 1/RM/22), February 1992, Amended in September 2008, published by Environment Canada.

Test	Purpose and Results	Description	Biological Test Method and Cost
	of the test population (statistical endpoint is LC ₅₀).	and the controls. The test requires approximately 40 L of effluent.	
Rainbow Trout (<i>Oncorhynchus mykiss</i>) Embryo development Test	Evaluate effects of effluent exposure to an early life stage of fish. Results are expressed as the concentration at which embryo viability is reduced by 25% (statistical endpoint is an EC ₂₅).	Rainbow Trout are common in clean, cold-water streams in North America. In some areas, they are not a native species but have been introduced to the watershed. Rainbow Trout are cultured throughout the country by commercial hatcheries. Adult Rainbow Trout migrate to shallow water to spawn in clean gravel. They bury their eggs in the rocks and gravel, where the young live until the yolk sac is absorbed. The embryo development test involves exposing recently fertilized Rainbow Trout eggs to a series of concentrations of the effluent sample for seven days at 14°C. Test exposure solutions are renewed every day. Dead embryos are counted and removed during the test. At the end of the test, the embryos' viability is assessed and numbers of healthy embryos are counted for statistical comparison between test concentrations and the control. The test requires approximately 80–90 L of effluent.	Toxicity Tests Using Early Life Stages of Salmonid Fish (Rainbow Trout) (Reference Method EPS 1/RM/28), July 1998, published by Environment Canada.
Fertilization assay using echinoids (sea urchins and sand dollars)	Evaluate effects of effluent exposure on egg fertilization success of echinoids. Results are expressed as the concentration at which the fertilized egg number is reduced by 25% (statistical sublethal toxicity endpoint is the IC ₂₅).	Echinoids are considered to be structurally advanced and complex invertebrates. Seven species of sea urchins and three species of sand dollars are commonly found in the coastal marine waters of Canada. Mature and gravid male and female echinoids are stimulated to spawn by injecting potassium chloride. Semen from at least 3 males is pooled and numbers are adjusted to the desired sperm:egg ratio. Eggs from at least 3 females are pooled and numbers are adjusted to 2000 eggs/millilitre (ml). Sperm is exposed for 10, 20 or 60 minutes (depending on the test option chosen) to a series of concentrations of the effluent sample. Eggs are then added to the test vessels for a 10- or 20-minute additional exposure. Adding formalin terminates the test. Preserved eggs are counted (in the range of 100 to 200 eggs) and classified as either fertilized or not fertilized, under a microscope at 100x magnification. For a valid test, the fertilization rate in the controls will be ≥ 50%, but < 100% and a positive and logical dose-effect curve should be obtained. The test requires approximately 1 litre (L) of effluent.	Fertilization Assay using Echinoids (Sea Urchins and Sand Dollars) (Reference Method EPS 1/RM/27), December 1992, amended in November 1997, published by Environment Canada.
<i>Ceriodaphnia dubia</i> Reproduction and survival test	Evaluate effects of effluent exposure on the reproduction of an invertebrate. Results are expressed as the concentration at which the average number of young per female is reduced by 25% (IC ₂₅). If mortality is significant, it may be possible to calculate the lethal concentration for 50% of the test population (statistical sublethal toxicity endpoint is	<i>Ceriodaphnia</i> is a species of zooplankton abundant in lakes, ponds and quiescent sections of streams and rivers throughout North America. In the test, <i>Ceriodaphnia</i> are separated so that there is 1 female adult animal per test vessel and 10 replicates per concentration. Young ceriodaphnids, less than 24 hours old, are exposed to a minimum of 7 effluent concentrations and a control, at 25°C. The test is completed when at least 60% of the surviving control organisms have had 3 broods of neonates or at the end of 8 days, whichever occurs first. During each day of the test, adult survivorship is assessed, all young produced are removed and counted, and the test solutions are renewed. At the end of the test, the number of surviving adults and the number of young produced per adult in 3 broods are compared statistically between exposure concentrations and the controls. The test requires 3–4 L of effluent.	Test of Reproduction and Survival using the Cladoceran <i>Ceriodaphnia dubia</i> (Reference Method EPS 1/RM/21), 2nd edition, February 2007, published by Environment Canada.

Test	Purpose and Results	Description	Biological Test Method and Cost
	an LC ₅₀).		
Sexual reproduction assay using the red macroalga <i>Champia parvula</i>	Evaluate effects of effluent exposure on the sexual reproduction of a marine red macroalga. Result is expressed as the concentration at which the number of cystocarps is reduced by 25% (statistical sublethal toxicity endpoint is IC ₂₅).	Mature plant body of <i>Champia parvula</i> is hollow, septate and highly branched. New cultures can be propagated asexually from excised branches, making it possible to maintain clonal material indefinitely. Two sexually mature male and 5 female branches of <i>Champia parvula</i> are exposed in a static system for 2 days to a series of concentrations of the effluent sample, followed by a 5-7-day recovery period in control medium. The recovery period allows time for the development of cystocarps on the female branches resulting from fertilization during the exposure period. For a valid test, the female control mortality must be <20%, and the average number of cystocarps per female control plants is ≥ 10. The test requires approximately 2 L of effluent.	Short Term Methods for Estimating Chronic Toxicity of Effluent and Receiving Waters to Marine and Estuarine Organisms (3rd Edition) (Reference Method EPA-821-R-02-014), October 2002, published by the U.S. EPA. For the U.S. EPA method, the minimum reporting outlined in Environment Canada test methodologies should be followed.
Algal growth inhibition test using <i>Pseudokirchneriella subcapitata</i>	Evaluate effects of effluent exposure on the growth of a unicellular freshwater alga. Result is expressed as the concentration at which the number of cells is reduced by 25% (statistical sublethal toxicity endpoint is an IC ₂₅).	<i>Pseudokirchneriella subcapitata</i> is a non-motile, unicellular, crescent-shaped (40-60 micrometres ³ [µm ³]) green alga found in most freshwaters in North America. Its uniform shape makes it ideal for enumeration with an electronic particle counter. Clumping seldom occurs, because <i>Pseudokirchneriella</i> is free of complex structures and does not form chains. Growth is sufficiently rapid to accurately count cell numbers after 72 hours. Axenic (i.e., aseptically prepared stock cultures containing only the test species), exponentially growing <i>Pseudokirchneriella</i> are exposed to the test solutions in a static, 96-well microplate. The algae are exposed to a dilution series of filtered effluent sample over several generations under constant temperature (24°C), with continuous light for 72 hours. The number of algal cells in the test concentrations is compared with the number in the control solutions. An effluent is considered toxic when a statistically significant, dose-dependent inhibition of algal growth occurs. The test requires < 1 L of effluent.	Growth Inhibition Test using a Freshwater Algae (Reference Method EPS 1/RM/25), 2nd edition, March 2007, published by Environment Canada. Or: Détermination de la toxicité : inhibition de la croissance chez l'algue <i>Pseudokirchneriella subcapitata</i> (Reference Method MA. 500-P.sub 1.0, Rév. 1, 2007), September 1997, published by the Centre d'expertise en analyse environnementale du Québec, Ministère du Développement durable, de l'Environnement et des Parcs du Québec.
Macrophyte growth inhibition test using <i>Lemna minor</i>	Evaluate effects of effluent exposure on the growth of a freshwater plant. Results are expressed as the concentration at which frond number and frond dry weight is reduced by 25% (statistical endpoint is an IC ₂₅).	<i>Lemna minor</i> (Lesser Duckweed or Common Duckweed) is a small vascular, macrophyte plant, found at or just below the surface in freshwater (ponds, lakes, stagnant waters and quiet areas of streams and rivers). It is a common macrophyte with nearly worldwide distribution from tropical to temperate zones and grows in most regions of Canada. Its growth is rapid and occurs by lateral branching. Seven- to 10-day-old, rapidly growing plants (typical size of frond is 1 cm) are exposed to a series of concentrations of the effluent sample diluted with growth medium for 7 days. During the test, the plants are incubated at 25°C under continuous light and static conditions. Plants are acclimated to the test media for 18–24 hours before testing. The leaves are counted and weighed at the end of the test and growth is compared statistically to the controls. For a valid test, the number of leaves on control plants will increase by 8-fold at the end of the test. The test requires approximately 1–2 L of effluent.	Test for Measuring the Inhibition of Growth using the Freshwater Macrophyte, <i>Lemna minor</i> (Reference Method EPS 1/RM/37), 2nd edition, January 2007, published by Environment Canada.

6.8 Dilution Water in Freshwater Sublethal Toxicity Testing

6.8.1 Dilution Water Selection

The sublethal toxicity test methods required for the metal mining EEM program clearly define the culture conditions and test procedures that need to be followed (Environment Canada 1992a, 1992b, 1998, 2007a, 2007b, 2007c; US EPA 1994a, 1994b, 1995, 2002). Some testing decisions are left to the discretion of the individual laboratories, as long as the standard test acceptability criteria can be achieved. For example, standard methods for testing *Ceriodaphnia*, Fathead Minnows, *Pseudokirchneriella* and Rainbow Trout allow the use of uncontaminated ground or surface water, dechlorinated tap water or reconstituted water as a source for the culture or the test control/dilution water, as long as the water of choice supports a healthy culture and provides a valid test result.

Most laboratories in Canada use “standard laboratory” water for routine culturing and testing requirements. This water is generally supplied to the laboratory through a natural groundwater system (well) or a local municipal water source, which must be dechlorinated and may be buffered to meet acceptable culturing criteria. Deionized water reconstituted to targeted water quality parameters is also used. Advantages of using laboratory water include the following:

- It can be maintained at a consistent quality with minimal risk of contamination by undesirable and/or harmful chemicals or biota.
- Regular monitoring of water chemistry and culture health, as well as reference toxicant testing, ensure that the water is of acceptable quality for toxicity testing.
- Since cultures are maintained in laboratory water, no additional acclimation is needed for testing effluents or chemicals when laboratory water is used as the control/dilution water.
- Laboratory water, normally used in regulatory testing across Canada, provides a measure of the inherent toxicity of the effluents and allows comparison of effluent quality over time.

During the metal mining EEM program, most sublethal toxicity tests will likely be performed using laboratory water as control/dilution water, in order to attain comparable results among different laboratories and over time. It is also likely that in many sublethal toxicity tests where there is measurable effluent toxicity, this toxicity can be attributed to inorganic substances such as metals and ammonia, and the toxicity of the effluent may also be affected by site-specific characteristics such as pH, alkalinity and hardness. These characteristics can be controlled and reproduced in the laboratory for cases in which test results, reflective of the site conditions, are desired. However, a mine may decide to test its effluent using unexposed surface water (as control and dilution water), providing the sample is not exposed to effluent. Alternatively, a reference area of similar physicochemical characteristics to a mine site could be used to supply control/dilution water.

The use of unexposed surface waters can be especially helpful in obtaining the following information.

Estimating the mitigating or stimulatory effect of unexposed surface site water as dilution water on the expression of toxicity from the effluent discharge or effluent-exposed surface water

Although parallel testing of effluents and effluent-exposed surface waters using site water and hardness-adjusted laboratory water have produced similar results (BEAK 1998, 1999), it is impossible to simulate all the physicochemical characteristics of site water using laboratory water. Therefore, if characteristics of site water, other than hardness, alkalinity and pH, are suspected to influence in the expression of toxicity, it may be useful to perform toxicity testing using site water in order to account for site-specific effects.

Unexposed surface water includes mine-site-collected water that has been collected upstream of a mine effluent discharge or from a nearby reference area. Unexposed surface water from the mine site area may vary in physical, chemical and biological characteristics over time.

Disadvantages of using unexposed surface water as dilution/control water include the following:

- Relatively large volumes of unexposed surface water may be needed for testing, thus additional expense is incurred for the collection, shipment and storage of site-water samples.
- Some laboratory organisms will need acclimation to the unexposed surface water if it is significantly different in physicochemical characteristics from the laboratory water (refer to section 6.9.1).
- Mandatory screening of the water through 60- μ m mesh is required to ensure indigenous populations of micro- or macro-organisms present in surface water do not compete with or impair the health of laboratory test organisms.

In spite of its practical technical disadvantages as control/dilution water, unexposed surface water may provide more site-specific toxicity information. The advantages include the following:

- It reflects the physical/chemical characteristics of the receiving environment.
- It could indicate the potential for non-discharge-related effects.
- Tests conducted may better reflect the influence of receiving environment characteristics on toxicant potency than tests conducted with laboratory water.

In Canada, this practice has been logistically constraining due to the large volumes of water needed to be shipped far distances. However, recently it has been shown that individual sites can identify 1 or 2 test organisms most likely to detect site-specific changes in effluent quality, so tests using receiving water could be conducted on just 1 or

2 species (Taylor et al. 2010). For example, the *Pseudokirchneriella* test requires smaller volumes of water, making this test an ideal candidate for evaluating the effects of receiving-water chemistry on effluent toxicity (Taylor et al. 2010). Sites should determine which type of water best suits their study objectives.

It should be noted that when using site water, problems may arise with reference water exhibiting sublethal responses. Beak International Inc. (BEAK 1998) attributed this problem to indigenous populations of micro-organisms infecting the laboratory organisms. BEAK staff found that boiling the water prior to use before testing was successful in reducing mortality.

Sprague (1997) prepared an extensive review of studies that compared toxicity test results to receiving-water impact, and concluded that effects measured in sublethal toxicity tests correlate with environmental effects most of the time, especially if water collected upstream of the effluent discharge is used as the control/dilution water.

When the purpose of a sublethal toxicity test is to estimate site-specific effects of contaminants, unexposed surface water from the vicinity of the mine site is recommended for use as control/dilution by Environment Canada, the U.S. EPA and the American Society for Testing and Materials (ASTM) in their method and associated guidance documents (Environment Canada 1992a, 1992b, 1998, 2007a, 2007c; US EPA 1994a; ASTM 1998).

Table 6-3: Specifications of the Environment Canada test methods and recommendations for collection, storage and use of site-collected dilution waters

Criteria	<i>Ceriodaphnia dubia</i>	Fathead Minnow	<i>Pseudokirchneriella subcapitata</i>	<i>Lemna minor</i>	Rainbow Trout embryo
Acceptable Dilution Water					
<i>For Culturing</i>	Uncontaminated groundwater, surface water, dechlorinated water, reconstituted water, dilute mineral water, or receiving water	Uncontaminated groundwater, surface water or dechlorinated water	Growth medium	Hoagland's E+ medium	Groundwater, surface water, reconstituted water, dechlorinated water or receiving water
<i>For Testing</i>	Reconstituted, dechlorinated, uncontaminated groundwater or surface water, receiving water	Reconstituted, dechlorinated, uncontaminated groundwater or surface water, receiving water	Reagent water, uncontaminated receiving water, groundwater, surface water or reconstituted water	Modified APHA growth medium, SIS growth medium, receiving water	Groundwater, surface water, reconstituted water, dechlorinated water or receiving water
Site Water					
<i>Collection Point</i>	Upstream from or adjacent to source but removed from effluent exposure	Upstream from or adjacent to source but removed from effluent exposure	Upstream from or adjacent to source but removed from effluent exposure	Upstream from or adjacent to source but removed from effluent exposure	Upstream from or adjacent to source but removed from effluent exposure
<i>Collection Procedure</i>	As for effluent	As for effluent	As for effluent	As for effluent	As for effluent
<i>Acclimation Procedure</i>	Recommends acclimating at least 2 generations of brood organisms before collecting neonates for tests	Recommends acclimation of breeding stock prior to testing	None	Recommends placing plants in site-collected dilution water 18 to 24 hours prior to testing	None
<i>Acclimation Rationale</i>	Recommended hardness $\pm$ 20% of culture water range, alkalinity range $\pm$ 20% of culture water	When outside hardness, alkalinity range $\pm$ 20% of culture water recommended	N/A	Needs to be done for all types of test medium	N/A
<i>Treatments</i>	Recommend filter 60 μ m, boiling if necessary	Recommend filter 60 μ m, boiling if necessary	Filter 0.45 μ m	Filter 1 μ m, then filter 0.22 μ m, nutrient-spiked	Recommend filter 60 μ m or boiling if necessary
<i>Storage</i>	Preferably no more than 14 days; maximum of 1 month at 4°C with no headspace	Preferably no more than 14 days; maximum of 1 month at 4°C with no headspace	Preferably no more than 14 days; maximum of 1 month at 4°C with no headspace	Preferably no more than 14 days; maximum of 1 month at 4°C with no headspace	Preferably no more than 14 days; maximum of 1 month at 4°C with no headspace
<i>During Toxicity Testing</i>	Include lab water control. If screening test shows impairment, treat water by boiling. If impairment remains, use hardness-adjusted lab water.	Include lab water control. If screening test shows impairment, treat water by boiling. If impairment remains, use hardness-adjusted lab water.	Include reagent water control	Include lab water control	Include lab water control. If screening test shows impairment, treat water by boiling. If impairment remains, use hardness-adjusted lab water.

The widespread acceptance of unexposed surface dilution water for predicting site-specific effects is based on knowledge regarding the interaction of contaminants with water quality characteristics. For example, metal toxicity is well known to be influenced by the physicochemical characteristics of water, such as pH, alkalinity, hardness (reviewed by Wang 1997 and Sprague 1995). However, studies comparing results of toxicity tests on effluents and effluent-exposed surface waters from 4 different mine sites indicated that a similar estimation of toxicity could be obtained from tests using unexposed surface water or laboratory water for dilution, especially if the laboratory water was adjusted to the hardness, alkalinity and pH of the site water (BEAK 1998, 1999). This finding indicates that the use of site-collected dilution water may not always be necessary, because laboratory waters can be prepared to reflect site-water characteristics such as hardness, pH and alkalinity.

Naturally elevated levels of organic carbon is one important aspect of some site waters; organic carbon is known to mitigate the effects of metal toxicity by reducing metal bioavailability. Ranges of dissolved organic carbon (DOC) have been measured as high as 58 mg/L in some Ontario lakes (Neary et al. 1990, as cited in Welsh et al. 1993), and such waters may influence the expression of metal toxicity in effluent or effluent-exposed surface water matrices. However, in parallel toxicity tests of mine effluent conducted using site water of moderately high organic carbon (total organic carbon of 9.4 mg/L) and hardness-adjusted laboratory water, there was no significant difference in organism response (BEAK 1999). Of note is that high organic carbon content may not always result in reduced metal bioavailability. In waters of high hardness, concentrations of calcium and magnesium may be high enough to bind with humic acid (which makes up the majority of organic carbon). Therefore, humic acid binding sites would be limited and unavailable to tie up free metal ions. In this situation, decreases in effluent sublethal toxicity due to high receiving water DOC would not occur (Winner 1985). In addition, the type of organic matter will also influence metal bioavailability (i.e., more allochthonous-like organic matter decreases Cu toxicity better than autochthonous-like natural organic matter; Schwartz et al. 2004), which complicates the interpretation of DOC's role in reducing metal toxicities to aquatic biota.

Comparing toxicity of the discharge or effluent-exposed surface water relative to an impaired upstream water

If upstream water is contaminated by nonpoint or upstream-point sources of pollution that are unrelated to the mine operation, a mine may decide to use that water for test dilution purposes in toxicity testing in order to provide an appropriate comparison of test organism responses, as long as the upstream water can support the health of the test organisms. If the upstream water cannot support health of the test organisms, it could be tested separately in a dilution series to quantify its effects with uncontaminated reference site or regular laboratory water used for test control and dilutions.

6.9 Collection, Shipment and Storage of Samples for Sublethal Toxicity Testing

The procedures for the collection, shipment and storage of site-collected dilution water are outlined in each of the Environment Canada and EPA test methods (Environment Canada 1992a, 1992b, 1998, 2007a, 2007b, 2007c; US EPA 1994a, 1994b). Table 6-4 provides estimates of the volumes of site water needed for performing a suite of EEM tests, and includes estimates for effluents or effluent-exposed surface water volume. As recommended by the U.S. EPA (1994a), site-collected dilution water samples should be representative of the water body and be unaffected by recent runoff or erosion events that may cause the water to have a higher total suspended solids concentration.

Table 6-4: Dilution/control water and corresponding effluent volumes for sublethal toxicity tests*

Test	Dilution Water Volume (L)	Effluent Volume (L)
Fathead Minnow	45	21
Rainbow Trout (embryo test)	300	125
<i>Ceriodaphnia dubia</i>	10	4
<i>Pseudokirchneriella subcapitata</i>	1	1
<i>Lemna minor</i>	5 for static, 12 for static-renewal	2 for static, 5 for static-renewal
Testing Series	Dilution Water Volume (L)	Effluent Volume (L)
Fathead, <i>Ceriodaphnia</i> , <i>Pseudokirchneriella</i> , <i>Lemna</i> (static)	65	30
Fathead, <i>Ceriodaphnia</i> , <i>Pseudokirchneriella</i> , <i>Lemna</i> (static-renewal)	72	35
Rainbow Trout embryo, <i>Ceriodaphnia</i> , <i>Pseudokirchneriella</i> , <i>Lemna</i> (static)	325	140
Rainbow Trout embryo, <i>Ceriodaphnia</i> , <i>Pseudokirchneriella</i> , <i>Lemna</i> (static-renewal)	330	150

All volumes are calculated assuming 1 control and 7 test concentrations, except for the Rainbow Trout embryo test where volumes are calculated assuming 1 control and 5 test concentrations. Fathead Minnow assumes 500 ml test volume and 3 replicates, and *Lemna minor* assumes 150 ml test volume and 4 replicates.

* Estimated effluent volumes for marine/estuarine sublethal toxicity tests are outlined in Table 6-2.

6.9.1 Test Organism Acclimation

Pre-acclimation of culture organisms is recommended prior to exposure to site water. As the purpose of using site water for test control and dilutions is to more accurately predict receiving-water impact, the most accurate prediction should be achieved using organisms adapted to the physicochemical conditions of the receiving environment. For example, Lloyd (1965) found that fish cultured in hard water need to lose calcium before they are

as sensitive to metals as fish cultured in soft water. Therefore, if the site water to be used for test dilution is softer than the lab water, pre-acclimating fish to the site water conditions would provide enough time for loss of calcium prior to test initiation. Alternatively, if site-collected water is higher in hardness, then pre-acclimation would allow fish species to accumulate calcium prior to testing.

The Environment Canada methods for *Ceriodaphnia dubia* and Fathead Minnows recommend that cultures be maintained in water of similar hardness, alkalinity and pH (i.e., within 20%) to the site water used for test dilution (Environment Canada 1992a, 2007a). Since many Canadian mines are located beside rivers or lakes of low hardness, it is likely that some acclimation of laboratory cultures would be necessary. BEAK developed a pre-acclimation procedure during the 1997 Aquatic Effects Technology Evaluation (AETE) Program study, later refined in the 1999 study, for Fathead Minnows and *Ceriodaphnia dubia* (BEAK 1998, 1999). Based on the expected hardness, alkalinity and pH of the site water, cultures are gradually introduced to laboratory water of decreasing hardness over several days until the appropriate hardness is reached. This procedure was adapted from that used by B.A.R. Environmental Inc. (BAR) during its 1996 AETE Program study, in which cultures were gradually acclimated to hardness-adjusted laboratory water and site water if screening of un-acclimated organisms showed impairment to laboratory organisms (BAR 1997).

For pre-acclimation of cultures, laboratory water of reduced hardness may be prepared by diluting standard laboratory water with deionized water. Hardness can be increased by adding salts, used for the preparation of reconstituted water in the appropriate amounts (Table 6-5). When hardness-adjusting water, it is important to keep the alkalinity level appropriate to the hardness, because alkalinity affects the speciation of metals (US EPA 2002; Laurén and McDonald 1986). Appropriate hardness and alkalinity relationships are available in Table 6-5, and additional values may be interpolated (US EPA 1994b).

Detailed procedures for pre-acclimation of *Ceriodaphnia dubia* and Fathead Minnow are described below. References to hardness assume a corresponding change in alkalinity and pH. No pre-acclimation procedures are described for the Rainbow Trout test, since eggs are delivered from the hatchery and used in testing within 24 hours (Environment Canada 1998). Similarly, *Pseudokirchneriella* and *Lemna* cultures are maintained in standard culture media that are different from standard testing media, although *Lemna* does allow for some pre-acclimation of cultures, since plants are transferred to the test medium 18 to 24 hours prior to initiation of testing (Environment Canada 2007c).

Table 6-5: Preparation of different hardness/alkalinities

Water Type	Reagent Added (mg/L) ¹			Final Water Quality			
	NaHCO ₃	CaSO ₄ ·2H ₂ O	MgSO ₄	KCL	pH ²	Hardness ³	Alkalinity ³
Very soft	12.0	7.5	7.5	0.5	6.4–6.8	10–13	10–13
Soft	48.0	30.0	30.0	2.0	7.2–7.6	40–48	30–35
Moderately hard	96.0	60.0	60.0	4.0	7.4–7.8	80–100	60–70
Hard	192.0	120.0	120.0	8.0	7.6–8.0	160–180	110–120
Very hard	384.0	240.0	240.0	16.0	8.0–8.4	280–320	225–245

Source: US EPA (1994a)

¹ Add reagent-grade chemicals to deionized water.

² Approximate equilibrium pH after 24 hours of aeration.

³ Expressed as milligrams (mg) CaCO₃/L.

Ceriodaphnia dubia

Ceriodaphnia cultures are initiated and maintained according to the Environment Canada standard method. To pre-acclimate cultures to the hardness of the site-collected dilution water, a brood of neonates, less than 24 hours old, is initiated in water, reduced in hardness by 20% from that of laboratory water by addition of deionized water. Each day, organisms are transferred to new solutions, decreased by a further 20-30% in hardness. Once the desired hardness is reached (within approximately 1 week) and the culture organisms pass the health criteria of the Environment Canada method (i.e., production of at least three broods, total neonate production of at least 15 per adult, less than 20% adult mortality), a new culture is initiated. The second-generation cultures are maintained in the hardness-adjusted water until organisms pass the health criteria (approximately 1 week). Selenium and vitamin B12 are added to the hardness-adjusted culture water if low in hardness, as recommended by the standard method.

Fathead Minnows

Acclimation and pre-acclimation procedures carried out by BAR and BEAK during the 1996 and 1997 AETE studies used breeding tanks of Fathead Minnows, gradually changed in hardness and alkalinity to that of the applicable site water (hardness-adjusted laboratory water or HALW) (BAR 1997; BEAK 1998). Eggs were collected from the adjusted-water tanks and reared at the HALW until hatching. However, preliminary work completed by BEAK in early 1999 showed that eggs could be hatched out in HALW without prior acclimation of the breeding tanks. By omitting breeding tanks from the pre-acclimation procedure, the volume of HALW needed for pre-acclimation is reduced while maintaining a supply of eggs for hatching in a number of different water types.

Eggs are hatched according to the Environment Canada standard method, with fresh water renewal every 24 hours.

6.9.2 Screening Tests

Once organisms are pre-acclimated to the physicochemical conditions of the site water, they may be exposed to the site water in screening tests. If organisms exposed to the site water meet the test method control acceptability criteria, the site water may be considered suitable for use as dilution water provided the site water meets the control validity criteria for the test method. Comparison of the site-water response to the laboratory response in a screening test may reveal a statistically significant reduction in reproduction and/or survival, even though the site-water-exposed group is within the control acceptability criteria for the test. As long as the site water meets the control validity criteria for the test method, it may be considered suitable for testing. For additional information, consult the Environment Canada test methods.

6.9.2.1 Screening Test Impairment

If organisms are given an adequate opportunity to gradually acclimate to the physicochemical characteristics of the site water (by exposure to hardness-adjusted laboratory water), impairment observed during screening tests should not be due to shock exposure to different water quality characteristics such as hardness, alkalinity or pH. Therefore, impairment would likely be due to the presence of harmful biological agents or toxicants.

Special attention should be paid to any 100% site-water exposures showing significant mortality in one or more replicates. Microbiological organisms present in the site water may impair the health of test organisms, and anecdotal evidence from several ecotoxicity laboratories indicates that impairment by indigenous micro-organisms usually occurs after a few days of exposure. For example, this has been manifested in Fathead Minnow tests conducted at BEAK as a sudden onset of significant mortality in the site-water control, often in one or two replicates only. Occasionally, evidence of fungal or bacterial growth may be observed in the test vessels. If such contamination is indicated by a screening test, acclimation of the organisms will most likely not result in a removal of impairment. Therefore, either the site water should be treated by a suitable means to remove the impairment (i.e., boiling or ultraviolet treatment—see below), or hardness-adjusted laboratory water should be used as a surrogate.

Limited laboratory trials of site water showed that impairment could be removed by boiling site water gently for 10 minutes and cooling it prior to use in testing. Other treatments have been reported in the literature, such as ultraviolet light and 0.45-µm filtration (Grothe and Johnson 1996; Kszos et al. 1997). If site water is collected during late spring to early fall, some form of biological contamination should be expected (unless experience with a particular site water indicates otherwise), and precautions such as boiling should be taken.

If the impairment is due to chemical contamination, the suitability of the site water for use as a dilution and control water is questionable, even though cultures may be acclimated to naturally high levels of metals in site water (see section 6.9.1). If the cultures are exposed to higher levels of contaminants, post-acclimation can result in either higher or lower sensitivity of laboratory organisms, depending on the contaminant, organism and the water characteristics, and the utility of the post-screening test acclimation procedure becomes questionable. If impairment is detected in a screening test, the recommended procedure is to attempt treatment by boiling or to use hardness-adjusted laboratory water as a surrogate dilution water. If it is suspected that site water is contaminated, including a boiled-site-water exposure in screening tests would resolve the question of the effectiveness of that treatment for the site water of interest.

6.9.3 Effluent or Effluent-Exposed Surface Water Toxicity Tests

Table 6-3 summarizes the recommendations for collecting, storing and using site-collected dilution waters in sublethal toxicity testing.

Once site water has been deemed suitable for testing, toxicity tests may be initiated as per the appropriate Environment Canada testing method. In addition to the site-water control, an additional control, using water from the laboratory culture, needs to be included in testing to serve as a check of culture health and site-water quality. No additional control is necessary when the test-control dilution water is the same as the culture water. In the case of *Lemna minor* and *Pseudokirchneriella subcapitata* tests, the laboratory control would be the standard-test growth medium, specified in the standard method.

If site-collected water is deemed unsuitable for use as test control and dilution water, a treatment such as boiling should be attempted in order to remove the impairment. If successful, the treated site-collected dilution water should be used in testing. However, boiling large volumes of water may not be practical for tests such as the Rainbow Trout test, and hardness-adjusted laboratory water may be a more suitable alternative. If no practical treatment can be found to remove impairment to test organisms caused by the site-collected dilution water, hardness-adjusted water should be used as the dilution water with pre-acclimated organisms. Care should be taken to match the pH of the characteristics of the site-collected dilution water as closely as possible.

6.10 Use of Sublethal Toxicity Testing in Resolving Confounding Influences

Sublethal toxicology data also have the potential utility to aid in the resolution of confounding factors. A multistakeholder group on metal mining toxicology has elaborated on this third use for sublethal toxicity data.

Estimating the relative contribution of mine effluent releases and other natural and/or anthropogenic influences on sublethal toxicity in the same receiving water body

During any phase of the EEM program, sublethal toxicity test data can be used to deal with situations where there are confounding influences. Sometimes the site characteristics do not permit full determination of the mine's effluent effects even with an adapted study design. Information from sublethal toxicity testing may then help in the interpretation of field results. The choice of when to use sublethal toxicity tests in this application is up to the mine operator and the nature of the confounding influences. However, the confounding influence scenario where sublethal toxicity testing would be most relevant is the multiple-point-source discharge and/or nonpoint-source input situation. Sublethal tests or frequency monitoring should be determined based on the site-specific nature of confounding influence situations.

Estimates of sublethal toxicity can help in understanding the relative contribution of diverse industrial or municipal discharges to effects on aquatic organisms in the receiving water, whether the discharges are from upstream point or nonpoint sources (e.g., municipal landfill leachate, agricultural runoff) or the mine's property. The upstream contribution of an observed environmental effect can be estimated, given surface water sublethal toxicity data, discharge flow, and features of dispersion into the receiving environment. If plumes from different discharges at a mine site overlap, more effort is necessary to distinguish the toxic contributions of the mine's discharge sources vs. upstream sources. Samples of surface water from key locations in the high effluent exposure areas could be tested, to estimate the combined toxic contribution of the sources.

The following is a 3-step procedure for assessing the relative contribution from different sources of sublethal toxicity to the high effluent exposure receiving environment:

- Conduct a battery of sublethal toxicity tests on samples collected from all significant discharge sources from the mine's property. Use standard laboratory water for test dilutions and control, or unexposed site water. This estimates the absolute sublethal toxicity of each mine-site discharge. Repeat the sampling and testing on any discharge that is known to be variable in toxicity, in order to obtain an estimate of the degree of variability.
- Conduct a parallel battery of sublethal toxicity tests for each discharge to a river, using water collected directly upstream from the discharge point for dilution and control. For lakes or estuaries, carry out the parallel battery of tests by collecting control/dilution water from outside the zone immediately affected by the discharges. Separate and simultaneous controls should be run using standard uncontaminated water as a QA measure. It should be recognized that "upstream" sources of control/dilution water might already be contaminated by other effluent discharges or sources of toxicants. Accordingly, the upstream dilution water might contribute to significant effects on growth or reproduction in concentrations of the effluents being studied or even in control vessels. This would not invalidate the results, because the purpose of the investigation is to evaluate the relative contributions of discharges to the total toxicity of the receiving water.

- Confirmation of the relative contribution of discharges is recommended, and can be achieved by conducting sublethal toxicity tests on samples of surface water from the water body receiving the discharges (so-called “ambient” tests). This can aid in:
 - a) confirming whether an effluent has a measurable toxicity after mixing into the receiving water;
 - b) estimating the persistence in the receiving water of toxicity from all contributing sources; and
 - c) determining the combined toxicity resulting from the mixing of all point and nonpoint sources, as an estimate of the overall effect on the receiving environment.

Testing samples of surface water, which receives discharges or toxicants from multiple sources, should be done synoptically and ideally during low-flow or worst-case periods. At a minimum, sampling should be carried out over as short a period of time as possible (e.g., 1 or 2 days). Repeated rounds of sampling and testing would be desirable if the toxicity of the discharges were variable. The above guidance on conducting toxicity assessment studies to estimate the contribution of multiple-discharge sources to instream effects is based on the 8 site investigations conducted under the U.S. EPA Complex Effluent Toxicity Testing Program (CETTP). Detailed reports on these studies were prepared by Mount and Norberg-King (1985, 1986), Mount et al. (1984, 1985, 1986a, 1986b, 1986c) and Norberg-King and Mount (1986). The results of CETTP testing, including independent critiques and re-analyses, were reviewed by Sprague (1997) during a project commissioned by the Aquatic Effects Technology Evaluation (AETE) Program. Sprague concluded that the U.S. EPA CETTP studies provided valid findings that should be considered by Canadian metal mining companies when designing aquatic environmental monitoring programs at their mine sites.

Mines may elect to conduct additional investigations where the most sensitive species in the effluent produced an IC_{25} of less than 30%. Below are additional recommended investigations. At a minimum, the most sensitive test species can be used to estimate the geographic extent of the potential response. Alternatively, the results of the toxicity test(s) may lead to or trigger other recommended laboratory or field monitoring tools.

A tiered approach to resolving the confounding influences is recommended, starting with these additional recommended investigations:

- 1) re-testing with the sublethal test that provided the most sensitive IC_{25} result, using upstream or reference-site water for test control and dilutions; or
- 2) receiving-water toxicity testing with samples collected from the area where a sublethal response is predicted.

6.11 References

[ASTM] American Society for Testing and Materials. 1998. Standard guide for conducting acute toxicity tests on aqueous ambient samples and effluents with fishes, macroinvertebrates, and amphibians, designation E1192-97. Conshohocken (PA): Annual book of ASTM standards. Section 11: Water and environment technology.

[BAR] B.A.R. Environmental Inc. 1997. Toxicity assessment of mining effluents using upstream or reference site waters and test organism acclimation techniques. Aquatic Effects Technology Evaluation Program. AETE Project 4.1.2a.

[BEAK] Beak International Incorporated. 1998. Additional tool evaluations. Aquatic Effects Technology Evaluation Program. Beak Reference 20776.1.

[BEAK] Beak International Incorporated. 1999. Final report: effects of dilution water on the results of sublethal toxicity tests. Report to Natural Resources Canada. Beak reference 33940.

Chapman PM. 2000. Whole effluent toxicity testing – usefulness, level of protection, and risk assessment. *Environ Toxicol Chem* 19(1):3-14.

Environment Canada. 1992a. Biological test method: test of larval growth and survival using fathead minnows. Ottawa (ON): Environmental Technology Centre. Report EPS 1/RM/22, February 1992, Amended in September 2008.

Environment Canada. 1992b. Biological test method: fertilization assay with echinoids (sea urchins and sand dollars). Ottawa (ON): Environmental Technology Centre. Report EPS 1/RM/27, December 1992, Amended in November 1997.

Environment Canada. 1998. Biological test method: toxicity tests using the early life stages of salmonid fish (rainbow trout). Report EPS 1/RM/28, 2nd ed., July. Ottawa (ON): Environmental Technology Centre.

Environment Canada. 1999. Recommended procedure for the importation of test organisms for sublethal toxicity testing. Ottawa (ON): Environmental Technology Centre.

Environment Canada. 2001. Revised procedures for adjusting salinity of effluent samples for marine sublethal toxicity testing conducted under environmental effects monitoring (EEM) programs. Ottawa (ON): Method Development and Applications Section, Environmental Technology Centre.

Environment Canada. 2007a. Biological test method: test of reproduction and survival using the Cladoceran *Ceriodaphnia dubia*. Ottawa (ON): Environmental Technology Centre. Report EPS 1/RM/21, 2nd edition, February.

Environment Canada. 2007b. Biological test method: growth inhibition test using a freshwater alga. Ottawa (ON): Environmental Technology Centre. Report EPS 1/RM/25, 2nd edition, March 2007.

Environment Canada. 2007c. Biological test method: test for measuring the inhibition of growth using the freshwater macrophyte *Lemna minor*. Ottawa (ON): Environmental Technology Centre. Report EPS 1/RM/37, 2nd edition, January 2007.

[ESG] ESG International Inc. 1999. AETE synthesis report of selected technologies for cost-effective environmental monitoring of mine effluent impacts in Canada. AETE Project 4.1.4, March 1999. Ottawa (ON): Canadian Centre for Mineral and Energy Technology, Natural Resources Canada.

Grothe DR, Johnson DE. 1996. Bacterial interference in whole effluent toxicity tests. *Environ Toxicol Chem* 15(5):761-764.

Kszos LA, Stewart AJ, Sumner JR. 1997. Evidence that variability in ambient fathead minnow short-term chronic tests is due to pathogenic infection. *Environ Toxicol Chem* 16(2):351-356.

Laurén DJ, McDonald DG. 1986. Influence of water hardness, pH and alkalinity on the mechanisms of copper toxicity in juvenile rainbow trout, *Salmo gairdneri*. *Can J Fish Aquat Sci* 43:1488-1496.

Lloyd R. 1965. Factors that affect the tolerance of fish to heavy metal poisoning. In: *Biological problems in water pollution, third seminar, 1962*. Washington (DC): U.S. Public Health Service. Publ. 999-WP-25. p. 181-187.

[MDDEP] Ministère du Développement durable, de l'Environnement et des Parcs du Québec. 2007. Détermination de la toxicité : inhibition de la croissance chez l'algue *Pseudokirchneriella subcapitata*. Centre d'expertise en analyse environnementale du Québec. MA. 500-P.sub 1.0, Revised in September 2007).

Mount DI, Norberg-King TJ, editors. 1985. Validity of effluent and ambient toxicity tests for predicting biological impact, Scippo Creek, Circleville, Ohio. Washington (DC): U.S. Environmental Protection Agency. EPA 600/3-85/044.

Mount DI, Norberg-King TJ, editors. 1986. Validity of effluent and ambient toxicity tests for predicting biological impact, Kanawha River, Charleston, West Virginia. Washington (DC): U.S. Environmental Protection Agency. EPA 600/3-86/006.

Mount DI, Thomas NA, Norberg-King TJ, Barbour MT, Roush TH, Brandes RW. 1984. Effluent and ambient toxicity testing and instream community response on the Ottawa River, Lima, Ohio. Washington (DC): U.S. Environmental Protection Agency. EPA 600/3-84/080.

Mount DI, Steen AE, Norberg-King TJ. 1985. Validity of effluent and ambient toxicity testing for predicting biological impact on Five Mile Creek, Birmingham, Alabama. Washington (DC): U.S. Environmental Protection Agency. EPA 600/8-85/015.

Mount DI, Norberg-King TJ, Steen AE. 1986a. Validity of effluent and ambient toxicity tests for predicting biological impact, Naugatuck River, Waterbury, Connecticut. Washington (DC): U.S. Environmental Protection Agency. EPA 600/8-86/005.

Mount DI, Steen AE, Norberg-King TJ. 1986b. Validity of effluent and ambient toxicity tests for predicting biological impact, Back River, Baltimore Harbor, Maryland. Washington (DC): U.S. Environmental Protection Agency. EPA 600/8-86/001.

Mount DI, Steen AE, Norberg-King TJ. 1986c. Validity of ambient toxicity tests for predicting biological impact, Ohio River near Wheeling, West Virginia. Washington (DC): U.S. Environmental Protection Agency. EPA 600/3-85/071.

Neary BP, Dillon PJ, Munro JR, Clark BJ. 1990. The acidification of Ontario lakes: an assessment of their sensitivity and current status with respect to biological damage. Toronto (ON): Ontario Ministry of the Environment. 171 p.

Norberg-King TJ, Mount DI, editors. 1986. Validity of effluent and ambient toxicity tests for predicting biological impact, Skeleton Creek, Enid, Oklahoma. Washington (DC): U.S. Environmental Protection Agency. EPA 600/8-86/002.

Schwartz ML, Curtis PJ, Playle RC. 2004. Influence of natural organic matter source on acute copper, lead, and cadmium toxicity to rainbow trout (*Oncorhynchus mykiss*). Environ Toxicol Chem 23(12):2889-2899.

Sprague J. 1995. Factors that modify toxicity. In Rand G, editor. Fundamentals of aquatic toxicology. 2nd edition. Washington (DC): Taylor and Francis.

Sprague J. 1997. Review of methods for sublethal aquatic toxicity tests relevant to the Canadian metal-mining industry. Ottawa (ON): Aquatic Effects Technology Evaluation Program, Canadian Centre for Mineral and Energy Technology, Natural Resources Canada.

Taylor LN, Van der Vliet LA, Scroggins RP. 2010. Sublethal toxicity testing of Canadian metal mining effluents: national trends and site-specific uses. Hum Ecol Risk Assess 16(2):264-281.

[US EPA] United States Environmental Protection Agency. 1994a. Short-term methods for estimating the toxicity of effluents and receiving water to freshwater organisms. Third edition. Cincinnati (OH): Environmental Monitoring Systems Laboratory, U.S. Environmental Protection Agency. EPA /600/4-90/027.

[US EPA] United States Environmental Protection Agency. 1994b. Interim guidance on determination and use of water-effect ratios for metals Office of Water, U.S. Environmental Protection Agency. EPA 823-B-94-001.

[US EPA] United States Environmental Protection Agency. 1995. Short-term methods for estimating the toxicity of effluents and receiving water to West Coast organisms. First edition. Cincinnati (OH): Environmental Monitoring Systems Laboratory, U.S. Environmental Protection Agency. EPA 600/R-95/136.

[US EPA] United States Environmental Protection Agency. 2002. Short-term methods for estimating chronic toxicity of effluent and receiving waters to marine and estuarine organisms. Third edition. Cincinnati (OH): Environmental Monitoring Systems Laboratory, U.S. Environmental Protection Agency. EPA/821/R-02/014, October 2002.

Wang W. 1997. Factors affecting metal toxicity to (and accumulation by) aquatic organisms - an overview. Environ Internat 13:437-457.

Welsh PG, Skidmore JF, Spry DJ, Dixon DG, Hodson PV, Hickie BE. 1993. Effect of pH and dissolved organic carbon on the toxicity of copper to larval fathead minnows (*Pimephales promelas*) in natural lake waters of low alkalinity. Can J Fish Aquat Sci 50:1356-1362.

Winner RW. 1985. Bioaccumulation and toxicity of copper as affected by interactions between humic acid and water hardness. Water Res 19:449-455.

Table of contents

7.	Sediment Monitoring	7-1
7.1	Overview	7-1
7.2	Collection of Sediment Samples	7-2
7.2.1	Field Measurements and Observations	7-2
7.2.2	Criteria for Selection of a Sample Collection Device	7-2
7.2.3	Collection Device Penetration Depth	7-4
7.2.4	Sample Volume	7-4
7.2.5	Grab Sampler Operation	7-4
7.2.6	Sub-sampling of Sediment Grab Samples	7-5
7.2.7	Criteria of Acceptability of Samples	7-5
7.2.8	Replicate Samples	7-6
7.2.9	Sediment Variables that May be Measured in the Field	7-6
7.3	Sample Handling and Analysis	7-7
7.3.1	Procedures for Handling of Sediment Samples	7-7
7.3.2	Compositing Sediment Grab Samples	7-7
7.3.3	Sample Containers	7-7
7.3.4	Transportation and Storage of Sediment Samples	7-8
7.3.5	Laboratory Test Sample Preparation	7-9
7.3.6	Prevention of Sediment Sample Contamination	7-9
7.4	Sediment Variables	7-10
7.4.1	Determination of Particle Size Distribution	7-10
7.4.2	Determination of Total Organic Carbon Content	7-12
7.4.3	Determination of Total Metal Concentrations	7-12
7.5	Sediment Toxicity Testing	7-14
7.6	Confounding Factors	7-14
7.7	References	7-15

7. Sediment Monitoring

7.1 Overview

As part of each benthic invertebrate community survey, mines collect sediment samples for analysis of total organic carbon content and particle size distribution, if it is possible to sample sediment (*Metal Mining Effluent Regulations* [MMER], Schedule 5, section [s.] 16a)(iii)). Sediment samples are collected at the same sampling stations and at the same time as benthic invertebrate samples.

More sampling stations within each area may help to better understand potential contaminant concentrations in the exposure area. Each study design for benthic invertebrate community surveys should identify the sediment sample collection and laboratory analysis methods to be used (field and laboratory methodologies selected). The results of these analyses, including calculation of the mean, median, standard deviation, standard error, minimum and maximum values for each sampling area, are included in the interpretative report. The results of analyses of particle size distribution and total organic carbon are used to determine if there are habitat differences between the exposure and reference areas, in order to aid in the interpretation of the results of benthic invertebrate community surveys. The overall purpose of sediment monitoring is to answer the question, “Are there habitat differences that may contribute to effects in the benthic invertebrate community?”

For monitoring programs where the sampling of benthic invertebrates is conducted in an erosional habitat, sediment sampling may not be possible as a standard supporting environmental variable; in these cases the sediment monitoring data would not be reported. Some methods for retrieving sediments from erosional zones require elaborate equipment or two field visits, one for the placement and one for the collection of sediment traps. However, site-specific conditions may warrant the consideration of sediment sampling in some erosional habitats, as useful information regarding exposure can be obtained with these methods. These approaches could be considered during the study design exercises for magnitude and geographic extent or investigation of cause as an additional supporting variable or tool for determining effects.

When a benthic invertebrate community survey is conducted as part of magnitude and geographic extent or investigation of cause, it is recommended that each sediment sample collected also undergo chemical analysis, (e.g., metals). The study design for that study should identify the parameters for which sediment samples will be analyzed and the laboratory methods to be used, and the results of analyses should be reported in the interpretative report.

When investigation of cause (IOC) is conducted to identify the causes of the effect on the benthic invertebrate community, detailed studies of sediment may be appropriate as a tool to help determine the cause of effect. Chapter 12 contains extensive technical guidance

on the conduct of detailed sediment studies (e.g., sediment mass transport, depositional rate, coring, chemistry, sediment toxicity testing, sediment quality triad, pore water analysis, toxicity identification evaluation and toxicity reduction evaluation) recommended as part of IOC.

7.2 Collection of Sediment Samples

This section provides guidance on the collection, handling, storage and transportation of sediment samples, and on field measurements and observations. This guidance applies to sediment samples collected in all phases of the environmental effects monitoring (EEM) program.

7.2.1 Field Measurements and Observations

Field measurements and observations are critical to any sediment collection study. It is recommended that the following information (Mudroch and MacKnight 1991) be recorded at the time each sediment sample is collected from a sampling station:

- sample number, replicate number, station number, site identification (e.g., name)
- time and date of the collection of the sample
- ambient weather conditions, including wind speed and direction, wave action, current, tide, vessel traffic, temperature of both the air and water, thickness of ice if present
- sampling area location (e.g., positioning information) and location of any replicate samples
- type of platform/vessel used for sampling (e.g., size, power, type of engine)
- type of sediment collection device and any modifications made during sampling
- the water depth at each sampling station and the sediment sampling depth
- name of personnel collecting the samples
- details pertaining to unusual or unpredicted events that might have occurred during the operation of the sampler (e.g., possible sample contamination, equipment failure, unusual appearance of sediment integrity, control of vertical descent of the sampler)
- description of the sediment, including texture and consistency, colour, odour, presence of biota, estimate of quantity of recovered sediment by a grab sampler, or length and appearance of recovered cores (photographs provide a good permanent record of a retrieved sample)
- deviations from standard operating procedures

7.2.2 Criteria for Selection of a Sample Collection Device

There are numerous methods and procedures reported in the literature that describe how to collect various types of sediment samples and to help determine the most appropriate

sampling devices for different types of environments, including freshwater, marine or estuarine environments (for reviews see Baudo et al. 1990; Mudroch and MacKnight 1991; Environment Canada 1994; ASTM 1992; Burton 1992). Environment Canada (1994) Baudo et al. (1990) and Håkanson and Jansson (1983) suggest several factors that should be considered for the selection of sediment samplers and sampling location. The ideal sediment sampler should for the most part:

- permit free water passage during descent, to avoid a pressure wave
- have a sharp-edged cutting surface, a small-edge angle, smooth inside surface, and small wall thickness to minimize disturbance
- close tightly for the ascent
- allow sub-sampling
- have the capability of adjusting weight for penetration of different substrates
- be able to retrieve a volume of sediment large enough to meet the analytical test requirements
- effectively and consistently retrieve sediments from various water depths
- effectively and consistently retrieve sediments from the desired sampling depth
- not contaminate or influence the nature of the sediment
- require a minimum of supportive equipment
- be easy and safe to operate and not require extensive training of personnel
- be easily transported to and assembled at the sampling site

Most sediment samplers are designed to consistently isolate and retrieve a volume of sediment to a required depth below the sediment surface with minimum disruption to the integrity of the sample and no contamination of the sample. Maintaining the integrity of the collected sediments is of primary concern in most studies, since disrupting the structure of the sediment may change the physico-chemical and biological characteristics, which in turn could influence the partitioning, complexation, speciation and bioavailability of the toxicants. Sometimes it is also important to maintain the profile if sectioning is required at different depths. These issues become even more important during IOC monitoring studies when sediment may be collected for toxicity tests or more complex analytical methods. In general, it is recognized that it is difficult to collect a sediment sample with most sampling devices without some degree of disruption.

There are three main types of sediment samplers: grab, core and dredge samplers. For the first biological monitoring studies and studies conducted to confirm presence or absence of effects, grab samplers are recommended. Grab samplers are used to collect surficial sediments for the determination and assessment of the horizontal distribution of sediment characteristics. Details on this topic can be found in de Groot and Zschuppe (1981), Baudo et al. (1990), ASTM (1992), Burton (1992) and Sly and Christie (1992).

Core samplers collect a column of sediment to examine the historical or vertical distribution of the physical and chemical characteristics of the sediment (Environment Canada 1994). Corers are preferred in cases where the integrity of the profile is essential, as they are the

least disruptive. For these reasons, corers should be considered for magnitude and geographical extent and IOC studies. For additional information on core samplers, refer to Environment Canada (1994) and Chapter 12.

Dredges are used primarily for the collection of benthos, since they are usually equipped with net sides designed to filter out fine-grained sediments and retain coarse sediments and fauna. It is virtually impossible to accurately measure the surface area covered by the dredge sampler, or judge the depth to which the sediment sample has been collected. In addition, sediment integrity is disrupted, pore water excluded, and fine-grained sediments lost during ascent using dredge samplers. For these reasons, only grab (first biological monitoring studies and studies aim to confirm presence or absence of effects) and core samplers (magnitude and geographic extent and IOC) are being recommended for the collection of sediments.

7.2.3 Collection Device Penetration Depth

The desired depth of sediment penetration is a decision that depends upon the type of sampling device, the nature of the sediment, and the volume of sediment required. The actual depth of penetration depends primarily on the type of sampling device and the nature of the sediment. Generally, the most recently introduced contaminants of concern and most infaunal organisms are found in the upper 2 cm. Epifaunal organisms also have access to this horizon (Burton 1992). Therefore, a preferred penetration depth of 10-15 cm and a minimum penetration depth of 6-8 cm are recommended to ensure minimum disturbance of the upper layer during sampling. This depth is also appropriate for monitoring studies where historical contamination is not a priority (upper 0-5 cm of sediment).

7.2.4 Sample Volume

The recommended minimum volume or weight of sediment needed for each end use should be determined on a case-by-case basis and is available in Table 3 of Environment Canada (1994). Before commencing a sampling program, the type and number of analyses and tests should be determined, and the required volume or weight of sediment per sample calculated. Each physico-chemical test requires a specific amount of sediment. After the sample size is determined, it is important to compare the sample size required with the capacity of the sampler to deliver the desired amount of sediment, and reassess the number of replicate samples per station. The volume or weight requirements might dictate further sample handling such as sub-sampling, compositing, or sample splitting.

7.2.5 Grab Sampler Operation

When collecting bottom sediments with grab samplers, the speed of descent of the sampling device should be controlled and the sampler should not be permitted to “free fall.” To minimize twisting during the descent, a ball bearing swivel should be used to attach the sampler to the cable. The sampler should contact the substrate or be positioned just above it

and only its weight or piston mechanism should be used to force it into the sediment. The winching system should be in place to control both the ascent and descent of the sampling device, especially in deep water. After the sample is contained, the sampling device should be lifted slowly off the bottom, then steadily raised to the surface at about 30 cm/s. When the sampler is brought to the surface, the outside of the sampler should be carefully rinsed with water from the sampling station to remove material that could potentially contaminate the sample during transfer. The sampler should be inspected to ensure that it has closed properly. The standard operating procedures specific for each grab sampler should be followed in order to ensure proper operation of the sampler.

Regardless of the type of samplers used, standard operating procedures for each device should be immediately accessible, and all personnel involved with the collection of samples should be familiar with these procedures. The sampling vessel or platform should be stationary, and sufficiently stable to permit inspection and handling of the retrieved sample. Field notes should accompany each sample that is collected. The sampling device should be cleaned thoroughly between sampling stations and between within-station samples by dipping the sampler into and out of the water at a rapid speed to wash the sediment off. Alternatively, a hose can be used to wash the sediment off of the sampler with water from the sampling station. The sampler should be rinsed with water from the next sampling station before collecting a sample.

7.2.6 Sub-sampling of Sediment Grab Samples

If sediment grab samples are to be sub-sampled, access to the surface of the sample without a loss of water or fine-grained sediment is a prerequisite for selection of the sampler.

The non-turbid overlying water, if present, should be gently siphoned off before the sediment is sub-sampled, using a flat, clean scoop (e.g. Teflon[®] or a similarly inert, non-contaminating, non-reactive material) or a suitable hand-coring device. Ideally, each sub-sample should be placed into a clean, separate, pre-labelled container. The labelled sample container should be sealed and the air excluded.

In the event that the collection device does not allow access to the surface, the following procedures should be followed. Upon retrieval of the sample, the contents should be carefully deposited into a clean, inert container that is the same shape as the sampler. The sampler is placed into the container and the jaws opened slowly to allow the sample to be deposited into the container with as little disturbance as possible. Once the sample is in the container, sub-samples can be collected from the sample with a hand corer or scoop. The edges of the sample where the sediments may be disturbed during removal from the sampler should be excluded during sub-sampling.

7.2.7 Criteria of Acceptability of Samples

All samples should be visually inspected to ensure that:

- the desired depth of penetration has been achieved
- there is no evidence of incomplete closure of the grab sampler, or that the grab sampler was inserted on an angle or tilted upon retrieval (i.e., loss of sediment)

If the collected sample fails any of the criteria listed above, then the sample should be rejected and another sample collected at the site. The location of consecutive attempts should be as close to the original attempt as possible while avoiding any overlap and, where the direction of the current is known, consecutive attempts should be located in the opposite direction of the current, or “upstream.” Rejected sediment samples should be discarded in a manner that will not affect subsequent samples at that station or other possible sampling locations.

7.2.8 Replicate Samples

A single sediment sample from a sampling station will impart little information on the variability. Environment Canada (1994) therefore recommends the following for the minimum number of replicate samples:

- When replicate samples from a sampling station are recommended, the collection of a minimum number of five replicate samples within a sampling station is recommended unless determined otherwise from preliminary sampling and analysis.
- The collection of replicate samples is needed as part of the QA/QC of any good sampling program and should comply with the data quality objectives.
- The number of replicate samples should be higher at stations located close to a source of contamination (Skei 1992).

Collecting separate replicate samples at each sampling station allows for quantitative statistical comparison within and among different stations (Holland et al. 1993). The collection of separate samples within a sampling area can impart valuable information on the heterogeneity of the sediments. Separate sub-samples from the same grab can be used to measure the variation within a sample but not necessarily within the sampling station.

The number of replicates needed per sampling station is a function of the need for sensitivity or statistical power. Typically, the smallest deviation from the null hypotheses that is considered scientifically or environmentally important to detect should be decided a priori, together with the power of the test that is desired for the specific alternative (Green 1989).

7.2.9 Sediment Variables that May be Measured in the Field

In addition to the required determination of total organic carbon and particle size distribution, it is recommended that the following sediment variables be measured in the field, particularly during magnitude and geographic extent and investigation of cause:

- temperature and pH of the sediment at the sediment-water interface
- a measure of the redox potential of the sediments to determine if the sediments are oxic or anoxic, or to determine the depth of the interface between these conditions in the sediments. Dissolved oxygen is recommended for freshwater sediment and redox potential (Eh) is recommended for marine sediments.

These measurements could be useful for the interpretation of the analytical results.

7.3 Sample Handling and Analysis

7.3.1 Procedures for Handling of Sediment Samples

Any time that sediment samples are handled, it is recommended that the following procedures be observed:

- Sediment might contain a mixture of hazardous substances, so it is prudent to avoid skin contact with sediments by wearing protective clothing and equipment (e.g., gloves, boots, lab coats or aprons, safety glasses, and respirator) during sampling, sample handling, and the preparing of test substances.
- Handling of samples should be performed in a well-ventilated area (e.g., outside, in a fume hood, or in an enclosed glove box) to minimize the inhalation of sediment gases.
- Work surfaces should be covered with Teflon[®] sheets, high-density polyethylene trays, or other impervious or disposable, similarly inert material.
- A spill control protocol should be in place in the laboratory or sampling vessel, and participants in the project should be familiar with all standard operating procedures and recommendations.

7.3.2 Compositing Sediment Grab Samples

If the objective of the study dictates compositing sub-samples from separate grabs within a sampling station, the sub-samples may be placed into one clean sample container and, when full, sealed without trapped air. Compositing of sediment samples or sub-samples may also be performed in the laboratory.

7.3.3 Sample Containers

Environment Canada (1994) provides information concerning the storage and transportation of field-collected sediment samples.

Whole-sediment samples may be transferred directly from a sampler into a clean, large-volume (e.g., > 1 L) container. If smaller volumes of sediment are collected or sub-sampled, containers with wide mouths and Teflon[®]-lined lids are recommended for volumes ranging from 250 to 1000 ml.

If samples are to be stored at 4°C, sample containers should be filled to the rim and air excluded during capping. If samples are to be frozen for storage, glass containers should not be filled completely. A space of approximately 2.5 cm should be left to accommodate expansion of the sample when frozen; however this will depend on the size of the container and the percent moisture of the sample. The headspace in the container should be purged with nitrogen before capping tightly. Clear glass containers may be wrapped with an opaque material (e.g., clean aluminium foil) to eliminate light and reduce accidental breakage.

7.3.4 Transportation and Storage of Sediment Samples

The recommended procedures and conditions for the transportation and storage of sediment samples are as follows (Environment Canada 1994):

- The transport container should be refrigerated to $4 \pm 2^\circ\text{C}$ or contain ice or frozen gel packs that will keep the field samples below 7°C during transport to the laboratory.
- If field-collected samples are warm (e.g., $> 6^\circ\text{C}$), they should be cooled to between 1 and 6°C with ice prior to placement in the transport container.
- Samples should not freeze during transport.
- Ideally, a maximum/minimum thermometer or a continuous temperature recorder should be placed inside the transport container and the container sealed. Deviations in temperature should be reported.
- Light should be excluded from the transport container.
- All field-collected samples that require further processing before storage should be transported to the laboratory within 72 h, preferably within 24 h, of collection.

Where these conditions cannot be met due to operational constraints, the storage method and conditions adopted should strive to compromise the integrity of the sample as little as possible (Mudroch and MacKnight 1991).

Each sample container should be properly labelled and stabilized in an upright position in the transport container. Labelling of each sample container should include, at a minimum,

the site, station location or identification, the sample type, the method of collection, the name of the collector, and the date and time of collection.

7.3.5 Laboratory Test Sample Preparation

Sediment samples should be prepared in a well-ventilated area (e.g., fume hood) and the appropriate health and safety precautions should be followed. For first and second biological monitoring and magnitude and geographic extent, the preparation of samples under anoxic conditions is not a concern. However, for investigation of cause techniques such as toxicity testing, preparation of anoxic test sediments should be performed in a glove box in the presence of a controlled flow of an inert gas, if it is desirable to maintain these anoxic conditions. Below are some details on sediment preparation techniques used to allocate sediment to test containers:

Homogenizing: Mixing by hand or mechanical means may be used to achieve homogeneity of colour, texture and moisture; however, the efficacy of the method should be demonstrated, a priori, and the mixing time standardized to ensure consistency and minimize alterations in the size distribution of sediment particles.

Mixing of sediments should take place in the sample/storage container.

Partitioning: Coning or caking and quartering are the recommended techniques for partitioning the sediment for distribution among test containers. If a sediment splitter is used, its efficacy should be demonstrated and documented and it should be made of an appropriately inert material.

Drying: The recommended methods for drying sediment are oven-drying sediment sub-samples (1-5 g of wet sediment) at low temperatures (40-60°C) until a constant weight is reached or freeze-drying sediment subsamples.

Crushing/Grinding: Commercially available ball and pebble mills are recommended for fine-grinding small volumes of sediment (Mudroch and MacKnight 1991); however, it should be noted that grinding could change the chemistry of the material. Crushing can usually be achieved with a mortar and pestle.

Dewatering: Centrifugation with subsequent decanting of the supernatant is the recommended method for dewatering sediment samples. The centrifugation speed depends on the sample size and particle size (e.g., sediment weight or volume).

7.3.6 Prevention of Sediment Sample Contamination

When sediment samples are to be collected for chemical analysis, the procedures for the collection, handling, transportation and storage of samples are much the same as those outlined above. However, in such cases it is important that appropriate measures be taken to

ensure that sediment samples are not contaminated.

When sediment samples are to be collected for chemical analysis, sample collection devices should not be made of copper, zinc, brass or galvanized material, since collection devices made of unprotected metallic material can potentially affect the concentrations of metals in sediment samples. If this is not possible, when sub-sampling, the sediment that is in direct contact with the sides of the sampler should be excluded. The sub-sampled sediment should be transferred to clean containers made of inert material that will neither contaminate nor influence the characteristics of the sediment sample. The container should be tightly sealed and air should be excluded.

All sample containers should be pre-treated prior to receiving a field sample (Environment Canada 1983, 1989). New glass and most plastics should be pre-treated to remove residues, and/or leachable compounds, and to minimize potential sites of adsorption. Pre-treatment includes the following sequence of activities (adapted from Environment Canada [1989]):

- Scrub with phosphate-free detergent and hot water.
- Rinse with high-pressure hot water.
- Subject to a 72-h acid bath with 8 M HNO₃ (50 ml of HNO₃ per litre of water).
- Rinse four (4) times with hot water.
- Rinse three (3) times with DDW (double distilled water).
- Wash bottle caps (Teflon[®] or Teflon[®]-lined) with detergent and hot water, and rinse with DDW.

The acid bath will leach trace metals (e.g., Cu, Fe, Mo, Ni, Zn) from plastics. The triple rinse with distilled water is necessary because the acid treatment can activate adsorption sites on polymers which are then capable of binding trace metals in the field sample.

7.4 Sediment Variables

7.4.1 Determination of Particle Size Distribution

Where it is possible, the determination of sediment particle size distribution should be conducted each time that a benthic invertebrate community survey is conducted. Particle size should be determined for a minimum of one sample from each benthic sampling station.

Particle size determination is important in the interpretation of the results of chemical or biological analyses. Most importantly, from the point of view of using this data to aid in the interpretation of the results of benthic invertebrate community surveys, particle size has a significant impact on the structure of benthic invertebrate communities. It may also provide insight into the origin of sedimentary materials and about the dynamic conditions of sediment transport and deposition. From particle size analysis, specific surface, expressed as m²/g, can be determined, and with this, the adsorptive capacity of metals

and organic substances can be assessed.

Many different classifications of particle sizes exist; however, the following breakdown based on the Wentworth (1922) classification is recommended for the interpretation of EEM data.

<u>Classification</u>	<u>Particle Size (in mm)</u>
Gravel	16.0–2.0
Coarse Sand	2.0–0.2
Fine Sand	0.2–0.062
Silt	0.062–0.0039
Clay	< 0.0039

Procedures for methods of sediment particle size analysis can be found in ASTM (2003). Particle size analysis or grain size analysis is generally performed in two parts: sieve analysis and hydrometer analysis. The sieve analysis classifies particles greater than 0.06–0.075 mm in size (actual minimum size depends on the sieve set used). This is done by wet-sieving the sample through a set of at least four sieves, ranging in size from 0.06 mm to 16 mm. The material retained on the sieves is dried and weighed. Particles passing through the 0.06-mm sieve are collected and transferred to a 2-L container, together with the wash water. A hydrometer is used to determine the quantity of particles in this fraction from 0.06 mm down to 0.0014 mm. The data from these two tests are then tabulated and calculated to produce a particle size distribution curve. This curve graphically defines the percentage of material in the different fractions based on the total sample weight.

It is also possible to determine particle size distribution using laser diffraction, and this method is increasingly available. This method is more efficient and provides higher resolution results than the above methods. A laser diffraction instrument uses light from a low-power helium-neon laser (the analyzer beam). Particles from sediment samples enter the beam via a dispersion tank that pumps the material, carried in water, through a sample cell. The light scattered by the particles is incident onto the receiver lens, which focuses the scattered light onto a diode composed of numerous concentric rings. Through a process of constrained least squares fitting of theoretical scattering predictions to the observed data, the computer calculates a volume size distribution that would give rise to the observed scattering characteristics. No a priori information about the form of the size distribution is assumed, allowing for the characterization of multi-modal distributions.

The efficiency of laser diffraction is also a major benefit. A typical measurement takes only a few seconds, and the data are saved digitally and are instantly available for plotting and other calculations. Often, the entire distribution can be accounted for in a single measurement. Depending on the instrument used, a laser particle size analyzer can measure all sizes ranging from 0.05µm to 2000µm. For samples with a size range greater than 2000µm, sieve data can be merged with the laser results. Finally, the results using

laser diffraction are very high resolution, and are easily reproducible—overcoming a major shortcoming of the hydrometer and sieve methods.

7.4.2 Determination of Total Organic Carbon Content

As with particle size distribution, the amount of sediment total organic carbon (TOC) should be determined each time that a benthic invertebrate community survey is conducted. TOC should be determined for a minimum of one sample from each benthic sampling station.

Carbon is present in sediment in several organic forms such as humic matter; chemical, plant and animal matter; as well as inorganic carbonate forms. Organic carbon in sediment and the water column causes a decrease in dissolved oxygen by using up available oxygen, hence creating a more anoxic environment. Also, at certain pH levels, humic substances form complexes with metals, increasing metal solubility in the water column. Two methods are commonly used to analyze TOC in sediment. The elemental analyzer method, valid for samples of 0.5–25 mg, is based on the use of thermal conductivity. The oxidizing furnace method requires samples of 0.25–0.5 g and is based on the use of infrared spectrophotometry.

Elemental analyzer: Inorganic carbon is first eliminated by treatment with hydrochloric acid. TOC is then oxidized to carbon dioxide in the presence of a catalyst. The gas produced is separated by chromatography and quantified with a thermal conductivity detector.

Oxidizing furnace: Inorganic carbon is first eliminated by treatment with hydrochloric acid. TOC is then oxidized in the oxidizing furnace in the presence of manganese dioxide. The carbon dioxide formed from the organic carbon is measured directly by infrared absorption at the characteristic wavelength for carbon dioxide.

Procedures for these methods of analyzing TOC in sediment are described in US EPA (1986) and APHA (1995).

7.4.3 Determination of Total Metal Concentrations

The determination of total metal concentrations in sediments is not required as part of the EEM program. However, mines are encouraged to determine total metal concentrations in sediments when benthic invertebrate community surveys are conducted. Information regarding metals in sediment can be important to the interpretation of the results of benthic invertebrate community surveys, and to the design of subsequent surveys. If effects at a site are suggestive of eutrophication, consideration should also be given to determination of sediment nutrient concentrations.

Sediments are an integral component of aquatic ecosystems, and hence, a frequent aspect

of many environmental monitoring programs. They originate from the differential settling of both suspended terrigenous particles that have been introduced into aquatic ecosystems and precipitates that have resulted from chemical and biological processes within aquatic systems. Suspended particles entering the aquatic system may already contain contaminants. Alternately, non-contaminated particles suspended in water may accumulate soluble contaminants present in the waters of aquatic systems. Precipitation processes are also capable of scavenging contaminants. As a result, sediments can be viewed as either a reservoir or a sink for contaminants.

Contaminated sediments from point-source inputs such as mining effluents can become bioavailable and enter aquatic food-webs, therefore affecting the quality of the habitat. Measuring sediment quality helps identify which contaminants are entering the exposure area. Sediments provide a better integrator of average long-term environmental conditions than single-event water chemistry samples.

The selection of parameters for sediment chemistry analysis will be determined on a site-specific basis. Where historical data exist on sediment quality, they should be used in conjunction with effluent characterization and water quality data to help determine parameters to analyze.

Total, or bulk, sediment chemistry provides information on the loading rates of particular elements, and on depositional patterns. Techniques used for determination of metals in sediments include atomic absorption spectrophotometry (AAS), X-ray fluorescence (XRF), instrumental neutron activation analysis (INAA), inductively coupled atomic absorption spectrophotometry (ICP-AES) and ICP-mass spectrometry (ICP-MS). Because of the high concentrations of metals in sediment, particularly in mining areas, analytical techniques with higher detection limits (e.g., ICP, ICP-AES) are generally acceptable for sediment chemistry analysis.

Bulk sediment samples are digested using either aqua regia or a mixture of perchloric, nitric and hydrochloric acids for extraction of total metals. Metal concentrations can be influenced by sediment particle size and organic carbon content. Smaller particles and organic material have a higher affinity and more binding sites for metals than coarser grained material. Therefore, with all other factors being equal, total metal concentrations tend to be higher in fine organic substrates. To account for this influence, it is recommended that sediment samples be sieved through a 63- μm mesh screen, and that only the fraction less than 63 μm be analyzed.

Alternatively, metal concentrations in sediments can be normalized for particle size or organic content when comparing results between areas. Sediment metal data can be normalized to percent fines (silt + clay fractions) using the following equation (ESP 1996):

$$\text{Metal}_{\text{NF}} = \frac{\text{Metal}}{\text{Fines}}$$

where: Metal_{NF} = metal concentration normalized to fines
 Metal = reported metal concentration in sediment (mg/kg)
 Fines = proportion of fines in sediment

7.5 Sediment Toxicity Testing

Sediment toxicity testing is not a required element of EEM. There are many uses of sediment toxicity testing, including evaluating potential contamination in aquatic environments, verifying alterations seen in benthic invertebrate communities that may be due to toxicity of sediments and not other physical or biological factors, and possibly to interpret confounding factors (see below). Chapter 12 contains extensive guidance on sediment toxicity testing.

7.6 Confounding Factors

Sediment toxicity testing can help interpret situations where field effects are inconclusive due to confounding factors such as historical contamination or multiple dischargers to the same watercourse. Whole-sediment toxicity tests, conducted in the lab, typically use a standard overlying water, thereby isolating the effects of the sediment. Research underway by Environment Canada (Lisa Taylor, personal communication, Ecotoxicology and Wildlife Health Division, Environment Canada) is looking at the modifying effects of water chemistry parameters on sediment toxicity and whether it can be accounted for by using water collected from the study site as the overlying water. Site waters could include upstream receiving water, downstream receiving water (i.e., effluent mixed with receiving water), full-strength effluent, and/or a “clean” reference site water. The choice of overlying water should be decided based on the objectives of the study. Examples of study objectives include isolating effects due to historically contaminated sediments from those due to current effluent, determining whether the current effluent is modifying the bioavailability of contaminants within the sediment, or verifying whether the water upstream of a discharge point influences the toxicity of sediment collected below the discharge point. If the site water is confounded due to multiple dischargers to the same watercourse, it may also be more useful to use a standard overlying water that is simulated in the laboratory to match the field conditions. Parameters likely to ameliorate toxicity, such as pH, hardness, alkalinity or dissolved organic matter, can be adjusted in laboratory water to approximate the situation in the field.

7.7 References

[APHA] American Public Health Association. 1995. Standard methods for the examination of water and wastewater. 19th edition. Washington (DC): American Public Health Association.

[ASTM] American Society for Testing and Materials. 1992. E 1391-90, Standard guide for collection, storage, characterization and manipulation of sediments for toxicological testing. In 1992 Annual book of ASTM Standards, Vol. II.04, Section 11. Philadelphia (PA): American Society for Testing and Materials. p. 1134-1153.

[ASTM] American Society for Testing and Materials. 2003. D422-63, Standard test method for particle-size analysis of soils. In Annual book of ASTM Standards, Vol. 04.08. West Conshohocken (PA): American Society for Testing and Materials. p. 10-17.

Baudo R, Giesy JP, Muntau H, editors. 1990. Sediments: chemistry and toxicity of in-place pollutants. Chelsea (MI): Lewis Publishers, Inc.

Burton GA Jr, editor. 1992. Sediment toxicity assessment. Chelsea (MI): Lewis Publishers Inc.

de Groot AJ, Zschuppe KH. 1981. Contribution to the standardization of the methods of analysis for heavy metals in sediments. Rapp. P.-v. Reun. Cons. Int. Explor. Mer. 181:111-122.

[ESP] Ecological Services for Planning. 1996. Aquatic effects technology evaluation, 1996 field evaluation. Final survey report for Dome Mine, Ontario. Ottawa (ON): Prepared for Aquatic Effects Technology Evaluation Program, Natural Resources Canada.

Environment Canada. 1983. Sampling for water quality. Ottawa (ON): Environment Canada, Inland Waters Directorate, Water Quality Branch. xi + 55 pages.

Environment Canada. 1989. Bottle washing procedures. Burlington (ON): Environment Canada, National Water Research Institute, Inland Waters Directorate, National Water Quality Laboratory.

Environment Canada. 1994. Guidance document on collection and preparation of sediments for physiochemical characterization and biological testing. Ottawa (ON): Environment Canada. Environmental Protection Series Report EPS 1/RM/29.

Green RH. 1989. Power analysis and practical strategies for environmental monitoring. Environ Res 50:195-205.

Håkanson L, Jansson M. 1983. Principles of lake sedimentology. Berlin (DE): Springer-Verlag.

Holland PT, Hickey CW, Roper DS, Trower TM. 1993. Variability of organic contaminants in inter-tidal sandflat sediments from Manukau Harbour, New Zealand. Arch Environ Contam Toxicol 25:456-463.

Mudroch A, MacKnight SD. 1991. CRC handbook of techniques for aquatic sediments sampling. Boca Raton (FL): CRC Press.

Skei JM. 1992. A review of assessment and remediation strategies for hot spot sediments. Hydrobiologia 235/236:629-638.

Sly PG, Christie WJ. 1992. Factors influencing densities and distributions of *Pontoporeia hoyi* in Lake Ontario. Hydrobiologia 235/236:321-352.

[US EPA] United States Environmental Protection Agency. 1986. Test methods for evaluating solid waste - physical and chemical methods, SW-846. Washington (DC): United States Environmental Protection Agency.

Wentworth CK. 1922. A scale of grade and class terms for clastic sediments. J Geol 30:377-392.

Table of Contents

8.	Data Assessment and Interpretation	8-1
8.1	Overview.....	8-1
8.2	Understanding the Definition of Effect, and Meaning of Data Interpretation, within EEM.....	8-2
8.3	Data Assessment and Interpretation for the Fish Study.....	8-2
8.3.1	Preparing the Analyses	8-5
8.3.1.1	Immature Fish	8-5
8.3.2	Summary Statistics.....	8-6
8.3.3	Analysis of Variance (ANOVA) and Analysis of Covariance (ANCOVA)	8-6
8.3.3.1	ANOVA	8-7
8.3.3.1.1	Normality and Homogeneity of Variances	8-7
8.3.3.1.2	Independence (Pseudo-replication).....	8-8
8.3.3.2	ANCOVA	8-9
8.3.3.2.1	Analysis of Residuals.....	8-10
8.3.3.2.2	Independent Variable	8-10
8.3.3.2.3	Linear Regression	8-11
8.3.3.2.4	Slopes of the Regression Lines	8-11
8.3.3.2.5	Options for Non-parallel Regression Slopes	8-12
8.3.3.2.6	Non-parametric Alternatives to ANCOVA	8-13
8.3.4	Transformations	8-13
8.3.5	Level of Replication.....	8-14
8.3.6	Effect and Supporting Endpoints.....	8-14
8.3.6.1	Size-at-Age	8-14
8.3.6.2	Gonad Weight, Liver Weight, Condition and Fecundity	8-14
8.3.6.3	Mean Age.....	8-15
8.3.6.4	Age-at-Maturity	8-15
8.3.7	Statistical Analysis for Non-lethal Sampling.....	8-16
8.3.8	Data Quality Assurance / Quality Control and Analysis (Errors and Outliers).....	8-16
8.4	Effects on Usability of Fisheries Resources	8-18
8.4.1	Mercury in Fish Tissue	8-18
8.4.2	Physical Abnormalities	8-19
8.5	Data Assessment and Interpretation for the Benthic Invertebrate Community Study	8-20
8.5.1	Study Design and Statistical Procedures	8-20
8.5.2	Data Treatment.....	8-22
8.5.3	Reference Condition Approach	8-25
8.5.4	Supporting Endpoints	8-26
8.6	The Role of Power Analysis, α , β and Critical Effect Size in Determining	

Effects.....	8-27
8.6.1 Setting α and β	8-27
8.6.2 Power Analysis: Determination of Required Sample Size, Power and Appropriate Critical Effect Size	8-29
8.6.2.1 A Priori Power Calculations	8-30
8.6.2.2 Post Hoc Power Analyses	8-32
8.7 Critical Effect Sizes	8-33
8.8 Statistical Considerations for Mesocosm Studies	8-33
8.9 References.....	8-35
Appendix 1: Step-by-Step Guidance through Statistical Procedures	8-39
Appendix 2: Graphical and Tabular Representation of Data.....	8-59
Appendix 3: Case study – ANCOVA and Power Analysis for Fish Survey	8-75

List of Tables

Table 8-1: Required fish survey measurements, expected precision and summary statistics.....	8-3
Table 8-2: Fish survey effect indicators and endpoints for various study designs and the appropriate statistical analyses	8-4
Table 8-3: Supporting endpoints to be used for supporting analyses.....	8-4
Table 8-4: Summary of effect endpoints analyzed using ANCOVA	8-9
Table 8-5: Fish Tissue effect and supporting endpoints and statistical procedures	8-18
Table 8-6: Statistical procedure used to determine an effect for each of the seven study designs	8-20
Table 8-7: Sample sizes required to detect difference of ± 2 SD	8-29

8. Data Assessment and Interpretation

8.1 Overview

As part of the environmental effects monitoring (EEM) requirements under the *Metal Mining Effluent Regulations* (MMER), after biological monitoring studies are conducted, an interpretative report shall be prepared (MMER, Schedule 5, section 17). The owner or operator shall submit to the Authorization Officer reports of the results of the studies in writing. The role of the interpretative report within the EEM program is to summarize study results (including difficulties or confounding factors encountered), conduct applicable spatial analyses (and when sufficient data are available, temporal trend analyses), specify any identified “effects,” and make recommendations for subsequent EEM program monitoring. Data interpretation or the role of the report does *not* include determining the ecological, economic or social significance of results. The content of the interpretative report is available in Chapter 10 of this document and in the MMER.

The purpose of this chapter is to provide general guidance on how to assess and interpret EEM data, specifically:

- 1) which effect endpoints to use and report;
- 2) the statistical (or other) approach to use for each effect endpoint in order to determine the presence or absence of an effect; and
- 3) the role of power analysis, α , β and critical effect size (CES) in determining effects.

EEM involves iterative phases of monitoring and reporting. For each phase it is required to report the results of the data assessment made under Schedule 5, s. 16. The report must include the identification of any effects on fish populations, fish tissue or the benthic invertebrate community, the overall conclusions of the biological monitoring studies based on the results of the statistical analysis, and a summary of the results of previous monitoring. More specifically, the data generated for each mine should be analyzed to determine whether there are significant differences in certain effect indicators between reference and exposure areas or along an exposure gradient (i.e., determination of effect). In addition to the within-phase (spatial) analysis, a comparison of effects between phases (temporal comparisons) is recommended in order to determine whether any effects identified previously are lessening or worsening.

For EEM purposes, only specified data (the effect indicators) generated from the fish survey, benthic invertebrate community survey and fish usability studies are used to assess the presence of effects. Other EEM data are only used to help interpret effects on fish and benthos (e.g., effluent characterization and water quality monitoring) or to help characterize any changes in effluent quality over time (e.g., sublethal toxicity testing). The tables in the following sections summarize the recommended data analysis procedures for the effect indicators for each monitoring requirement (Tables 8-2, 8-3 and

section 8.5). Also, refer to the relevant sections of this chapter for further details. Many of the data interpretation issues are the same for the fish survey, fish usability and benthic invertebrate community sections that follow (e.g., assumptions and interpretation of statistical techniques common to more than one of these sections). Several of these common issues are discussed in the fish section below, and are not repeated in the following sections on fish usability and benthic invertebrate communities.

8.2 Understanding the Definition of Effect, and Meaning of Data Interpretation, within EEM

Understanding 1) the types of data analyses that are relevant and 2) what is meant by the definition of “interpretation” is integral to the EEM program, particularly when writing an interpretation report. In order to address both issues, it is important to define “effect.”

Within EEM, an effect is defined generally as a statistically significant difference in fish or benthic invertebrate community effect indicators measured between an area exposed to effluent and a reference area, or a statistically significant difference in these effect indicators within an exposure area along a gradient of effluent concentrations. For fish tissue analysis (which is conducted to determine the usability of fisheries resources), an effect is defined as measurements of concentrations of total mercury that exceed 0.5 µg/g wet weight in fish tissue taken in an exposure area and that are statistically different from and higher than the measurements of concentrations of total mercury in fish tissue taken in a reference area (Schedule 5, Interpretation, s.1).. In cases where it is not feasible to examine wild fish or field distribution of benthic invertebrates in areas exposed to effluent and reference areas, an alternative monitoring approach for fish or fish habitat may be used to determine if the effluent is causing an effect (Chapter 9).

Given the above definition of effect, it is important to recognize that not all effects identified in EEM represent damage to fish, fish habitat or the usability of fisheries resources. However, effects as defined above do represent scientifically defensible differences or gradients that may reflect changes to the ecosystem associated with the effluent. As a result, detailed information on the effects, including the magnitude, geographic extent and possible cause of the effect, may contribute to the understanding of the ecosystem and could be used in the management of the aquatic resources.

8.3 Data Assessment and Interpretation for the Fish Study

The data collected during the fish population study will include indicators of growth, reproduction, condition and survival (when it is possible to obtain data to establish the indicators), that include the length, total body weight and age of the fish, the weight of its liver or hepatopancreas, and, if the fish are sexually mature, the egg weight, fecundity and gonad weight of the fish (MMER Schedule 5, s. 16).

The overall procedure that should be followed and reported can be divided into the following stages: 1) preparing the analyses, 2) initial summary statistics, 3) analysis of variance (ANOVA) analyses, 4) analysis of covariance (ANCOVA) analyses, and 5) power analyses. Appendix 1 provides a step-by step-guidance through the statistical procedures for the fish survey.

The required fish survey measurements, expected precision, and summary statistics are described in Table 8-1. Table 8-2 outlines the effect indicators for various study designs and the appropriate statistical analyses that are applicable for the fish population study. Table 8-3 outlines the supporting endpoints.

Table 8-1: Required fish survey measurements, expected precision and summary statistics

Measurement Requirement (MMER Schedule 5, Part 2, s. 16)	Expected Precision***	Reporting of Summary Statistics (MMER Schedule 5, Part 2, s. 16) and other general reporting
Length (fork or total or standard)*	+/- 1 millimetres (mm)	Mean, median, standard deviation (SD), standard error, minimum and maximum values for sampling areas
Total body weight (fresh)	+/- 1.0%	Mean, median, SD, standard error (SE), minimum and maximum values for sampling areas
Age	+/- 1 year (10% to be independently confirmed)	Mean, median, SD, SE, minimum and maximum values for sampling areas
Gonad weight (if fish are sexually mature)	+/- 0.1 grams (g) for large-bodied fish species and 0.001 g for small-bodied fish species	Mean, median, SD, SE, minimum and maximum values for sampling areas
Egg size(if fish are sexually mature)	+/- 0.001 g	Weight, (recommended minimum sub-sample sizes of 100 eggs), mean, SE, minimum and maximum values for sampling areas
Fecundity** (if fish are sexually mature)	+/- 1.0%	Total number of eggs per female, SE, minimum and maximum values for sampling areas
Weight of liver or hepatopancreas	+/- 0.1 g for large-bodied fish species and 0.001 g for small-bodied fish species	Mean, median, SD, standard error, minimum and maximum values for sampling areas
External condition	n/a	Presence of any lesions, tumours, parasites or other abnormalities
Sex	n/a	

* If caudal fin forked, use fork length (from the anterior-most part to the fork of the tail). Otherwise, use total length, and report type of length measurement conducted for each species. In cases where fin erosion is prevalent, standard length should be used.

** Fecundity can be calculated by dividing total ovary weight by weight of individual eggs. Individual egg

weight can be estimated by counting the number of eggs in a sub-sample. The sub-sample should contain at least 100 eggs.

*** For small-size fish weights, use at least a 3-decimal scale.

Table 8-2: Fish survey effect indicators and endpoints for various study designs and the appropriate statistical analyses

Effect Indicator	Effect Endpoint and Statistical Procedure		
	<i>Standard Survey</i>	<i>Non-lethal Sampling</i>	<i>Wild Molluscs</i>
Growth (Energy Use)	Size at age (body weight against age) (ANCOVA)	Size (length and weight) of young of the year (age 0+) at end of growth period (ANOVA)	Whole-animal wet weight (ANOVA)
Reproduction (Energy Use)	Relative gonad size (gonad weight against body weight) (ANCOVA)	Relative abundance of YOY (% composition of YOY) (See Chapter 3, section 3.4.2.2)	Relative gonad size (gonad weight against body weight) (ANCOVA)
Condition (Energy Storage)	Body weight relative to length Relative liver weight (liver weight against body weight) (ANCOVA)	Body weight relative to length (ANCOVA)	Whole-animal dry weight, dry shell or soft tissue weight related to shell length (ANCOVA)
Survival	Age (ANOVA)	Length frequency distribution (2-sample Kolmogorov-Smirnov test)	Length frequency analysis (2-sample Kolmogorov-Smirnov test)

Table 8-3: Supporting endpoints to be used for supporting analyses¹.

Effect Indicator	Supporting Endpoint	Statistical Procedure
Energy Use	Body weight (whole)	ANOVA
	Length	ANOVA
	Size-at-age (length against age)	ANCOVA
	Relative gonad size (gonad weight against length)	ANCOVA
	Relative fecundity (# of eggs/female against body weight)	ANCOVA
	Relative fecundity (# of eggs/female against length)	ANCOVA
	Relative fecundity (# of eggs/female against age)	ANCOVA
	YOY survival	See Chapter 3, section 3.4.2.2
Energy Storage	Relative liver size (liver weight against length)	ANCOVA
	Relative egg size (mean egg weight against body weight)	ANCOVA
	Relative egg size (mean egg weight against age)	ANCOVA

Note: these analyses are for informational purposes, and significant differences between exposure and reference areas are not necessarily used to designate an effect.

¹ For the ANCOVA analyses, the first term in parentheses is the endpoint (dependent variable, Y) that is analyzed for an effluent effect. The second term in parentheses is the covariate, X (age, weight or length).

8.3.1 Preparing the Analyses

Upon completion of the field and laboratory measurements, the data should be promptly entered into a computer spreadsheet and quality assurance / quality control (QA/QC) should be conducted. Values entered into the spreadsheet should be double-checked with the original handwritten data sheet to prevent typographical errors. A data matrix with the location identifier (area), variables in columns, and observations in rows operates as the fundamental working unit. In this spreadsheet, include a column for comments on the physical condition and any abnormalities noticed during the sampling process. These comments may prove to be useful in identifying unusual observations and help to determine whether data should be removed from an analysis. A location identifier for area or site should be chosen—one that can be easily distinguished as reference or exposure. This will allow for easier interpretation for others who are not familiar with the location identifier codes. If an insufficient number of fish were collected at an exposure site but were collected at the reference site, be sure to make special note of this.

Failure to identify transcription errors can invalidate further analyses. Assuming the data have been entered correctly, data that will be necessary for interpretation should be summarized, screened for erroneous values and outliers, and assessed for normality and transformed if necessary; and, any significant confounding factors should be summarized.

Differences between sexes in growth rate, body weight, condition factor, gonad size and liver size are common, due to differences in overall energetic requirements between male and female fish. Therefore, for all parameters, sexes should initially be treated separately when conducting the analyses. In addition, sexually immature fish should not be mixed with sexually mature fish for analyses.

8.3.1.1 Immature Fish

It should be confirmed that all fish which are assumed to be adults are undergoing gonadal development for the next spawning season. The inclusion of immature fish into statistical analyses can provide misleading results. Immature fish devote proportionally more energy toward growth, so the body size-gonad relationship for immature fish is different than that of adult fish. For data analysis, fish identified as immature in the spreadsheet should be removed. The gonadosomatic index (GSI) = $\text{gonad weight} / \text{body weight} \times 100$ can be useful in identifying immature fish. As a general rule, for many fish species, immature fish can be categorized as having a GSI of $< 1\%$, although there are some notable exceptions, such as guarding species like the Brown Bullhead. A plot of gonad weight vs. body weight, and using this general rule for GSI, can be most useful in identifying immature fish. Comments from the field observations may also assist in identifying unusual observations that are suspected to be immature (e.g., comments such as “weighed only one testis”). The sampling period has to be adjusted to the biology (life history) of the species to avoid capturing fish prior to gonadal development for the

upcoming reproductive season. However, when non-lethal sampling is to be carried out and age-frequency distributions are used to assess reproductive success, the timing of sampling is less important. Data analysis on immature and mature fish should be conducted separately, except, for obvious reasons, when comparing the proportion of non-spawning fish among sites.

8.3.2 Summary Statistics

The descriptive statistics (mean, median, standard deviation [SD], standard error [SE]) and the minimum and maximum values will be determined, when it is possible to obtain data, to establish the indicators of growth, reproduction, condition and survival that include the length, total body weight and age of the fish, the weight of its liver or hepatopancreas, and, if the fish are sexually mature, the egg size, fecundity and gonad weight of the fish (MMER Schedule 5, s. 16). The fish survey measurements to determine effects in fish growth, reproduction, condition and survival, the expected precision, and summary statistics are described in Chapter 3.

The summary statistics should be calculated by species and sex for each area being summarized (e.g., reference area and exposure area). Before calculating summary statistics, the data should be graphed using box plots for examination of extreme outliers. The summary statistics should be presented in graphical and tabular format for all variables. The data should be examined for normality and equality of variances (basic statistical assumptions). Note that slopes and adjusted means and associated error terms should also be reported for ANCOVA, as outlined below.

Visual screening techniques such as box and whisker plots, normal probability plots, and stem-and-leaf diagrams can be used to identify extreme values (true outliers and/or data entry errors). Most statistical software packages provide data summary modules capable of generating appropriate summary statistics and graphics. These summary statistics are usually needed for presentation, and aberrantly high or low values can indicate errors. Extreme values or outliers should not be removed from the data set (unless they are obvious sampling, measurement or data entry errors) (Grubbs 1969; Green 1979), because mistakenly removing valid data will result in the loss of statistical power in the fish survey. Instead, extreme values should be identified in the report and the influence of the extreme value(s) on the results should be determined by reanalyzing the data without the extreme value.

8.3.3 Analysis of Variance (ANOVA) and Analysis of Covariance (ANCOVA)

In addition to descriptive statistics, an analysis of the results must be conducted to determine if there is a statistical difference between the sampling areas (MMER Schedule 5, s. 16(c)). This is usually conducted using ANOVA or ANCOVA. However, in some instances, other statistical procedures (e.g., non-parametric methods) may be used. The

analyses (for ANOVA and ANCOVA) that are used to determine whether statistically significant effects have occurred should follow these three steps of data inspection, analysis and interpretation (Appendix 1 provides a step-by-step guidance through the statistical procedures for the fish survey):

- 1) The data should be inspected to see whether they satisfy the assumptions of ANOVA or ANCOVA. These procedures are robust enough to allow for moderate violations of some assumptions and, in some cases, data transformation will help to remedy departures from the assumptions. In cases where data transformations do not sufficiently rectify departures from the assumptions, it may be necessary to use non-parametric procedures, in which case the methods of power analysis discussed in section 8.6 would not apply. These issues are further discussed below, and the standard statistical texts (e.g., Sokal and Rohlf 1995) should be consulted for a more complete discussion.
- 2) Following inspection of the data and any necessary transformations, the actual statistical comparisons are carried out.
- 3) After the statistical comparisons are made, key results for the effect indicators (Table 8-1) should be presented in a clear fashion so as to indicate whether there has been effects and, if so, the nature of the effects (including the direction and magnitude of the effects). An effect is declared if the p-value is less than the a priori α value determined, as outlined in section 8.6.

8.3.3.1 ANOVA

ANOVA is used to test for site differences in length, weight and age. The assumptions for ANOVA are that:

- the data for reference and exposure populations are normally distributed;
- the variances are equal between the reference and exposure populations; and
- the error terms are independently distributed.

A one-factor ANOVA is used to test for differences in the mean response (length, weight or age) using the factor site (e.g., reference or exposure). A residual plot can be useful in identifying outliers. Observations with studentized residuals with a magnitude greater than 4 typically warrant investigation. Non-parametric alternatives for ANOVA include the Kruskal-Wallis test, or, if comparing two sites, the Mann-Whitney test (non-parametric alternative to the two-sample T-test).

8.3.3.1.1 Normality and Homogeneity of Variances

The assumptions of normality and homogeneity of variance should be assessed before applying most parametric procedures. However, most univariate normal distribution-

based statistical methods are quite robust and can support moderate violations of the assumptions. Transformation of original data will help normalize the data or homogenize the variances. Logarithmic transformations are often preferred because most biological measures are considered to operate on a log or exponential scale (Peters 1983) and such a transformation is biologically meaningful. It should be noted that for the purposes of the fish EEM survey, 1 should not be added to values before logging because it has undesirable effects on the calculated variances when changing measurement units. If the transformations are unable to produce data that meet the assumptions, a plot of the residuals may reveal problematic data points that may warrant investigation. Most of the univariate statistical methods are robust under moderate violations of assumptions, with some exceptions such as analyses with small and unequal samples. For serious violations, non-parametric statistics can be considered.

8.3.3.1.2 Independence (Pseudo-replication)

When designing experiments, it is desirable to ensure that replicates are randomly allocated to different treatment levels, such that the responses of each replicate are independent of other replicates. This element of randomness provides some assurance that observed differences in responses among treatments results from treatment effects and not from other factors.

Lack of independence can occur when, for example, one person collects all the data from the exposure area while another person collects data from the reference area. This can bias the data if the two individuals consistently use slightly different sampling or sorting protocols. Generally, these kinds of problems can only be remedied by changing the method of conducting the sampling so as to remove the sources of bias.

Randomly allocating replicates to different treatment levels is a relatively easy procedure when conducting manipulative experiments (e.g., controlled laboratory tests), but is less obvious for observational field studies. Observational studies, such as environmental impact studies (e.g., single-stressor EEM studies) or environmental assessments (i.e., multiple stressors), test hypotheses about the presence and magnitude of effects. However, the strength of inferences from these types of experiments is limited, for two reasons (Paine et al. 1998):

- the stressor (e.g., mine outfall, hydroelectric dam) cannot be reproduced; and
- stressors cannot be applied randomly to replicates.

What this means is that the stressor or treatment is always partly or wholly confounded with space or time, and that the observed effects may or may not be caused by the stressor of interest. For example, when investigating whether effluent from an industrial plant is having an effect on downstream fish populations, it is not possible to replicate the treatment of effluent exposure (i.e., there is only one plant and outfall), or to randomly assign fish populations to the different treatment levels (reference vs. exposed). As such, when significant differences are observed between reference and exposed fish

populations, one can conclude that there are differences between these two populations, but not necessarily that the differences were caused by effluent exposure. Interpreting significant differences as treatment effects when either treatment is not replicated or replicates are not independent is referred to as pseudo-replication (Hurlbert 1984).

Before attributing cause to any specific stressor, it is critical that observations be confirmed, through replication over time, and that some effort be expended to confirm that the stressors of interest are involved in the responses.

8.3.3.2 ANCOVA

ANCOVA is used to test for site differences in condition, relative gonad weight, relative liver weight, weight-at-age, size-at-age, and relative fecundity. A summary of these analyses is provided below.

Table 8-4: Summary of effect endpoints analyzed using ANCOVA

Effect Endpoint	Response Variable	Covariate
Condition	Body weight	Length
Relative liver weight	Liver weight	Body weight
Relative gonad weight	Gonad weight	Body weight
Weight-at-age	Body weight	Age
Size-at-age	Length	Age
Relative fecundity	Eggs/female	Body weight

The assumptions for ANCOVA are that:

- the relationship between the response and covariate is linear;
- the slopes of regression lines among sites are parallel;
- the covariate is fixed and measured without error; and
- the residuals are normally and independently distributed with zero mean and a common variance.

It should be noted that ANCOVA is basically a two-step procedure consisting of:

- a) determining whether the slopes are approximately parallel; and
- b) if the slopes are parallel, going on to determine whether the elevations of the regressions are significantly different. This procedure is discussed more fully below.

ANCOVA is used to test for differences in a response among sites while taking into account the variability in test subjects by including a covariate in the analysis. This inclusion of a covariate in the analysis decreases the error term (by accounting for the variability explained by the regression of the response variable on the covariate) and thus

increases the power of the test (Huitema 1980).

It has been suggested that the range of the independent variable (covariate) should be approximately the same for each site. This will be difficult to assure in practice, but the violation of this should be considered when interpreting results from such cases. If there is reason to believe that there are issues with the overlap of the range of covariate values, perform a single-factor ANOVA on the covariate values between sites. If the covariate means do not significantly differ between sites, the results of the ANCOVA will probably be reliable (Quinn and Keough 2002). A significant difference in the mean covariate values between sites is a significant effect. In interpreting differences in the covariate means or ranges observed, take into consideration the consistency of sampling gear between sampling sites and the selection of samples. It may be appropriate to provide an analysis of a subset of the data, omitting unusually high or low covariate values in order to provide a reliable analysis.

The range of covariate values for the weight-at-age effect endpoint must be considered before performing an ANCOVA. For several small-bodied fish species, the range of the covariate (age) might only be between 2 and 3 or 2 and 4. An ANCOVA with only 2 or 3 values of the covariate can provide misleading results. In these cases it may be appropriate to perform a one-factor ANOVA on body weight, using site as the factor for each age group.

8.3.3.2.1 *Analysis of Residuals*

The preferred method of examining the residuals is to use graphical methods rather than relying on formal tests to assess normality and equality of variance. In fact, Day and Quinn (1989) have recommended against using formal tests. A good discussion of this topic can be found in Miller (1986). Draper and Smith (1981) review various methods of examining residuals, particularly residuals from regressions. Most statistical software packages also provide modules for examination of residuals. These methods are usually graphical, although diagnostic statistics are available as well. The primary advantage of these methods, compared to formal tests, is that they can identify the cause of violations of normality or equality of variances.

8.3.3.2.2 *Independent Variable*

The assumption that the independent variable is fixed is frequently violated, and Draper and Smith (1981) discuss the consequences of this violation. A non-fixed independent variable is likely to prove problematic, mainly in situations where the range of the independent variable is very small, i.e., when the range in size (or age) of the fish included in the regression is very small. In this case (very narrow size or age range), there is little to be gained by using ANCOVA with size or age as a covariate, and the data would be better analyzed as a simple ANOVA comparison of the exposure to reference area (i.e., no need to factor out the influence of the covariate).

8.3.3.2.3 *Linear Regression*

The assumption of a linear relationship can be tested for samples with multiple observations at different values of the independent variable. This may be possible for discrete variables such as age, but not for continuous independent variables such as body weight. At a minimum, linearity should be verified by visual inspection. Linearity can often be improved by transformation (e.g., the log-log transformation is used very widely for this purpose for the EEM fish ANCOVA analyses). The regression plots should also be inspected to ensure that the slopes are not unduly influenced by outliers. Scatter plots help identify outliers and unusual data. For example, when reproductive data are analyzed for fish, the plots aid in identifying potential “immature” fish that could affect the results. The scatter plots should be included in the interpretative report.

8.3.3.2.4 *Slopes of the Regression Lines*

A key assumption of ANCOVA is that the slopes of the regression lines for the reference vs. exposure areas are approximately equal. Therefore, the first part of an ANCOVA analysis is to test for differences in slopes between areas. A significant interaction term in the ANCOVA for covariate X vs. area (e.g., age*area or size*area) indicates significantly different slopes. In cases where the slopes are not significantly different (i.e., interaction term not significant), this indicates that the regression lines are approximately parallel to each other. Using the weight-at-age ANCOVA as an example, parallel slopes would indicate that weight gain over age is similar for both areas. The next step in this example is to proceed with the ANCOVA model, and test for differences in adjusted means (elevation) to investigate whether fish are proportionately heavier at any age in one area than in another.

It is possible that the slopes of regressions may differ. For example, fish from the reference area may be gaining weight more rapidly with increasing age (steeper slope) than fish from the exposure area. If the slopes of the regressions are significantly different, the ANCOVA cannot be completed. In this case, using the weight-at-age example, the effect would not be a proportional difference in weight at any age; rather, the rate of weight gain with increasing age would be significantly different among areas. This is considered a statistically significant EEM effect for the fish survey. That is, an effect would be determined as a significant difference in slope among areas rather than a significant difference in elevation. For this situation, it is also a good idea to plot separate regression lines to obtain a better qualitative understanding of the weight-at-age relationship for each area over the entire data range of the X covariate (e.g., where do the lines intersect?). It should be noted that, even when the slopes of the regressions significantly differ among areas, it is still possible to make further comparisons over a particular range of values for the X covariate (i.e., a particular age or size range) (Sokal and Rohlf 1995). This kind of comparison would be appropriate if it is judged that that particular age or size range is of particular concern.

It is also preferable that the range of the independent variable be approximately the same for each “treatment” (i.e., area). This may be difficult to assure in practice, but any violation of this should be considered when interpreting results from such cases. For example, if the size range used as the X covariate for the reference area does not show much overlap with the size range for the exposure area, use of the ANCOVA results requires the assumption that the regression slopes would still be parallel for overlapping size ranges and may not be appropriate in this situation.

8.3.3.2.5 Options for Non-parallel Regression Slopes

When the assumption of parallel regression slopes is not met, ANCOVA cannot proceed, because adjusted treatment means cannot be correctly interpreted. In this case there is a covariate by treatment interaction, and differences in the response variable among treatments vary at different values of the covariate. There are a few options for dealing with non-parallel regression slopes in ANCOVA. These are discussed below in the order that the methods should be applied to data sets with non-parallel slopes. The first two options provide mechanisms by which the slopes can be treated as being parallel, thus allowing a full ANCOVA and comparison of adjusted means. The third option provides an alternative methodology for calculating measured effects when the slopes cannot be treated as being parallel, even after applying options 1 and 2.

1. Influential Points (from Barrett et al. 2010)

Influential points are observations with high leverage (outliers in the covariate space) that have the potential to dominate conclusions by producing substantial influence on the regression coefficients (Fox 1997). If one or more points is highly influencing the slope of a regression line and causing non-parallel slopes, removal of this (these) point(s) may remove the evidence against fitting the data to the parallel model. Influence can be assessed using the Cook’s distance statistic (Cook 1977, 1979), which is incorporated into many statistical software packages. It is calculated using studentized residuals (outliers in the response variable) and a measure of leverage called “hat values” (outliers in the predictor variable) as a measure of impact for each observation (Fox 1997). A plot of Cook’s distance vs. the covariate is most useful in identifying high-influence observations. A numerical cut-off of $4 / (n - k - 1)$, where n is the total number of observations and k is the number of predictors in the regression model, can also be used to assess high-influence observations (Fox 1997).

2. Coefficients of Determination (from Barrett et al. 2010)

The coefficient of determination (R^2) expresses the proportion of the total variability in the response variable that is explained by its linear relationship with the independent variable, and is a measure of the association between the two variables (Quinn and Keough 2002). When the regression slopes are found to be non-parallel, the R^2 of the full regression model (model with the interaction term included) can be compared to the R^2 of the reduced regression model (model with the interaction term removed). When the R^2 of

the parallel (reduced) model is high (greater than 0.8) and only slightly (less than 0.02) lower than that of the full model, the parallel model can provide a sufficient representation of the data and can be used to proceed with the analysis.

3. Estimating Effects for Different-sized Fish (from Lowell and Kilgour 2008)

When the above two methods cannot be applied to the data set (i.e., when the slopes remain non-parallel even after applying the above two methods), the following method can be used to estimate measured effects for smaller (or younger) and larger (or older) fish. First determine the minimum and maximum values of the covariate within the range of covariate overlap for the two regressions (reference and exposure areas). Then, determine the predicted values of the response variable for each area regression line at these two covariate values (minimum and maximum). An estimate for the effect at the minimum covariate value (i.e., the effect on smaller or younger fish) will be the difference in predicted values, calculated as exposure-predicted value minus reference-predicted value, expressed as a percentage of the reference-predicted value. If the data were log-transformed, the predicted values must be anti-logged (i.e., x expressed as 10^x) before calculating the percent difference. The calculation is the same for larger (or older) fish, but using the maximum value of the covariate where the ranges for each area overlap. Each of these two measured effects (percent differences for small/young fish and large/old fish) can then be compared to CESs in the same way as is done for measured effects calculated from means (from ANOVA) or adjusted means (from ANCOVA).

8.3.3.2.6 *Non-parametric Alternatives to ANCOVA*

ANCOVA is robust to violations of the assumptions of the test when sample sizes are approximately equal (Huitema 1980; Hamilton 1977). When assumptions are seriously violated and sample sizes are unequal, non-parametric alternatives to ANCOVA could be considered. Several different non-parametric techniques using ranks have been proposed. Iman and Conover (1982) proposed a non-parametric alternative in which the response and covariate are replaced by their ranks. The analysis is the same as the parametric ANCOVA using the ranks as data, and is the simplest non-parametric alternative. Groups of tied ranks are replaced by the average rank for that grouping. Some other non-parametric alternatives are discussed in Shirley (1981) and Quade (1967).

8.3.4 Transformations

Transformations of the data can often help improve normality and homogenize variances (reduce some violations), and an examination of the relationship between the means and variances can help identify the most appropriate transformation (see Green 1979). Taylor's Power Law (Taylor 1961), which examines the relationship between treatment means and variances, can be used to determine the specific transformations in order to normalize data or homogenize variances (Green 1979). Logarithmic transformations are

often preferred because biological measures are frequently considered to operate on a logarithmic or exponential scale (Peters 1983). It should be noted that 1 should not be added to values before logging for the purposes of the fish EEM survey, because it has undesirable effects on the calculated variances when changing measurement units. If the transformations are unable to produce data that approximately meet the assumptions, it may be necessary to use non-parametric statistics.

8.3.5 Level of Replication

For each of the ANOVA and ANCOVA analyses, the level of replication (sample size, n) is the number of individual fish. The minimum sample size recommended is 20 sexually mature fish per sex (and an additional 20 sexually immature fish if small-bodied fish species are being sampled) for each of the 2 sentinel fish species in both the reference and exposure area. A power analysis should be conducted to determine sample size if the appropriate data are available.

8.3.6 Effect and Supporting Endpoints

8.3.6.1 Size-at-Age

Rates of growth are commonly described by the relationship of size (as weight or length) to age. Over the entire lifespan of a fish, this relationship is curvilinear, with the rate of increase declining as fish approach the limit of their lifespan (Ricker 1975). As only adult fish are often sampled, classical growth rates cannot be calculated. Nevertheless, for the purposes of the EEM program, fish growth can be inferred from size-at-age estimates determined for each area using ANCOVA. This calculation assumes that the relationship between size and age for adult fish is approximately log-linear (log size vs. log age) (Bartlett et al. 1984).

Size-at-age may be estimated by calculating the regression relationship between body size (weight or length) and age for each sampling area (reference and exposure). It is recommended that both length and weight be used to calculate size-at-age, in order to determine which provides the best fit and tightest regression.

8.3.6.2 Gonad Weight, Liver Weight, Condition and Fecundity

Relative gonad and liver size (and fecundity) are obtained by regression and analyzed using ANCOVA, using body weight as the covariate. Likewise, condition is obtained by regressing body weight against body length, and essentially describes how “fat” fish are at each area.

A variety of indices have been used in fisheries biology to describe the condition of fish (Bolger and Connolly 1989). Calculating the ratio of one variable to another has been used to derive many of them. Examples of a few common indices are):

- condition factor (k) = $100 (\text{body weight}/\text{length}^3)$;
- GSI = $100 (\text{gonad weight}/\text{body weight})$; and
- liver somatic index (LSI) = $100 (\text{liver weight} / \text{body weight})$.

In general, however, investigators have become cautious about using derived variables and ratios because they may have undesirable statistical properties (Green 1979; Jackson et al. 1990). Although these indices may be used for presentation purposes, it is preferable statistically to estimate (and analyze) the parameters from regressions of original variables (i.e., ANCOVA) rather than from ratios (Gibbons et al. 1993).

8.3.6.3 Mean Age

Calculation of mean age is meant as a gross reflection of the age distribution of adult fish collected from each area. Variability in mean age of fish can be estimated using ANOVA. The mean square error from the model is the best estimate of variability. Site difference in length and weight can also be analyzed in this fashion. It is essential that the sampling gear be consistent between the sampling areas, because most sampling methods select for certain age classes.

8.3.6.4 Age-at-Maturity

Age-at-maturity is a commonly used parameter in fisheries biology. However, few methods of calculation incorporate a measure of statistical confidence or variability. Therefore, it is recommended that age-at-maturity be estimated by traditional probit analysis, as is commonly used for determining median lethal concentration (LC_{50}) in toxicity tests. By determining the proportion (%) of mature individuals in each adult age class, and converting these data to probits (or plotting the data on probit paper), a straight-line relationship is generated (probit vs. log age) that allows one to estimate the age where 50% of the fish sampled are sexually mature. An estimate of variability in age-at-maturity among individual fish can be obtained from the slope of the line. The slope estimates $1/SD$. Therefore, the SD is estimated by $1/\text{slope}$. Using data collected over several phases, confidence limits can be calculated as an estimate of precision and statistical comparison of area values. Most statistical software packages can convert percentages to probits, and several small, independent packages are designed to conduct LC_{50} /probit analysis and generate the confidence limits. For more detailed information on conducting probit analysis, refer to Hubert (1980). For a discussion of factors to be considered when using probit analysis and other techniques for estimating age-at-maturity, refer to Trippel and Harvey (1991).

8.3.7 Statistical Analysis for Non-lethal Sampling

For non-lethal sampling, length-frequency distributions should be compared using a 2-sample Kolmogorov-Smirnov test. Gray et al. (2002) analyzed young-of-the-year fish separately, in order to assess age-specific variability in growth rates.

The Kolmogorov-Smirnov test is a robust analysis to determine if two data sets differ significantly, and can be used to look at relative distributions of data. This is a non-parametric, distribution-free test that assesses the similarity of two cumulative distribution functions of two data sets (Sokal and Rohlf 1995):

$$H_0: F(X) = F(Y); H_1: F(X) \neq F(Y)$$

Differences are considered significant at $p < 0.05$.

ANOVAs can be performed on length and weight. Data may need to be transformed. If appropriate, a post hoc analysis of differences between sites can be conducted using the Tukey Honestly Significant Difference test.

ANCOVAs should be performed for size-at-age (if possible) and condition factor (length vs. weight by site). The analyses should examine whether there were significant regressions, and if there was a significant interaction between areas. If slopes were equal, the data should be examined for a difference between areas, which area had the greatest values, what is the percentage area difference, and what was the p for slope or adjusted mean differences. If there is an interaction, the data should be plotted to see if the data are interpretable.

8.3.8 Data Quality Assurance / Quality Control and Analysis (Errors and Outliers)

Guidance on QA/QC for data analysis is provided below. The importance of ensuring data quality cannot be overemphasized. Each applicable chapter provides further guidance on QA/QC for study design, consistency of methods and measurements, and definitions of protocols and procedures.

There are various types of common entry errors, including data entry errors, entering the wrong species, missing or moved decimal places, and wrong sex or stage of maturity. It is critical to examine the data for errors and outliers prior to initiating analysis of data. Entry errors, transcription errors and invalid data are impossible to detect in final reports.

Data that have been entered incorrectly can sometimes be easily detected using scatter plots of length vs. weight, weight vs. gonad weight and weight vs. liver weight to look for points that are obviously different. Data entry errors are relatively easy to correct and can be re-entered. If the error cannot be reconciled because of obvious errors or

omissions in the original data sheet, the fish (data point) should be removed from the data set.

Errors and extreme observations inflate the variance and reduce the power to detect significant differences in the data set. Evaluation of outliers includes consideration of the raw data, the field conditions, and the data collection process. Data points that are different, but are not due to entry errors, can arise for a number of reasons. For example, fish may appear sick or damaged, the fish may be an outlier for no apparent reason, or the outlier may represent an important phenomenon that is part of the response to the stressors under study.

In the first case, there can be a small number of fish that are obviously sick or were damaged (in a manner unrelated to the stressors under consideration) and should not be considered part of the data set for interpretation. These usually appear as single points that are separate from the main data set. Examples of these include fish that are missing their tail due to predation wounds, fish that have a jaw deformity or injury that has affected their feeding, or fish that are blinded through injury and are thinner than other fish. In these cases, the fish should be removed from the comparison.

If there is no obvious reason for the presence of rare outliers, the analysis should be conducted with and without the suspect observation, to determine how much influence it has on the conclusions. If it has an impact on whether a relationship is significant or not, statistics textbooks should be consulted for advice on how to evaluate whether the measurement can be removed.

In the third case, there can be several fish that are obviously different but possibly part of the relationship being examined. In other cases, fish can have a delay in sexual maturity associated with environmental stressors. In this case, several fish would appear as outliers. As noted above, the analyses should be conducted both with the outliers (to see if there are differences between sites) and without the outliers (to see if the fish with gonadal development are showing normal levels of gonadal development).

There may be cases when some fish within a population are different—for example, in situations where some fish may skip a year of spawning. If one is evaluating impacts on spawning, the analysis should consider the potential impacts on spawners and non-spawners independently. Individuals that skip reproductive seasons can usually be identified as negative outliers in a plot of gonad weight vs. body weight, i.e., plots of residuals from ANCOVA will be skewed left, and will not be normally distributed. These individuals should be excluded from analyses of reproduction, and possibly all variables. The reductions in variance achieved will usually compensate for any loss of power from reduced sample sizes. If females skipping reproductive years are excluded, that exclusion should be made objectively (Environment Canada 1997). Also, the frequency of such individuals in reference vs. exposed areas should be provided, in case skipping reproductive years is related to exposure. It is much more difficult to identify males that might skip reproductive years, if in fact that ever occurs.

8.4 Effects on Usability of Fisheries Resources

The purpose of examining the usability of fisheries resources is to determine whether the effluent has altered fish in such a way as to limit the resources' use by humans. Fish usability can be affected by altered appearance, altered flavour, or odour (tainting), or tissue contaminant levels that exceed consumption guidelines for human health and levels found in the reference area. Table 8-5 outlines the effect and supporting endpoints and appropriate statistics (or guideline levels) that are applicable for usability of fisheries resources.

Table 8-5: Fish Tissue effect and supporting endpoints and statistical procedures

	Variable	Statistical Procedure
Effect Endpoint ¹	Contaminants in fish tissue (mercury)	ANOVA, and evaluate against tissue guideline levels
Supporting ² Endpoints	Physical abnormalities	Chi-square (separate test done for each class of abnormality; number of tests will depend on how many classes of abnormalities are present in the fish collected)
	Tainting	ANOVA

¹ Effect endpoint to be used for determining "effects" as designated by exceedence of tissue guideline levels. Statistically significant differences between exposure and reference areas may also be relevant (MMER Schedule 5, s. 9(c)).

² These analyses are for informational purposes, and significant differences between exposure and reference areas are not necessarily used to designate an effect.

8.4.1 Mercury in Fish Tissue

One of the methods for evaluating fish usability is by measuring concentrations of contaminants of concern in tissue from fish collected from the exposure and the reference areas. Contaminants may be identified as a concern if they are present in the effluent and there are applicable human health consumption guidelines for those contaminants. Local consumption and commercial fisheries should guide which fish species and edible tissues (e.g., liver, kidney, bones, flesh, or even entire fish) should be analyzed. Chapter 3 provides further guidance on methods for determining which (if any) contaminants should be included in the analyses. This determination depends, in part, on previously collected data on contaminant levels in fish tissue and the effluent.

Mines are required to measure levels of mercury in fish tissue if mercury is detected in the effluent (during effluent characterization – Chapter 5) above 0.10 µg/L. An effect in fish tissue, as defined in the MMER, means measurements of concentrations of total mercury that exceed 0.5 µg/g wet weight in fish tissue taken in an exposure area and that are statistically different from and higher than the measurements of concentrations of total mercury in fish tissue taken in a reference area (MMER, Schedule 5, section 1).

Other potential metal mine-related contaminants of concern on a site-specific basis include copper, zinc, manganese, cyanide, radium, and uranium.

Chapter 3 recommends that tissue analyses be performed on five composite samples (each composed of at least eight individual fish) of a single species (preferably one sex) for each of both areas. That is, the sample size (n) for the ANOVA is five. This would be sufficient replication to detect an effect size of ± 2 SD at power = 0.9, if α and β are set at 0.1 (see Section 3.0). Thus, careful consideration should be given to the appropriate effect size to use for the particular contaminant of concern and whether increased replication may be justified. If lesser effect sizes (i.e., less than 2 SD) or greater power levels are decided to be more appropriate for the contaminant, it will be necessary to increase sample size by analyzing more composite samples.

Percent lipid and percent moisture should also be reported for each tissue sample. This is for informational purposes only to aid in data interpretation. Statistical differences in percent lipid or moisture does not constitute an effect.

8.4.2 Physical Abnormalities

Fish usability can be affected by altered appearance of fish. The data collected during the biological monitoring studies shall be used to identify the sex of the fish sampled and the presence of any lesions, tumours, parasites or other abnormalities (MMER Schedule 5, s. 16(b)). Obvious abnormalities may include:

- tumours and/or lesions on the body surface (including the eyes, lips, snout, gills);
- spinal column malformations;
- eroded, frayed or hemorrhagic fins;
- other physical malformations; or
- obvious parasites.

For each class of abnormality that has been noted, a comparison between reference- and exposure-area fish should then be done using a chi-square goodness-of-fit test for relative frequencies. This information is used to help interpret effects, although, for EEM purposes, a significant difference does not necessarily signify an effect. The number of statistical tests that are necessary will depend on the number of classes of abnormalities that are noted in the collected fish. Sample size will have been determined by the number of fish collected for the fish survey. Cohen (1988) provides guidance on the power of a chi-square test that would result from that level of replication.

8.5 Data Assessment and Interpretation for the Benthic Invertebrate Community Study

The data collected during the benthic invertebrate community survey shall be used to determine the following effect indicators (MMER Schedule 5, s. 16(a)(iii)):

- 1) total benthic invertebrate density;
- 2) the evenness index;
- 3) taxa richness; and
- 4) the similarity index (referred to in this document as Bray-Curtis Index).

The above effect indicators are to be used for determining statistically significant differences between exposure and reference areas or along an exposure gradient. See Chapter 4 for additional information on these effect indicators. The mean, median, SD, SE, and minimum and maximum values are determined for each effect endpoint for the sampling areas. In addition, an analysis of the results shall be used to determine if there is a statistical difference between the sampling areas for each of the effect indicators (MMER Schedule 5, s. 16(c)).

8.5.1 Study Design and Statistical Procedures

Table 8-6 outlines the appropriate statistical procedures that are applicable for analysis for each of the recommended study designs. See Chapter 4 for additional information on these study designs. In contrast to the fish survey, the statistical procedure used to determine whether there has been an effect is dependent on which of the seven study designs is employed. For a given study, all four effect indicators are analyzed using the same study-design-determined statistical procedure. The one exception is the Reference Condition Approach, which uses a different set of statistical procedures that do not require inter-area comparisons of these four indicators, unless accompanied by ANOVAs; the procedures for this study design are outlined below and in Chapter 4.

Table 8-6: Statistical procedure used to determine an effect for each of the seven study designs

Study Design	Statistical Procedure
Control-Impact (C-I)	ANOVA
Multiple Control-Impact (MC-I)	ANOVA
Before/After Control-Impact (BACI)	ANOVA
Simple Gradient (SG)	Regression/ANOVA
Radial Gradient (RG)	Regression/ANOVA

Multiple Gradient (MG)	ANCOVA
Reference Condition Approach (RCA)	Multivariate/ANOVA

Note: Multivariate analyses can be performed on data collected using any of the designs in Table 8-6, to look for patterns that may be useful for highlighting potential areas of concern. Under certain circumstances, ANCOVAs may also be appropriate for any of these designs (e.g., to factor out the effect of a potentially confounding environmental variable).

Although it is possible to use ANOVA to analyze data collected under most of the study designs listed in Table 8-6, ANOVA is most applicable to the control-impact (C-I) and multiple control-impact (MC-I) designs. The simplest of these study designs is the C-I (or reference/exposure) design. In rivers, for example, this consists of one (usually upstream) reference area and one or more downstream exposure areas. Chapter 4 provides guidance on the different ways that C-I designs can be laid out. This type of study design employs ANOVA comparisons between reference and exposure areas, with a significant difference signifying an effect.

The MC-I design is similar to the C-I design, except that it employs additional reference areas that are located in adjacent watersheds or bays where the sampled habitat is comparable to that found within the exposure area. This type of design helps to reduce problems with confounding factors (e.g., when a single reference area differs from an exposure area with respect to several environmental variables in addition to the point-source effluent). Analogous to a C-I design, a significant difference between an exposure area and the mean of the reference areas, as determined by ANOVA, would represent an effect.

ANCOVA can also be used for both C-I and MC-I designs to factor out covariates that may create “noise” that makes it difficult to make simple ANOVA comparisons of reference to exposure areas. For example, without the use of ANCOVA, differences in depth among stations within the reference and exposure areas may mask effluent-related differences that may exist between those areas. This may occur when the benthic invertebrate indicators change along a continuum of increasing depth, and when it is not possible to take all samples at identical depths. In this example, ANCOVA can be used to factor out the effect of the depth covariate so as to focus on the effect of effluent exposure. The same approach can be used for other covariates that influence the benthic invertebrate indicators along a continuum.

An improvement to the above C-I and MC-I designs is possible when data can be collected both before and after initiation of effluent discharge into the receiving water area. This kind of monitoring design has been termed a before/after control-impact (BACI) design (Schmitt and Osenberg 1996). Use of a BACI design helps to distinguish effluent effects from natural differences between reference and exposure areas that may have existed before the initiation of effluent discharge.

In its simplest form, a BACI design entails collecting monitoring data at least once, both before and after initiation of effluent discharge in both a reference and exposure area, with the data analyzed using an area-by-time factorial ANOVA (Green 1979). In this situation, evidence for an effluent effect is inferred when the area-by-time interaction term in the ANOVA is significant. When the reference and exposure areas have been sampled repeatedly during both the before and after periods, it is possible to use a BACI paired series analysis, in which case the potential effects are investigated by testing for a change in delta (difference between reference and exposure) from the before to after period (Schmitt and Osenberg 1996). The design can be further improved by incorporating multiple reference areas (Schmitt and Osenberg 1996; Underwood 1997).

In contrast to the C-I and MC-I designs, the simple gradient (SG) and radial gradient (RG) designs are more amenable to regression analysis. The assumptions for regression analysis are applicable to the analysis of the benthic invertebrate community data, and have already been outlined in the section 8.3.3.2 discussion on ANCOVA (regression is one component of ANCOVA).

For additional information on study designs, refer to chapters 2 and 4.

8.5.2 Data Treatment

As for the fish survey, the data should be reported in both graphical and tabular format for each area (reference and exposure area(s)) being summarized. The reported data will include the descriptive statistics (mean, median, SD, SE, and minimum and maximum values) as well as the sample sizes. Gradient data should be presented graphically as scatter plots of variable vs. distance from the effluent outfall. For gradient designs with no discrete “areas,” tabular presentation prior to the main analysis would be applicable to station-by-station summary statistics, with the sampling unit being field sub-samples rather than stations. Station-by-station summary statistics are also applicable to C-I-type designs in cases where field sub-samples are not pooled prior to taxon enumeration, although the key summary statistics are those that are calculated for whole areas (to help with interpreting significant differences [“effects”] among areas).

The same three main analysis steps outlined in section 8.3.3 should be followed to determine whether statistically significant “effects” have occurred:

- 1) The data should be inspected to see whether they satisfy the assumptions of the statistical test or procedure being used (ANOVA, ANCOVA, regression or multivariate analyses).
- 2) The appropriate statistical procedure would be performed following data inspection and any necessary transformations (or non-parametric alternative).

- 3) The key results for the effect indicators should then be presented to clearly indicate whether there has been an effect, with details on the nature of the effect (including direction and magnitude). Again, an effect is declared if the p-value is less than the a priori α value determined, as outlined in section 8.6.

The same considerations and constraints discussed in section 8.3.3 for conducting ANOVA and ANCOVA analyses apply to benthic invertebrate community analyses using those two statistical procedures. Thus, data inspection, analysis and interpretation when using ANOVA or ANCOVA for the benthic invertebrate community survey should follow the generic recommendations provided in section 8.3.3.

Gradient designs are particularly useful for 1) situations where rapid effluent dilution precludes the selection of an exposure area that is comparatively homogeneous in terms of effluent concentration and 2) determining how far along an effluent path the effects are observed (i.e., determining the geographical extent of “effects”). The geographic extent of “effects” can be determined graphically by plotting the response variable(s) against distance from the effluent outfall, and inspecting the data for an inflection point where the response variable asymptotes to the reference condition. Data from sampling stations arrayed in this manner could also be used, together with measured physicochemical data, in a multivariate analysis (e.g., ordination or clustering) that is used to identify which more distant stations tend to group with reference stations and which tend to group with clearly affected stations.

Both of these approaches (graphical plotting and multivariate analysis) look for patterns in the data to qualitatively determine the approximate geographic extent of an effect. That is, they do not necessarily entail hypothesis testing, and therefore, in the context of the EEM program, are not used to designate an effect sufficient to warrant follow-up action, but rather are used for informational purposes.

Nevertheless, statistical tests are possible for some gradients. In the simplest case, an effect would be declared if the slope of the regression of the variable against distance from the effluent source is significantly different than zero, or if the correlation coefficient is statistically significant (data transformations may be necessary to satisfy assumptions of linearity). In this case, the effect is a relatively uniform gradient of variable values away from the point source, rather than an effect in a given discrete area.

An effect can also be signified by a significant exposure vs. reference ANOVA difference when comparing a group of stations along the gradient close to the mine to “reference” stations along the gradient far from the mine. This is analogous to the C-I approach, and assumes some degree of uniformity in exposure within the exposure group of stations and within the “reference” group of stations. Furthermore, the two groups of stations would need to be far enough apart to represent clear differences in exposure, and a sufficient number of stations would need to be available for each group to attain the desired level of power. Based on the power analysis discussion in the following section,

an initial recommendation is to have at least five fairly uniform stations relatively close to the mine (high effluent exposure area) and five fairly uniform stations far enough from the mine to approximate a “reference” area (i.e., minimally affected by the effluent). Providing intermediate stations would likely necessitate a total of at least 15 gradient stations overall.

Regardless of the method of analysis, overall statistical power is usually improved by emphasizing station replication on the 2 ends of the gradient. Again, emphasis should also be placed on extending the gradient sufficiently far from the mine (as much as is feasible) to allow sampling of stations that are as minimally affected as possible (and that serve as approximate “reference” stations).

Given sufficient sub-samples per station, it is also possible to use ANOVA to determine the presence or absence of an effect for a given station. This would entail using field sub-samples as replicates (treating stations as areas) and making station-by-station ANOVA comparisons of more high effluent exposure stations along the gradient to more distant reference stations. This method of analysis could be used to determine where along the gradient an effect disappears at the given α level of significance. This latter approach may, however, require extensive sampling effort, depending upon the number of stations along the gradient and the required (by power analysis) number of field sub-samples per station.

In cases where these kinds of statistical tests are not adequate for a given gradient design, a redesign of the monitoring program will be necessary to enable an appropriate statistical test during the next monitoring study. The redesign may entail increased replication focused on the key exposure and reference areas (or stations) that are to be compared (e.g., increased replication in the area of greatest effluent exposure and in the area with the lowest effluent exposure that best represents reference conditions).

In some cases, it may be necessary to compare exposure vs. reference gradients. This would be the case when a co-occurring (non-mine-related) environmental gradient (i.e., covariate) confounds effluent effects in the exposure area. By using a multiple gradient (MG) design, it may be possible to make statistical comparisons of the exposure area gradient to a similar (non-mine-related) environmental gradient in an unexposed reference area. The reference gradient should be as similar as possible in depth and habitat to the exposure area gradient. Potential effluent “effects” would be tested for by using ANCOVA to compare reference to exposure area regression elevations (or adjusted means), while factoring out the influence of the co-occurring environmental covariate.

For example, if the gradient in effluent exposure away from the mine was confounded by a co-occurring increase in depth, an ANCOVA comparison might be made to a reference area where the depth gradient is the same. If the slopes for the reference and exposure area regressions against the covariate ($X = \text{depth}$) are approximately equal, a significant difference in adjusted means would indicate an effect of the effluent on the effect

indicator Y (e.g., taxon richness). Again, section 8.3 provides further guidance for ANCOVA analyses and the different ways these analyses can be used to indicate an effect.

8.5.3 Reference Condition Approach

The reference condition approach (RCA) is a study design that combines inspection of multivariate patterns in the data with assessments of whether exposure stations fall outside a given ordination probability ellipse for reference stations. The fundamental concept of the RCA is to establish a database of stations that represent unimpaired conditions (reference stations) at which biological and environmental attributes are measured. This database is used to develop predictive models that match a set of environmental variables to biological conditions. These predictive models then allow a set of environmental measurements to be made at a new station and used in the model to predict the expected biological condition at the new station. An assessment of whether there has been an effect at the exposure station is enabled by a comparison of the actual biological condition at the new (exposure area) station with conditions at the reference stations to which the new station is predicted as belonging.

The reference condition database is established by an initial standardized sampling program at a wide variety of spatial scales. The same benthic macroinvertebrate sampling protocol is used in as many ecoregions and stream orders or lakes as are available in a catchment. A number of environmental variables are measured in conjunction with invertebrate sampling. The data are then subjected to a 3-step multivariate analysis in which:

- 1) a number of invertebrate groups are formed based on similarity of community structure;
- 2) biological data are correlated with environmental attributes, and an optimal set of environmental variables is identified that can be used to predict group membership; and
- 3) the biological condition of test (exposure) stations is assessed by using the optimal set of environmental variables to predict group membership.

How the test station fits, relative to the group to which it is predicted to belong, establishes whether and to what degree the station is different from the reference group. A station or group of stations that fall outside the statistically determined ordination probability ellipse for the reference stations signifies the presence of an effect. The boundaries of the reference ellipse should be set a priori based on some of the considerations discussed in section 8.6. A more complete discussion of the assumptions, procedures and interpretation of the RCA is available in Reynoldson et al. (1995, 2000) and Bailey et al. (2003).

It should be further noted that, depending on the timing and locations of an RCA sampling program, it may also be possible to use the resulting database to make ANOVA comparisons between reference and exposure areas in order to determine whether there has been an effect. This latter kind of analysis would be analogous to an MC-I design.

To summarize, an overall procedure similar to that outlined in section 8.3 should also be followed (with appropriate modifications) for the benthic invertebrate community survey. However, the power analysis is not applicable to graphical approaches and the RCA. Consequently, RCA studies should be designed in a way that provides an accurate and precise determination of reference conditions so as to maximize the likelihood of detecting departures from reference conditions at exposure stations, when they exist. The following elements may be included as part of an RCA study:

- 1) Preparing the analyses: QA/QC (including checks for data entry errors), summary of confounding factors, description of the sampling design and taxonomic level used, clear identification of the sampling units used for statistical comparisons (e.g., stations rather than field sub-samples), ensuring equivalence of sampling substrata, and sampling techniques among different reference and exposure areas being compared
- 2) Summary statistics (graphical and tabular presentation of means, etc., as described above)
- 3) Statistical analyses (hypothesis testing) to determine “effects” (ANOVA, ANCOVA, regression)
- 4) Graphical approaches (e.g., inspection of the shape of regression lines, which is used for inspecting patterns in the data rather than determining “effects”)
- 5) Multivariate statistical analyses used for determining a) patterns in the data and b) the position in multivariate space of exposure stations relative to reference ordination probability ellipses; only b) is used to determine “effects”
- 6) Power analyses (not applicable to graphical approaches and RCA)

8.5.4 Supporting Endpoints

The following benthic invertebrate community supporting endpoints should also be reported, including means, medians, SDs, SEs, minimum and maximum values, and sample sizes:

- Simpson’s diversity
- taxon (e.g., family) density
- taxon (e.g., family) proportion
- taxon (e.g., family) presence/absence

Unlike the effect endpoints (total benthic invertebrate density, the evenness index, taxa richness and the similarity index), the above-listed variables are included as supporting endpoints and are not statistically analyzed to determine “effects.” They may, however,

be used to interpret effects at later stages (e.g., determining the magnitude and causes of “effects”). These should be reported in both graphical and tabular format for each area (reference and exposure area(s)) being summarized. It should be noted that there may be other descriptors that may also be useful for the interpretation of monitoring data, on a site-specific basis (see Resh et al. 1995 for a review).

8.6 The Role of Power Analysis, α , β and Critical Effect Size in Determining Effects

8.6.1 Setting α and β

In testing whether exposure areas differ significantly from reference areas, a low probability of a Type I error (α) is usually allowed so that a normal population or community will not be mistaken for an affected one. However, the monitoring program should also be designed to provide a reasonably high probability of statistically detecting a predetermined critical effect size (CES) if it has occurred, i.e., the power of the test should be high. Power is $1-\beta$, where β is the Type II error (see below).

Type I error is partially kept in check by setting a broad margin for variation around what is considered “healthy.” Sufficient sampling effort should also be expended to reduce Type II error, taking into account the low probability allowed for Type I error. Thus, to determine what sampling effort is required, the CES and the Type I and Type II error will all be taken into account and set a priori. That is, decisions should be made about the magnitude of Type I and Type II errors that are acceptable for determining power and thus the sampling effort required to detect the recommended CES.

Type I error occurs (at probability α) if the null hypothesis that there is no effect is rejected when in fact it is true (e.g., an exposure area is declared as being different from reference when it is not).

Type II error occurs (at probability β) if the null hypothesis is accepted when it is false (e.g., the exposure area is declared as not being significantly different from reference when it is actually impaired). Therefore, α is the risk to industry and β is the risk to the environment.

The power of a statistical test is $1-\beta$, the probability associated with correctly rejecting the null hypothesis when it is false (e.g., the probability associated with correctly identifying an impaired area). In a well-designed, properly replicated monitoring program, the goal is to keep α and β low and power high.

As can be seen from the equation given later in this section, one way to increase power, given a fixed sampling effort (i.e., sample size), is to increase α , i.e., there are trade-off decisions to be made when setting α and β . Traditionally, α has been set at 0.05 for

experimental studies where, in many cases, the cost of a Type II error is not particularly high. That is, an α of 0.05 is typically used in situations where the primary concern is to have maximal confidence that a statistically significant effect is real. On the other hand, there is much less consensus and available literature on what is an appropriate level for β . Some studies have suggested using a minimal power of 0.8 (i.e., $\beta = 0.2$) (Alldredge 1987; Cohen 1988; Burd et al. 1990; Osenberg et al. 1994; Keough and Mapstone 1995).

In many cases, “this rule of thumb” can be traced back to Cohen’s seminal work on power analysis (see Cohen 1988), which is primarily geared toward applications in the behavioural sciences. For those types of applications, Cohen contended that Type I errors were likely to be more serious than Type II errors for cases where the biggest concern is to not propagate erroneous conclusions based on incorrect declarations of significant differences. Specifically, he suggested that, if Type I errors were to be considered four times more serious, it might be reasonable to set α at the traditional (in terms of experimental studies) 0.05 and β at $4 \times 0.05 = 0.2$. He cautioned, however, that this rule of thumb should be ignored for other types of studies where these assumptions are not applicable.

This latter caveat applies to environmental monitoring studies where, because of the potentially high cost (both ecological and monetary) of failing to detect negative impacts, many researchers in the field of biomonitoring argue that α should be set at least to the same level as β (e.g., Alldredge 1987; Underwood 1993; Mapstone 1995). That is, the argument has been widely made that, barring extenuating circumstances, the risk to the environment should not be set greater than that to industry. This suggests that the most reasonable starting point is to set $\alpha = \beta$, and this position has been adopted by the EEM program. On a site-specific basis, it may sometimes be decided to 1) set $\alpha > \beta$ if it can be shown that the risk to the environment is of greater concern than the risk to industry, or to 2) set $\alpha < \beta$ if it can be shown that the risk to industry is of greater concern.

After deciding to set $\alpha = \beta$, it is necessary to make a decision on an appropriate value for α and β . In many cases, this decision will be made within the context of the desired power of the test, the CES that the program is to be designed to detect, and the implications for sampling effort. This decision-making process can be illustrated using Table 8-7 for the benthic invertebrate survey, where the effects on sample size of setting α and β at different levels were examined for detecting a CES of ± 2 SD by using the following power analysis equation, which yields an approximate sample size (n) in one step for the most basic C-I ANOVA design (see also the discussion in the next section for further details) (Guenther 1981; Alldredge 1987):

$$n = (2(Z_{\alpha} + Z_{\beta})^2(SD/CES)^2) + 0.25(Z_{\alpha})^2$$

where:

n = sample size

Z_{α} = standard normal deviate for α significance level (Type I error)

Z_{β} = standard normal deviate for β significance level (Type II error)

SD = standard deviation

CES = critical effect size

Table 8-7: Sample sizes required to detect difference of ± 2 SD

α	$1-\beta$			
	0.99	0.95	0.90	0.80
0.01	14	11	10	8
0.05	11	8	7	5
0.10	9	7	5	4

Using Table 8-7 for guidance (and the recommendation that the benthic invertebrate community survey should minimally have sufficient power to detect a CES of ± 2 SD), the benthic invertebrate working group recommended α and β be initially set at 0.1. This implied that, in most cases, the sampling effort would require a sample size of 5, which is within the range used in many benthic surveys (Resh and McElravy 1993). Basic ANOVA power analysis calculations also indicate that α and β can be set equal to 0.1 for the fish survey effect endpoints as well, with very little effect (relative to $\alpha = 0.05$, $\beta = 0.2$) on the sample size required to achieve the resulting level of power ($1-\beta$). The use of an α or β level other than 0.1 would require appropriate justification by either the proponent or the Authorization Officer (e.g., setting a more rigorous, lower Type II error (β) when the risk to the environment is judged to be of greater concern). Consultation with the Authorization Officer may also be required in cases where power analysis recommends the use of unreasonably high sample sizes.

It should also be noted from Table 8-7 that, by increasing sample size, it is possible to obtain lower Type I and II errors (lower α and β) while maintaining α equal to β . For example, α and β can both be set at 0.05, resulting in 95% power to detect a CES of ± 2 SD, by increasing sample size to 8 (see Table 8-7). The same argument applies to the other components of EEM (e.g., the fish survey and fish usability components) for different desired CESs, although the required sample sizes will be different. Thus, setting α equal to β provides an economic incentive to carrying out a well-designed, well-replicated monitoring program, because providing sufficient replication will help reduce the probability of Type I errors (i.e., α is kept low), thereby reducing the probability of unnecessary follow-up studies. Furthermore, since α is linked to β , the power of the monitoring program to detect real effects will also be increased. This improvement in monitoring design helps to ensure a better understanding of what types of effects, if any, are occurring.

8.6.2 Power Analysis: Determination of Required Sample Size, Power

and Appropriate Critical Effect Size

Power analysis is used for two major purposes during EEM:

- 1) at the beginning of a monitoring study (a priori), to calculate the sampling effort (sample size) that will be required to detect a given CES at a given level of power; and
- 2) following a recently completed monitoring study (post hoc), to determine the level of power that was actually achieved.

Both of these uses of power analysis are briefly reviewed here to help clarify the relationship between the two.

8.6.2.1 A Priori Power Calculations

During the initial design phase of an EEM study, power analysis can be used to determine the sample size required to achieve a test adequate to detect an effect equal to a predetermined CES prior to sampling. Using the CES, the probability of Type I error “ α ,” the probability of type II error “ β ,” an estimate of reference variability (e.g., SD for the reference area), and making some assumptions about the distribution of the data being evaluated, a scientifically defensible sampling strategy can be devised. The discussion below outlines the most basic (i.e., C-I ANOVA or ANCOVA) procedure for determining required sample size. Sample size refers to the number of fish for the fish survey and the number of stations for the benthic invertebrate community survey. In cases where the required sample size calculated for one effect endpoint (e.g., invertebrate density/condition) is greater than that calculated for another (e.g., invertebrate taxon richness / relative gonad weight), the greater sample size should be used (unless, as discussed above, consultation with the Authorization Officer confirms that this would result in excessively high sample sizes).

Once CES has been determined, the levels of α and β have been selected, and SD for the particular mine location in question has been estimated, they are entered into the power analysis equation to calculate the sample size required to detect an impact of magnitude CES between or among areas at a given power level. For the case where CES is set at $\pm 2SD$, due to cancelling of terms the determination of SD is not required for the power analysis, and Table 8-7 above gives pre-calculated sample sizes for various values of α and β .

It should be noted that determination of required sample size assumes that the variability among replicates for the exposure area is similar to that for the reference area. Although ANOVAs are fairly robust with respect to violation of normality assumptions, if the variance within an exposure area is much higher (or lower) than within the reference area, ANOVA comparisons may not be appropriate unless the variances can be made homogeneous by transformation. For the case where the exposure and reference

variances remain significantly different following transformation, the power analysis outlined here may overestimate or underestimate the number of sampling stations required. Non-parametric tests may be used in this case; non-parametric power analyses would then be required to estimate required sampling effort (Thomas and Krebs 1997).

For a basic C-I ANOVA or ANCOVA design, the estimated sample size required to detect a given CES at a given power level can be calculated by arranging the standard power analysis equation as follows (Green 1989):

$$n = 2(t_{\alpha} + t_{\beta})^2(SD/CES)^2$$

where:

n = sample size

t_{α} = value of Student's t statistic (two-tailed) with (n-1) degrees of freedom (df) at a significance level of α

t_{β} = value of Student's t statistic (one-tailed) with (n-1) df at a significance level of β

SD = standard deviation

CES = critical effect size, represented in the measurement units of the response variable

The equation is solved iteratively by choosing an approximate value of n (usually 20 for the fish survey) to look up t_{α} and t_{β} and then using the solution to find a more accurate n; the procedure is repeated until arriving at a final estimate for n (see section A1-8 of Appendix 1). Alternatively, the equation given in section 8.6.1 can be used to approximately solve for n in one step. Pre-calculated tables of n (expanding upon Table 8-7) are available for a variety of values of α , β and CES (Alldredge 1987; Cohen 1988).

The reader is referred to the appropriate literature (e.g., Cohen 1988) for guidance on power analysis and tables for determining sample size for regression (simple gradient, radial gradient) and chi-square (analysis of physical abnormalities in fish) monitoring designs. A number of software programs are also available for conducting power analyses for a variety of statistical designs (Thomas and Krebs 1997). As for a basic C-I design, power analysis for these other designs will also require an a priori decision on an appropriate magnitude for CES. For regression analyses, Cohen (1988) gives a table for converting CESs from SD units to a correlation coefficient (r), and in some cases it may be acceptable to use this r to look up the approximate sample size required for a regression-type gradient design. For example, given certain assumptions, he shows that using a CES of 2 SD is equivalent to using $r = 0.707$ (or $r^2 = 0.5$). Although the exact equivalency depends on the assumptions involved, it may be acceptable to use this conversion (possibly with a correction factor) to obtain an approximate CES appropriate for use in regression-type analyses. Tables are provided in Cohen (1988) for looking up required sample sizes for various values of r, α and β .

CESs for the fish survey are percentages of the reference mean and are not represented in

the measurement units of the response variable, as these effect sizes would vary for different studies. Therefore, the coefficient of variation (COV), expressed as a percentage of the reference mean ($COV = SD / \text{reference mean} \times 100$) is used as a measure of variability in sample size calculations. For a basic fish survey C-I ANOVA design with untransformed data (e.g., as used for the age effect endpoint), the estimated sample size required to detect a given effect size at a given power level can be calculated by using a different version of the equation above. This equation is as follows (Green 1989):

$$n = 2(t_{\alpha} + t_{\beta})^2 (COV/CES)^2$$

where:

COV = coefficient of variation (expressed as a percentage using reference site data)

CES = critical effect size (expressed as a percentage of the reference mean)

For a basic C-I ANCOVA design using log-transformed data (e.g., as used for the relative gonad weight effect endpoint), the estimated sample size required to detect a given CES at a given power level can also be calculated by using a different version of the equation above. This equation is as follows (Green 1989):

$$n = 2(t_{\alpha} + t_{\beta})^2 (SD_z/CES_z)^2$$

where:

SD_z = standard deviation of the residuals using log-transformed data

$CES_z = \log(f+1)$, where f = CES represented as a fraction of the reference mean (e.g., for a CES of 25% $\Rightarrow f = 0.25$)

For both of the above equations, sample size must be solved iteratively by choosing an approximate value of n to start with as discussed above.

8.6.2.2 Post Hoc Power Analyses

After completion of a sampling program, if a non-significant result has been obtained, a post hoc power analysis can be used to calculate the actual power that was available to detect an effect and the minimum CES that could be detected for a given power (Quinn and Keough 2002). This is particularly important if any of the relevant parameters that could affect power (i.e., n , α , CES, SD) have changed since the beginning of the study. In addition, these calculations should be used to make sample size recommendations for the subsequent monitoring study. The post hoc power calculations can be performed by rearranging the formulas above to solve for t_{β} or the CES. For example, to calculate power for the previous two equations, we obtain:

$$t_{\beta} = \sqrt{\frac{n}{2(COV/CES)^2}} - t_{\alpha}$$

and

$$t_{\beta} = \sqrt{\frac{n}{2(SD_z / CES_z)^2}} - t_{\alpha}$$

Power can then be obtained from the calculated value of t_{β} .

8.7 Critical Effect Sizes

To ensure that increased monitoring efforts are focused in the appropriate areas, Environment Canada has developed CESs for key fish and benthic invertebrate survey effect endpoints. See Chapter 1 for the table on CESs and for additional information.

8.8 Statistical Considerations for Mesocosm Studies

Some considerations would be unique to a mesocosm-type study. For example, control over experimental considerations would likely result in lower levels of variability within reference and exposure treatments, as compared with field data. This may make it possible to attain equivalent levels of statistical power using smaller sample sizes than used in the field. In the same vein, it may be possible to attain higher power levels or to detect smaller effect sizes while using the same sample sizes as used in the field. In fact, it may be desirable to have sufficient power to detect smaller effect sizes in mesocosm studies than in field surveys, due to the shorter exposure times typical of mesocosm studies. That is (using hypothetical numbers), a 10% effluent-induced change over a 30-day exposure period in a mesocosm study may be equivalent to a 25% change over a much longer lifetime exposure in the field.

In addition, due to the possibility of caging artifacts, it may be necessary to switch from using individual fish as the sampling unit for replication (as in the field) to using individual experimental enclosures (mesocosms) as sampling units. Using two mesocosm units (one for reference and one for exposure) with 20 fish each may not be valid, because it may not be possible to separate effluent effects from the effects due to subtle differences in the experimental enclosures. This is an example of the potential for confounding effects due to pseudo-replication (Hurlbert 1984).

In comparison to the fish survey, it may be even more straightforward to substitute mesocosm studies for benthic invertebrate community field monitoring, at least in terms of statistical design and analysis. As for the fish survey, the same steps outlined for data preparation, presentation and analysis would apply. Furthermore, due to comparatively fast turnaround times for changes in invertebrate community structure within mesocosms, it may be possible to use the same effect endpoints as used in the invertebrate field

survey (section 8.5). The most likely study design would be analogous to the C-I design (Table 8-6), with ANOVA comparisons being made between replicated reference and exposure mesocosms. The sampling units would be the individual mesocosms (equivalent to “stations” in the field survey). As for fish mesocosms, control over variability under experimental conditions may make it possible to attain greater statistical power or to detect smaller effect sizes (in terms of percentage change) using the same sample sizes as typically used in the field. This increase in *precision* is one of the most frequently cited advantages of using mesocosms in place of field sampling, and is weighed against the disadvantage of a potential decrease in *accuracy* due to using a (hopefully realistic) simulation of actual field conditions.

Chapter 9 provides more extensive discussion on data assessment and interpretation for alternative methods.

8.9 References

Allredge JR. 1987. Sample size for monitoring of toxic chemical sites. *Environ Monit Assess* 9:143-154.

Bailey RC, Norris RH, Reynoldson TB. 2003. Bioassessment of freshwater ecosystems: using the reference condition approach. Boston (MA): Kluwer Academic Publishers.

Barrett TJ, Tingley MA, Munkittrick KR, Lowell RB. 2010. Dealing with heterogeneous regression slopes in analysis of covariance: new methodology applied to environmental effects monitoring fish survey data. *Environ Monit Assess* 166(1-4):279-291.

Bartlett JR, Randerson PF, Williams R, Ellis DM. 1984. The use of analysis of covariance in the back-calculation of growth in fish. *J Fish Biol* 24:201-213.

Bligh EG, Dyer W. 1959. A rapid method of total lipid extraction and purification. *Can J Biochem Physiol* 37:911-917.

Bolger T, Connolly PL. 1989. The selection of suitable indices for the measurement and analysis of fish condition. *J Fish Biol* 34: 171-182.

Burd BJ, Nemec A, Brinkhurst RO. 1990. The development and application of analytical methods in benthic marine infaunal studies. *Adv Mar Biol* 26:169-247.

Cohen J. 1988. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale (NJ): Lawrence Erlbaum Associates.

Cook RD. 1977. Detection of influential observation in linear regression. *Technometrics* 19:15-18.

Cook RD. 1979. Influential observations in linear regression. *J Amer Stat Assoc* 74:169-174.

Day RW, Quinn GP. 1989. Comparisons of treatments after an analysis of variance in ecology. *Ecol Monogr* 59:433-463.

Draper NR, Smith H. 1981. Applied regression analysis. 2nd ed. New York (NY): John Wiley & Sons, Inc.

Environment Canada. 1997. Fish survey expert working group report. EEM/1997/6. Ottawa (ON): Environment Canada.

Folch J, Lees M, Sloane Stanley GH. 1957. A simple method for the isolation and purification of total lipides from animal tissues. *J Biol Chem* 226:497-509.

Fox J. 1997. Applied regression analysis, linear model, and related models. Thousand

Oaks (CA): Sage Publications Inc.

Gibbons DW, Reid JB, Chapman RA. 1993. The New Atlas of Breeding Birds in Britain and Ireland: 1988–1991. London (UK): Poyser.

Gray MA, Curry AR, Munkittrick KR. 2002. Non-lethal sampling methods for assessing environmental impacts using a small-bodied sentinel fish species. *Water Qual Res J Can* 37:195-211.

Green RH. 1979. Sampling design and statistical methods for environmental biologists. New York (NY): Wiley-Interscience.

Green RH. 1989. Power analysis and practical strategies for environmental monitoring. *Environ Res* 50:195-205.

Grubbs F. 1969. Procedures for detecting outlying observations in samples. *Technometrics* 11:1-21.

Guenther WC. 1981. Sample size formulas for normal theory *T* tests. *Am Stat* 35:243-244.

Hamilton BL. 1977. An empirical investigation of the effects of heterogeneous regression slopes in analysis of covariance. *Educ Psychol Meas* 37:701-712.

Hubert JJ. 1980. Bioassay. Dubuque (IA): Kendall/Hunt Publishing.

Huitema BE. 1980. The analysis of covariance and alternatives. New York (NY): John Wiley & Sons, Inc.

Hurlbert SH. 1984. Pseudoreplication and the design of ecological field experiments. *Ecol Monogr* 54:187-211.

Iman, R.L., Conover, W.J. 1982. A distribution-free approach to inducing rank correlation among input variables. *Commun. Statist.-Simula. Computa.* 11, 311-334.

Jackson DA, Harvey HH, Somers KM. 1990. Ratios in aquatic sciences: statistical shortcomings with mean depth and the morphoedaphic index. *Can J Fish Aquat Sci* 47:1788-1795.

Keough MJ, Mapstone BD. 1995. Protocols for designing marine ecological monitoring associated with BEK mills.)Technical Report Series 11). National Pulp Mills Research Program. Canberra (AU): Commonwealth Scientific and Industrial Research Organisation.

Lowell RB, Kilgour BW. 2008. Interpreting effluent effects on fish when the magnitude of effect changes with size or age of fish. dans K.A. Kidd, R. Allen Jarvis, K. Haya, K.

Doe et L.E. Burrige (éd.), *Comptes rendus du 34^{ième} atelier annuel sur la toxicité aquatique: du 30 septembre au 3 octobre 2007, Halifax, Nouvelle-Écosse*. Can Tech Rep Fish Aquat Sci 2793:82-83.

Mapstone BD. 1995. Scalable decision rules for environmental impact studies: effect size, Type I and Type II errors. *Ecol Appl* 5:401-410.

Miller RG. 1986. *Beyond ANOVA: Basics of applied statistics*. New York (NY): John Wiley & Sons, Inc.

Osenberg CW, Schmitt RJ, Holbrook SJ, Abu-Saba KE, Flegal AR. 1994. Detection of environmental impacts: natural variability, effect size and power analysis. *Ecol Appl* 4:16-20.

Paine RT, Tegner MJ, Johnson EA. 1998. Compounded perturbations yield ecological surprises. *Ecosystems* 1:535–545.

Peters RH. 1983. *The ecological implication of body size*. New York (NY): Cambridge University Press. 329 pp.

Quade D. 1967. Rank analysis of covariance. *J Amer Stat Assoc* 62:1187-1200.

Quinn GP, Keough MJ. 2002. *Experimental design and data analysis for biologists*. Cambridge (UK): Cambridge University Press.

Randall R, Lee II H, Ozretich R, Lake J, Pruell J. 1991. Evaluation of selected lipid methods for normalizing pollutant bioaccumulation. *Environ Toxicol Chem* 10:1431-1436.

Resh VH, McElravy EP. 1993. Contemporary quantitative approaches to biomonitoring using benthic macroinvertebrates. In: Rosenberg DM, Resh VH, editors. *Freshwater biomonitoring and benthic macroinvertebrates*. New York (NY): Chapman and Hall. p. 159-194.

Resh VH, Norris RH, Barbour MT. 1995. Design and implementation of rapid assessment approaches for water resource monitoring using benthic macroinvertebrates. *Austral J Ecol* 20:108-121.

Reynoldson TB, Bailey RC, Day KE, Norris RH. 1995. Biological guidelines for freshwater sediment based on benthic assessment of sediment (the BEAST) using a multivariate approach for predicting biological state. *Austral J Ecol* 20:198-219.

Reynoldson TB, Day KE, Pascoe T. 2000. The development of the BEAST: a predictive approach for assessing sediment quality in the North American Great Lakes. In: Wright JF, Sutcliffe DW, Furse MT, editors. *Assessing the biological quality of fresh waters*:

RIVPACS and other techniques. Ambleside (UK): Freshwater Biological Association. p. 165-180.

Ricker WE. 1975. Computation and interpretation of biological statistics of fish populations. Bull Fish Res Board Can (23)2:519–529.

Schmitt RJ, Osenberg CW. 1996. Detecting ecological impacts: concepts and applications in coastal marine habitats. San Diego (CA): Academic Press. 401p.

Shirley EAC. 1981. A distribution-free method for analysis of covariance based on ranked data. Appl Stat 30:158-162.

Sokal RR, Rohlf FJ. 1995. Biometry. 3rd ed. New York (NY): W.H. Freeman.

Thomas L, Krebs CJ. 1997. A review of statistical power analysis software. Bull Ecol Soc Am 78:126-139.

Trippel EA, Harvey HH. 1991. Comparison of methods used to estimate age and length of fishes at sexual maturity using populations of white sucker (*Catostomus commersoni*). Can J Fish Aquat Sci 48:1446-1495.

Underwood AJ. 1993. The mechanics of spatially replicated sampling programmes to detect environmental impacts in a variable world. Austral J Ecol 18:99-116.

Underwood AJ. 1997. Experiments in ecology: their logical design and interpretation using analysis of variance. Cambridge (UK): Cambridge University Press. 504 pp.

Appendix 1: Step-by-Step Guidance through Statistical Procedures

The following provides statistical background and step-by-step guidance through the statistical procedures required for the environmental effects monitoring (EEM) fish survey. This background material and the step-by-step procedures are meant as general guidance, and can be adapted to the particular statistical software package procedures that are being used. Examples are taken from different data sets from pulp and paper EEM cycles to illustrate concepts where possible.

Analysis of covariance (ANCOVA) can be performed as multiple linear regression with indicator variables to represent sites. In an analysis with a reference (ref.) and an exposure (exp.) site, data can be fit to the regression model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 (x_1 \cdot x_2) \quad (1)$$

where y is the response, x_1 is the covariate, x_2 is an indicator variable for treatment (e.g., 0 for reference and 1 for exposure), and $x_1 \cdot x_2$ is a covariate by treatment interaction term which is equal to the product of the covariate and the indicator variable for each observation. This model fits the data to two regression lines with distinct intercepts and slopes, namely $y = \beta_0 + \beta_1 x_1$ for the reference site and $y = (\beta_0 + \beta_2) + (\beta_1 + \beta_3) x_1$ for the exposure site. A test for parallel regression slopes is equivalent to testing the significance of the coefficient of the $x_1 \cdot x_2$ interaction term (i.e., a test of whether $\beta_3 = 0$). If this coefficient is not significant (at the $\alpha = 0.05$ level of significance), the data can be described by two parallel lines with distinct intercepts. This model is

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2. \quad (2)$$

The test for differences in the response between treatments can proceed with (2). This test is equivalent to testing whether the two regression lines have equal intercepts (i.e., a test of whether $\beta_2 = 0$). If there is no significant difference in response between treatments, the data can be represented by a single regression line without the β_2 term.

Thus, analyzing data using ANCOVA is equivalent to fitting the data to (1) to assess parallel slopes, and testing for differences among sites is equivalent to testing the significance of the β_2 in (2). Comparisons to critical effect sizes are made by comparing the percentage difference in adjusted means (mean response adjusted to factor out differences in the covariate values) to predetermined critical effect sizes. This percentage difference can be easily calculated from (2). The coefficient β_2 in (2) is the vertical distance between the two regression lines (i.e., the difference in intercepts) and can be converted into a percentage difference in the responses variable as

$$\% \text{ difference} = (10^{\beta_2} - 1) \cdot 100\% \quad (3)$$

when the response variable is log-transformed. The adjusted means can be calculated by evaluating (2) using the grand mean of the covariate (average covariate value over all sites) for x_1 and using the appropriate indicator value for x_2 to obtain each adjusted mean if desired.

A1.1 Identifying Immature Fish

- Calculate gonadosomatic index (GSI) = gonad weight / body weight x 100. Immature fish can typically be identified as those with GSI < 1%.
- Plot gonad weight vs. body weight. Immature fish can usually be quickly identified.

Figure A1-1 illustrates a data set with several immature fish. A line representing GSI = 1% is added to help identify immature fish.

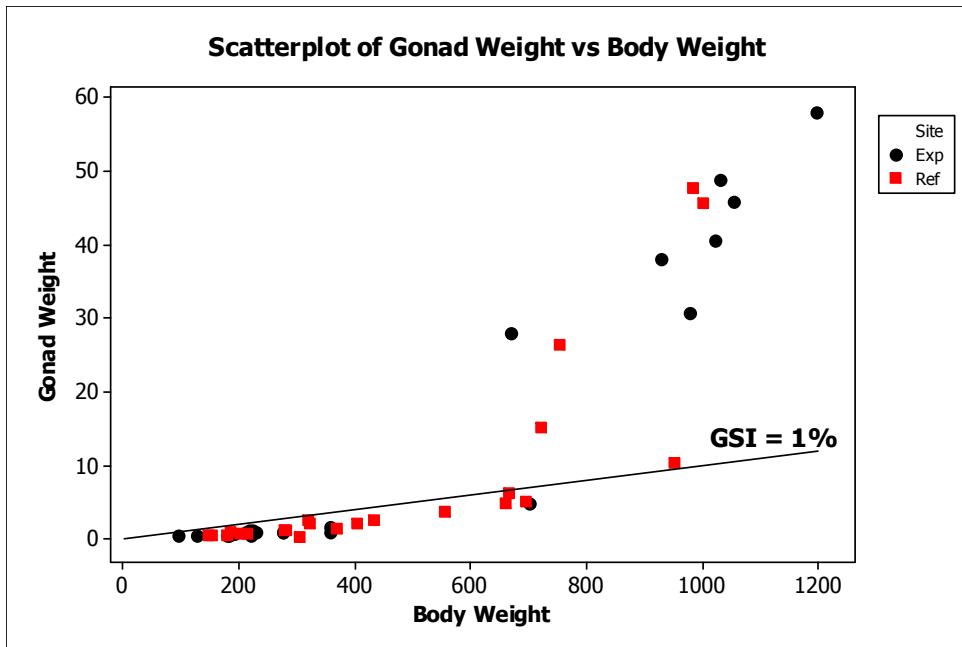


Figure A1-1: A plot of gonad weight vs. body weight for female *Catostomus macrocheilus*. Line represents GSI = 1%

Some fish species do not spawn every year. Some fish will not invest energy into reproduction every year. These species can be easily identified from plots of gonad weight vs. body weight where the data form two different groups corresponding to the spawning fish and non spawning fish. When a line of GSI = 1% is added to the plot, the spawning and non-spawning fish can be easily distinguished. See Figure A1-2.

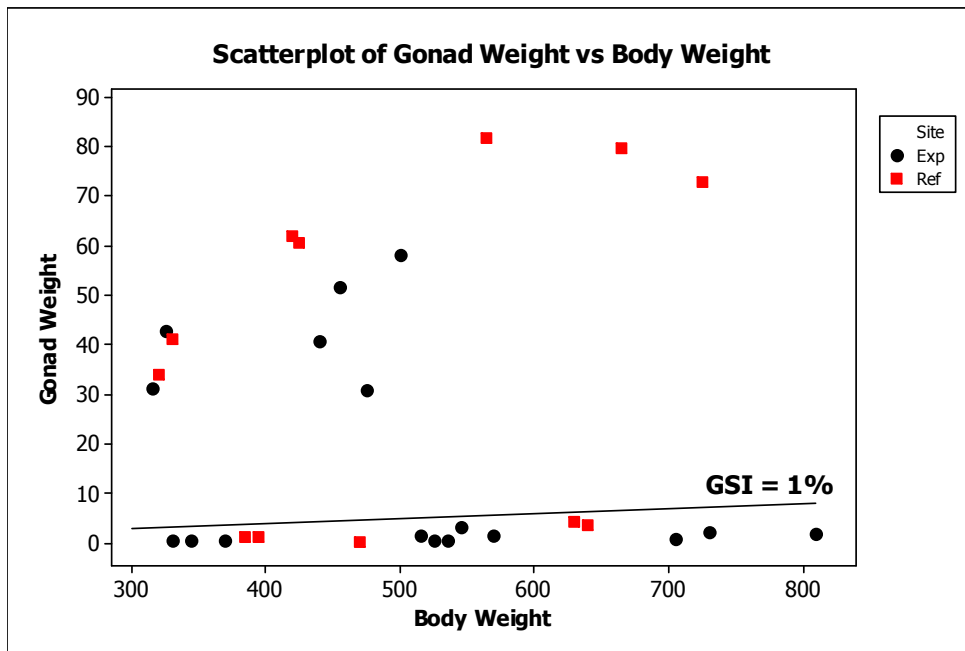


Figure A1-2: A plot of gonad weight vs. body weight for female *Lota lota*. Line represents GSI = 1%

A1.2 Summary Statistics

- Separate data by species, sex and site (e.g., reference or exposure).
- Plot each data set using a box plot and examine for obvious data entry errors or any unusual observations.

Box plots for the length variable for female *Catostomus commersoni* are shown in Figure A1-3. The box plot in A reveals an unusually long fish at the exposure site. A review of field notes and comments in the spreadsheet indicate that this fish was exceptionally longer than all other fish. This observation lies considerably far outside the range of values for “Length” and may be considered an outlier.

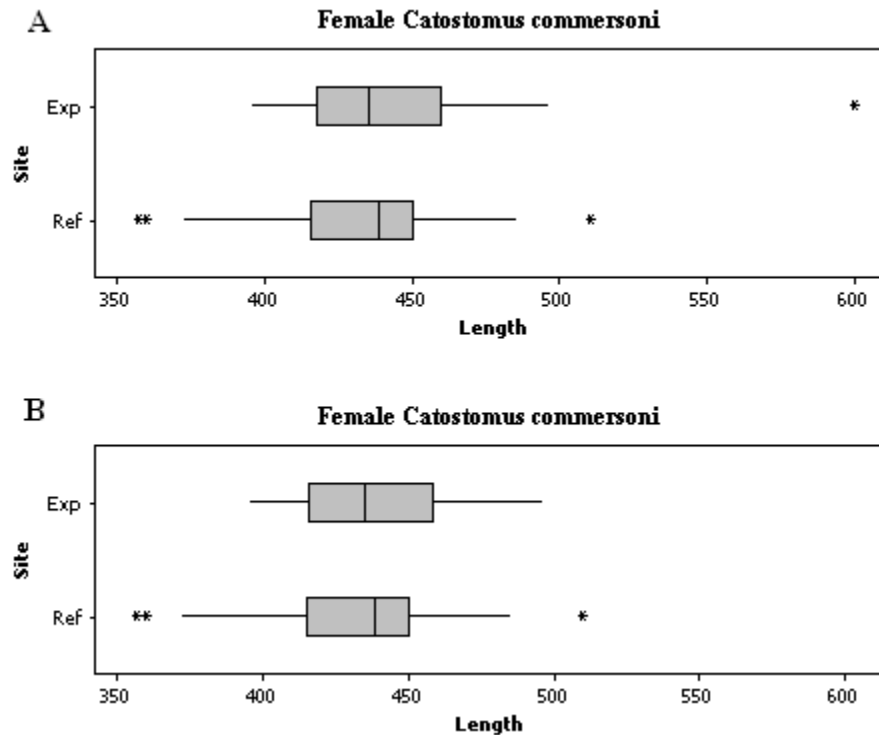


Figure A1-3: Box plots for female *Catostomus commersoni* by site
A. Outlier detected in exposure site.
B. Outlier is removed.

- Calculate and present summary statistics in a table.

Table A1-1: Summary statistics for “Length”

Species	Sex	Site	N	Mean	SD*	SE**	Min	Max
<i>Catostomus commersoni</i>	F	Exp	39	437.49	24.57	3.93	395	496
<i>Catostomus commersoni</i>	F	Ref	40	432.18	31.46	4.97	357	510
<i>Catostomus commersoni</i>	M	Exp	39	405.36	19.72	3.16	367	448
<i>Catostomus commersoni</i>	M	Ref	39	405.00	18.00	2.88	369	448
<i>Etheostoma exile</i>	F	Exp	33	3.7492	0.349	0.0440	3.0	4.5
<i>Etheostoma exile</i>	F	Ref	31	3.7129	0.556	0.0999	2.8	5.2
<i>Etheostoma exile</i>	M	Exp	37	3.5973	0.295	0.0485	3.0	4.1
<i>Etheostoma exile</i>	M	Ref	26	3.5346	0.277	0.0543	3.1	4.1

* Standard deviation

** Standard error

A1.3 Analysis of Variance

- Test all variables for normality.
- Test all variables for homogeneity of variances.
- Provide the statistical tests used and the p-value of the tests.

- If statistical assumptions are seriously violated or are violated and sample sizes are unequal, consider using a non-parametric alternative to analysis of variance (ANOVA) (e.g., Kruskal-Wallis test).
- Provide means (and medians if using non-parametrics) and pooled SD, as well as the test p-value.
- Plot residuals and check for outliers. Observations with studentized residuals of magnitude greater than 4 warrant investigation and potential removal. If any outliers are removed, provide both an analysis with all data and one with outlier(s) removed.

“Weight” - female *Catostomus commersoni*

Sample sizes = 40 (ref) and 39 (exp)

Normality (tested using Anderson-Darling test)

Female *Catostomus commersoni* exposure fish p-value = 0.257

Female *Catostomus commersoni* reference fish p-value = 0.340

Homogeneity of variances (tested using Levene’s test)

Female *Catostomus commersoni* fish p-value = 0.329

Statistical assumptions are met, therefore proceed with analysis of variance

Response: Weight

Factor: Site (exp, ref)

Results:

Exposure mean weight: 1274.6

Reference mean weight: 1255.6 p-value = 0.735

Pooled SD = 248.6

No unusual observations

“Age” - female *Catostomus commersoni*

Sample sizes = 40 (ref) and 39 (exp)

Normality (tested using Anderson-Darling test)

Female *Catostomus commersoni* exposure fish p-value = 0.056

Female *Catostomus commersoni* reference fish p-value < 0.005

Homogeneity of variances (tested using Levene’s test)

Female *Catostomus commersoni* fish p-value = 0.788

Assumption of normality was not met for reference fish. Sample sizes are 40 (ref) and 39 (exp). The sample sizes are approximately equal and the assumptions are not strictly

violated. Either the parametric ANOVA or a non-parametric alternative to ANOVA may be used. Here we use the non-parametric Kruskal-Wallis test.

Response: Age

Factor: Site (exp, ref)

Results:

Exposure median age: 10

Reference median age: 8

p-value = 0.001

“Length” - female *Catostomus commersoni*

Residual plot – studentized residuals vs. order (order data are entered in spreadsheet)

Outliers are typically regarded as observations with magnitude > 4 and can be easily identified in this plot.

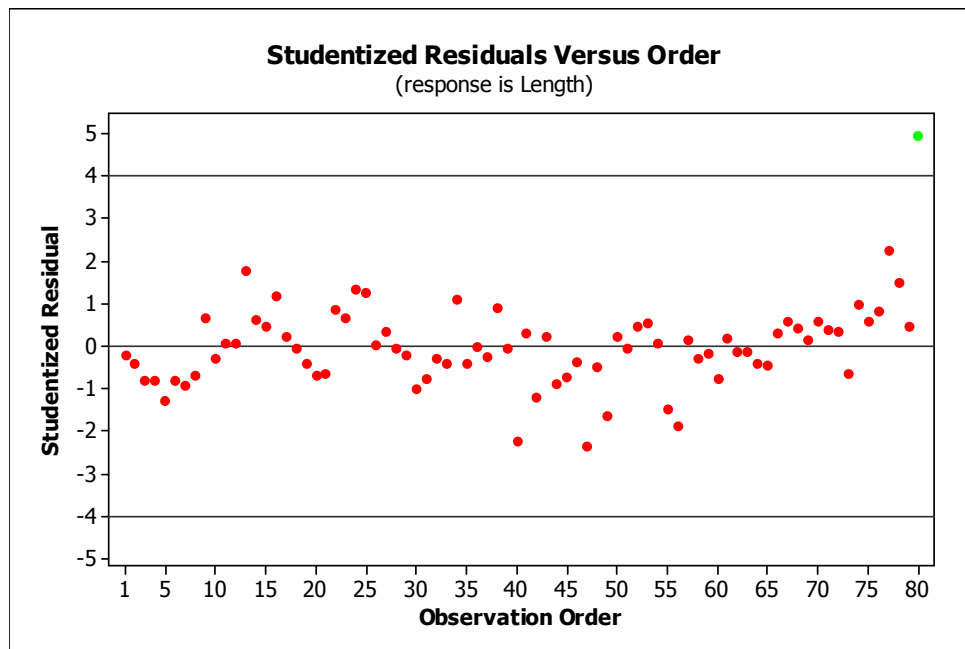


Figure A1-4: A plot of studentized residual vs. observation order (in spreadsheet) for the ANOVA on length for female *Catostomus commersoni*

A1.4 Analysis of Covariance (ANCOVA)

- Plot the response variable vs. covariate for all sites.
- Inspect plot for a linear trend and appropriate overlap of covariate values.
- Inspect plot for outliers—calculate studentized residuals from ANCOVA model.
- Consider removing outliers with magnitude > 4 (studentized residual).
- Test residuals for normality (each regression line).
- Test residuals for homogeneity of variances (among regression lines).

- Test homogeneity of regression slopes—fit data to regression model with interaction term and test significance of interaction term. Provide coefficient of determination “ R^2 ” for the regression model.
- Test for differences in the response—fit data to regression model without interaction term and test significant of the site (treatment) term. Provide coefficient of determination “ R^2 ” for the regression model and the pooled SD (of the residuals).
- Provide adjusted means for each site. Also take the anti-log of the mean if log-transformed data were used.
- Calculate the percent difference, calculated as a percent of the reference site (using anti-logs of adjusted means).

“Condition” - male *Rhinichthys cataractae*

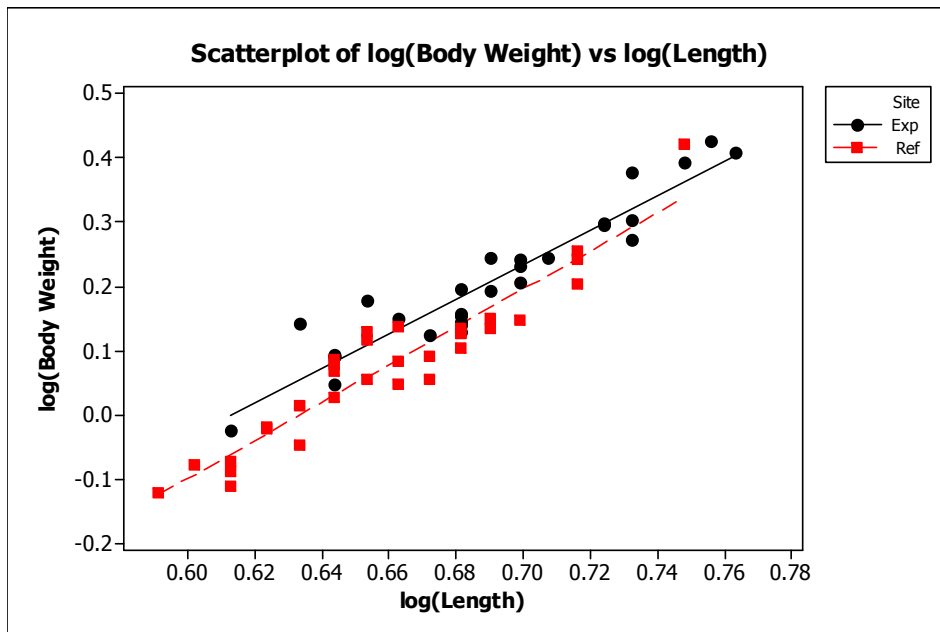


Figure A1-5: A plot of log(body weight) vs. log(length) for male *Rhinichthys cataractae*. Data are fit to two distinct regression lines, one for each site

Overlap of covariate values seems appropriate and there is a linear trend.

Sample sizes = 32 (ref) and 30 (exp)

Normality (tested using Anderson-Darling test)

Male *Rhinichthys cataractae* exposure residuals p-value = 0.262

Male *Rhinichthys cataractae* reference residuals p-value = 0.138

Homogeneity of variances (tested using Levene’s test)

Male *Rhinichthys cataractae* residuals p-value = 0.733

Homogeneity of regression slopes

Data fit to $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 (x_1 \cdot x_2)$ $R^2 = 0.9212$

β_3 not significant (p-value = 0.337), thus there is no evidence of non-parallel slopes.

Test for differences in the response

Data fit to $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$ $R^2 = 0.9199$

β_2 is significant (p-value = 0.0001), thus there is a significant difference in weight between sites.

Adjusted mean for reference weight: 1.3113 g

Adjusted mean for exposure weight: 1.4496 g

(Means are anti-logged to obtain original units when log-transformed—the anti-log of x is 10^x if the transformation was log base 10.)

Pooled SD = 0.0420164

Percent difference = 10.54% (calculated as percent of reference using adjusted means)

A1.5 Non-parallel Slopes in Analysis of Covariance

- Method 1

“Relative gonad weight” – male *Catostomus commersoni*

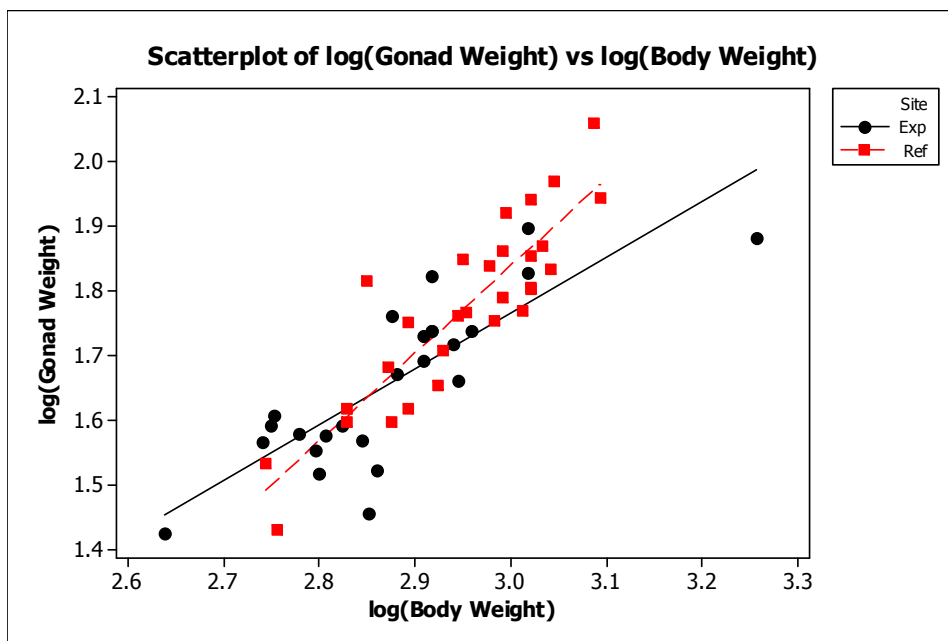


Figure A1-6: A plot of log (gonad weight) vs. log(body weight) for male *Catostomus*

commersoni. Data are fit to two distinct regression lines, one for each site

Overlap of covariate values seems appropriate and there is a linear trend. One observation warrants investigation in the exposure group.

A plot of the studentized residuals does not reveal any observations with extremely large magnitudes. See Figure A1-7.

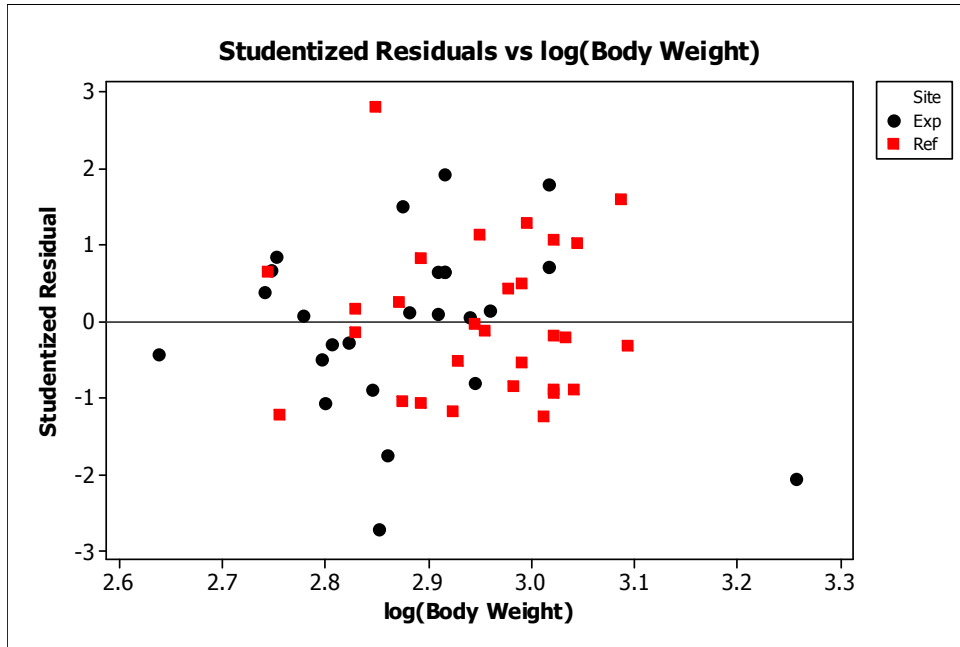


Figure A1-7: A plot of studentized residual vs. log(body weight) for male *Catostomus commersoni* data fit to the interaction model $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 (x_1 \cdot x_2)$

Sample sizes = 29 (ref) and 25 (exp)

Normality (tested using Anderson-Darling test)

Male *Catostomus commersoni* exposure residuals p-value = 0.543

Male *Catostomus commersoni* reference residuals p-value = 0.176

Homogeneity of variances (tested using Levene's test)

Male *Catostomus commersoni* residuals p-value = 0.882

Homogeneity of regression slopes

Data fit to $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 (x_1 \cdot x_2)$ $R^2 = 0.7710$

β_3 significant (p-value = 0.014), thus there is evidence of non-parallel slopes.

Assess influence by plotting Cook's distance vs. the covariate.

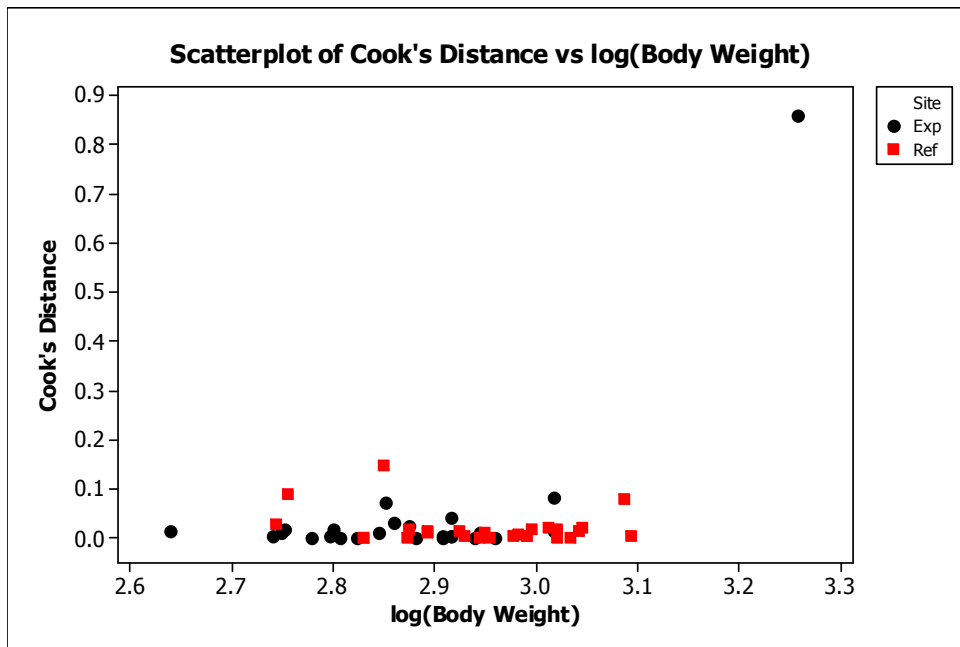


Figure A1-8: A plot of Cook's distance vs. log(body weight) for male *Catostomus commersoni* data fit to the interaction model $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 (x_1 \cdot x_2)$

One observation in the exposure group has a large Cook's distance. Remove and test assumptions again.

Sample sizes = 29 (ref) and 24 (exp)

Normality (tested using Anderson-Darling test)

Male *Catostomus commersoni* exposure residuals p-value = 0.408

Male *Catostomus commersoni* reference residuals p-value = 0.176

Homogeneity of variances (tested using Levene's test)

Male *Catostomus commersoni* residuals p-value = 0.852

Homogeneity of regression slopes

Data fit to $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 (x_1 \cdot x_2)$ $R^2 = 0.7841$

β_3 not significant (p-value = 0.205), thus there is no evidence of non-parallel slopes.

Continue with procedure.

- **Method 2**

“Condition” - male *Catostomus catostomus*

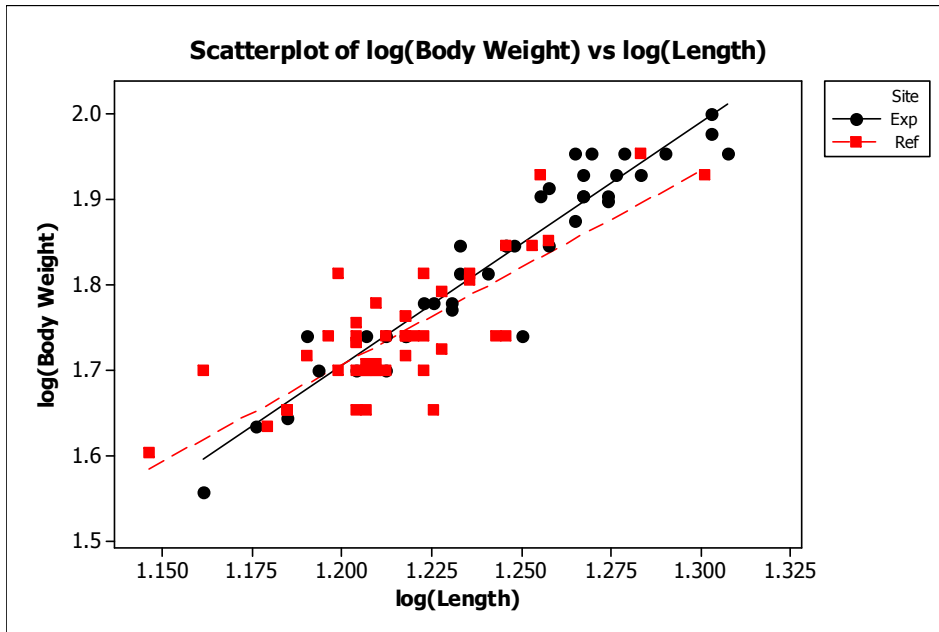


Figure A1-9: A plot of log(body weight) vs. log(length) for male *Catostomus catostomus*. Data are fit to two distinct regression lines, one for each site

Homogeneity of regression slopes

Data fit to $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 (x_1 \cdot x_2)$ $R^2 = 0.8530$

β_3 significant (p-value = 0.036), thus there is evidence of non-parallel slopes, but $R^2 > 0.8$, thus fit parallel model and compare coefficients of determination.

Data fit to $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$ $R^2 = 0.8450$

R^2 for parallel model is also > 0.8 and is less than 0.02 (i.e. 2 percentage points) less than R^2 for interaction model. Thus use parallel model to describe data and continue with analysis.

- Method 3

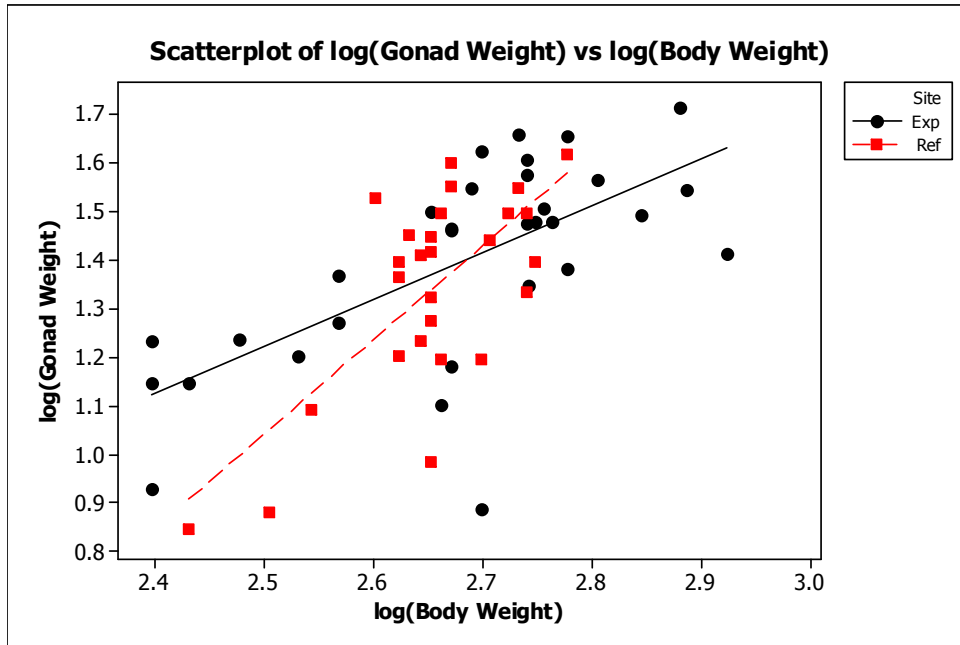


Figure A1-10a: A plot of log(gonad weight) vs. log(body weight) for male *Catostomus catostomus*. Data are fit to two distinct regression lines, one for each site

Homogeneity of regression slopes

Data fit to $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 (x_1 \cdot x_2)$ $R^2 = 0.4695$

β_3 significant (p-value = 0.036), a plot of Cook's distance vs. the covariate reveals no influential points, and $R^2 < 0.8$ thus application of method 2 cannot be attempted.

- Determine the maximum and minimum values of the range of the covariate for each site.
- Calculate the predicted values of the response for each site (regression line) at these two values of the covariate.
- Calculate a percentage difference (calculated as exposure – reference, expressed as a percentage of reference) at the two values of the covariate.

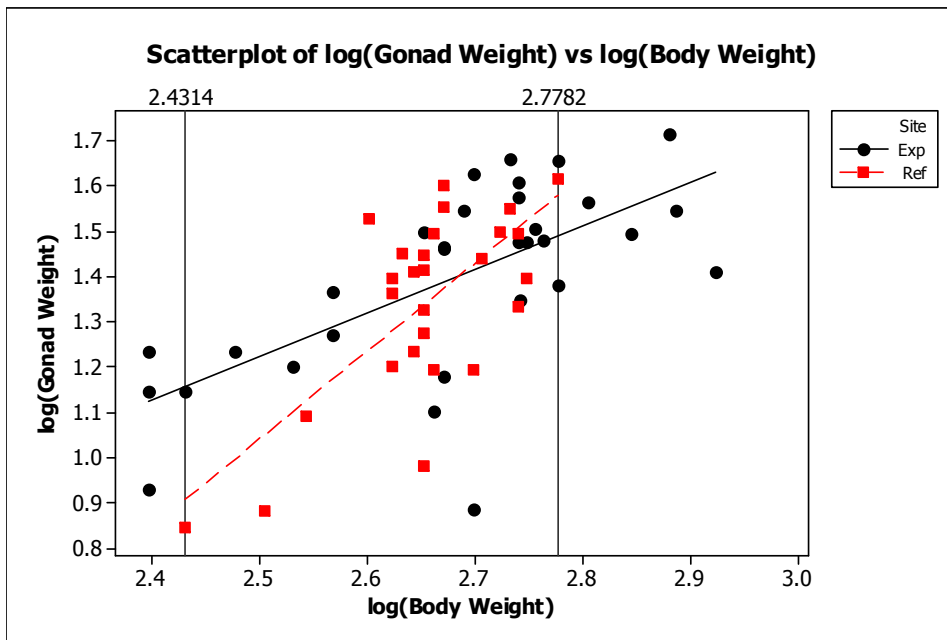


Figure A1-10b: The data from Figure 10a but with the minimum and maximum values of the range of overlap of the covariate between sites identified

Covariate values 2.4314 and 2.7782

For 2.4314: predicted values for the response are 1.0949 (ref) and 1.1544 (exp).

For 2.7782: predicted values for the response are 1.4963 (ref) and 1.4883 (exp).

Thus percent differences are calculated to be (after taking the anti-log of the predicted response values) 14.69% and -1.84% for the covariate values of 2.4314 and 2.7782, respectively. These will be the estimates of the effects for smaller and larger fish, respectively, and can be compared to a critical effect size.

A1.6 Non-parametric ANCOVA

“Relative gonad weight” – female *Catostomus commersoni*

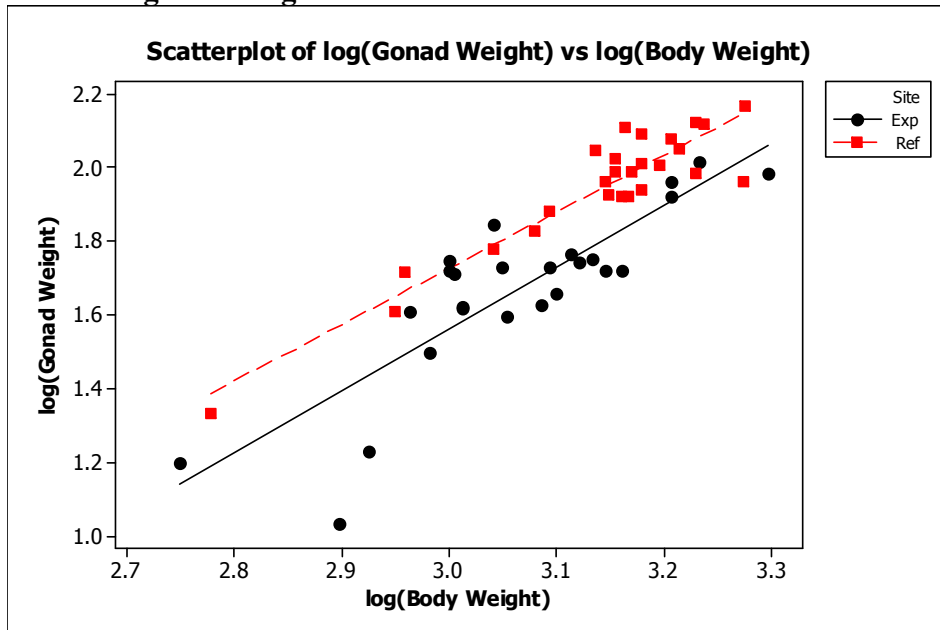


Figure A1-11: A plot of log(gonad weight) vs. log(body weight) for female *Catostomus commersoni*. Data are fit to two distinct regression lines, one for each site

Distribution of covariate values for two sites are not very similar.

Sample sizes = 26 (ref) and 25 (exp)

Normality (tested using Anderson-Darling test)

Female *Catostomus commersoni* exposure residuals

p-value = 0.476

Female *Catostomus commersoni* reference residuals

p-value = 0.596

Homogeneity of variances (tested using Levene's test)

Female *Catostomus commersoni* residuals

p-value = 0.024

- Only the assumption of homogeneity of variances is not met—sample sizes are almost equal, so parametric ANCOVA could be used—or the non-parametric ANCOVA on the ranks of the data.

Non-parametric ANCOVA on the ranks

Response: Gonad weight ranks

Covariate: Body weight ranks

Homogeneity of regression slopes

Data fit to $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 (x_1 \cdot x_2)$

$R^2 = 0.8299$

β_3 not significant (p-value = 0.364), thus there is no evidence of non-parallel slopes.

Test for differences in the response

Data fit to $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$ $R^2 = 0.8268$

β_2 is significant (p-value < 0.0001), thus there is a significant difference in gonad weight between sites.

Comparisons to critical effect sizes can sometimes be made by calculating a percentage difference using the adjusted mean ranks. This percent difference is simply the difference in adjusted mean ranks, calculated as exposure – reference, expressed as a percentage of the reference adjusted mean rank. The adjusted mean ranks can be calculated by evaluating Equation 2 (in the statistical background discussion at the beginning of this appendix) using the mean covariate rank for x_1 and using the appropriate indicator value for x_2 if the regression approach to ANCOVA is being used.

Adjusted mean for reference gonad weight rank: 30.8660

Adjusted mean for exposure gonad weight rank: 20.9394

Pooled SD = 6.31325 (ranks)

Percent difference = -31.16% (calculated as percent of reference using adjusted means using ranks)

Note: parametric ANCOVA will give a percent difference of -28.90%.

A1.7 Issues with the Range of the Covariate

Range of covariate values not similar between sites

- Look for a subset of the data where there is good overlap in the covariate values for each site. For example consider the data set for male *Pleuronectes americanus* relative gonad weight in Figure A1-12a. The range of covariate for the reference and exposure site is quite different where the reference has several smaller fish. We can take a subset of the data (exclude fish with $\log(\text{length}) < 1.375$) and obtain a data set with similar ranges of the covariate with good overlap. The analysis can be performed on this subset of the data (data set illustrated in Figure A1-12b). An analysis with all the data may be performed for comparison purposes but caution should be used in interpreting the results of the analysis using all the data.

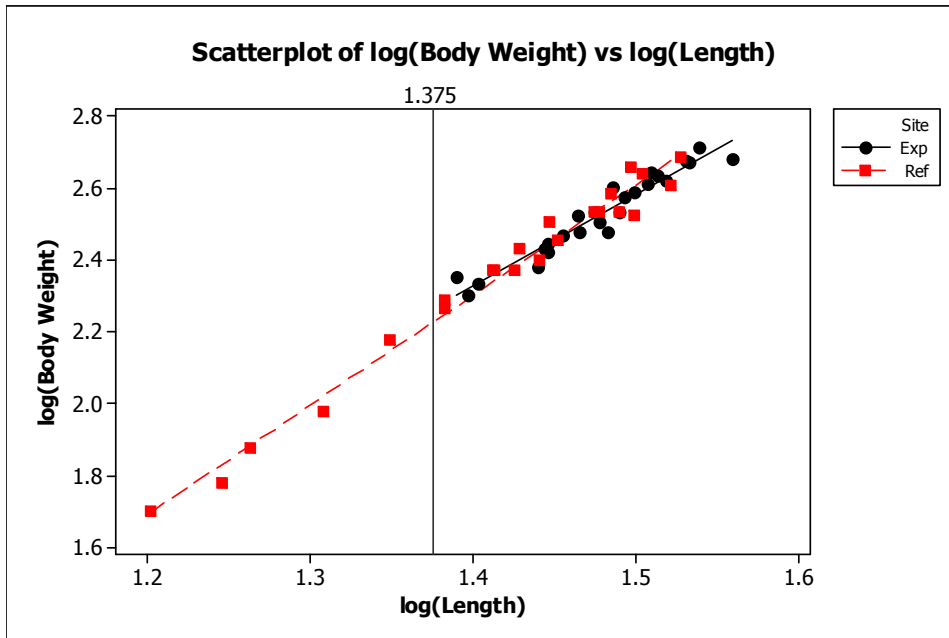


Figure A1-12a: A plot of log(body weight) vs. log(length) for male *Pleuronectes americanus*. Data are fit to two distinct regression lines, one for each site

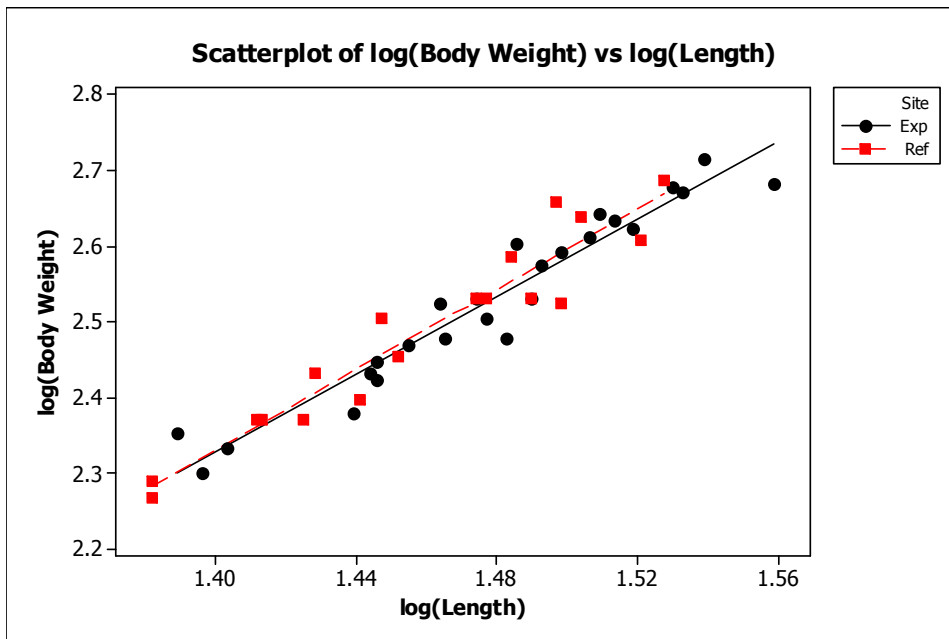


Figure A1-12b: A plot of log(body weight) vs. log(length) for male *Pleuronectes americanus*. A subset of the data in Figure 12a using only fish with log(length) > 1.375

Covariate observed only at a few values

- Figure A1-13 is an example of a data set where the covariate is only observed at a few values of the covariate. These data sets are typical for weight-at-age analyses

for small-bodied fish but may arise with other data sets. ANCOVA may be inappropriate.

- Perform a one-way ANOVA on body weight (factor: site) for fish aged 1.
- Perform a one-way ANOVA on body weight (factor: site) for fish aged 2.
- If sample sizes for an age group are too small for analysis, provide means and sample sizes.



Figure A1-13: A plot of log(body weight) vs. age for female *Fundulus heteroclitus*. Data are fit to two distinct regression lines, one for each site

A1.8 A priori Power Analyses

“Age” female *Perca flavescens*

We would like to determine what sampling effort is required to detect a 25% difference in age for female *Perca flavescens*. The following data are available from the fish survey from the previous cycle at the same mill. See section 8.6.2.1 for further explanations and definition of terms.

Species	Sex	Site	N	Mean	SD	SE	Min	Max
<i>Perca flavescens</i>	F	Exposure	30	4.100	1.094	0.200	3	8
<i>Perca flavescens</i>	F	Reference	29	3.759	1.300	0.241	2	7

- Suppose the probability of type I error (α) and the probability of type II error (β) are chosen to be 0.05 and 0.2, respectively (this is for illustrative purposes only; for most cases in the EEM program, type I and type II error should be set equal; $\alpha = \beta$).

- The coefficient of variation for the reference site can be calculated to be
 $COV = 1.300/3.759 \cdot 100 = 34.58\%$.
- Our critical effect size (CES) is 25%.
- We will start with $n = 20$ and solve the following iteratively for estimated n ($\hat{n}$)

$$\hat{n} = 2(t_{\alpha} + t_{\beta})^2 (COV/CES)^2$$

- Using $n = 20$, $\alpha = 0.05$, and $\beta = 0.2$ we obtain $t_{\alpha} = 2.093$ and $t_{\beta} = 0.861$
[t_{α} calculated as two-tailed with $(n-1)$ df, t_{β} calculated as one-tailed with $(n-1)$ df]
- $\hat{n} = 2(2.093 + 0.861)^2 (34.58/25)^2 = 33.39 = 34$
- Using $n = 34$, $\alpha = 0.05$, and $\beta = 0.2$ we obtain $t_{\alpha} = 2.035$ and $t_{\beta} = 0.853$
- $\hat{n} = 2(2.035 + 0.853)^2 (34.58/25)^2 = 31.9 = 32$
- Using $n = 32$, $\alpha = 0.05$, and $\beta = 0.2$ we obtain $t_{\alpha} = 2.040$ and $t_{\beta} = 0.853$
- $\hat{n} = 2(2.040 + 0.853)^2 (34.58/25)^2 = 32.03 = 32$
- $\hat{n} = n = 32$

Approximately 32 female *Perca flavescens* will be needed from each site (reference and exposure) to detect a difference of 25% in age.

“Relative gonad weight” female *Perca flavescens*

We would like to determine what sampling effort is required to detect a 25% difference in relative gonad weight for female *Perca flavescens*. The following results are available from the ANCOVA from the previous cycle at the same mill. See section 8.6.2.1 for further explanations and definition of terms.

- Sample sizes: 29 (ref), 30 (exp)
- Pooled SD (of residuals) using log transformed data = 0.0743033 (this is also equal to the square root of the mean square error term obtained from fitting the data to the parallel slope ANCOVA model).
- Suppose the probability of type I error (α) and the probability of type II error (β) are chosen to be 0.05 and 0.2, respectively (this is for illustrative purposes only; for most cases in the EEM program, type I and type II error should be set equal; $\alpha = \beta$).
- $SD_Z = 0.0743033$
- $CES_Z = \log(0.25 + 1) = \log(1.25) = 0.09691$
- We will start with $n = 20$ and solve with following iteratively for estimated n ($\hat{n}$)
$$\hat{n} = 2(t_{\alpha} + t_{\beta})^2 (SD_Z / CES_Z)^2$$
- Using $n = 20$, $\alpha = 0.05$, and $\beta = 0.2$ we obtain $t_{\alpha} = 2.093$ and $t_{\beta} = 0.861$
[t_{α} calculated as two-tailed with $(n-1)$ df, t_{β} calculated as one-tailed with $(n-1)$ df]
- $\hat{n} = 2(2.093 + 0.861)^2 (0.0743033/0.09691)^2 = 10.26 = 11$

- Using $n = 11$, $\alpha = 0.05$, and $\beta = 0.2$ we obtain $t_{\alpha} = 2.228$ and $t_{\beta} = 0.879$
- $\hat{n} = 2(2.228 + 0.879)^2(0.0743033/0.09691)^2 = 11.34 = 12$
- Using $n = 12$, $\alpha = 0.05$, and $\beta = 0.2$ we obtain $t_{\alpha} = 2.201$ and $t_{\beta} = 0.876$
- $\hat{n} = 2(2.201 + 0.876)^2(0.0743033/0.09691)^2 = 11.13 = 12$
- $\hat{n} = n = 12$

Approximately 12 female *Perca flavescens* will be needed from each site (reference and exposure) to detect a difference of 25% in relative gonad weight.

A1.9 Post hoc Power Analyses

“Condition” female *Catostomus commersoni*

In this example, a non-significant result is obtained for the condition effect endpoint for female *Catostomus commersoni*. An example of a post hoc power analysis is performed to determine the power of the test to detect the CES. We are given the following output from the ANCOVA using log (body weight) as the response variable and log(length) as a covariate. See section 8.6.2.2 for further explanations and definition of terms.

Source	Sum-of-Squares	Degrees of Freedom	Mean-Square	F-Ratio	p-value
log(length)	0.121190	1	0.119427	119.32	<0.001
Site	0.000046	1	0.000046	0.05	0.831
Error	0.027025	27	0.001001		
Total	0.148261	29			

- The CES for condition is 10% of the reference mean (converted to CES_Z in the following formula), and the probability of type I error (α) initially used for the above ANCOVA in this example was 0.05 (0.831 is greater than 0.05, so the exposure vs. reference comparison was declared as being non-significant).
- The power formula is
$$t_{\beta} = \sqrt{\frac{n}{2(SD_Z / CES_Z)^2}} - t_{\alpha}$$
- $SD_Z = \sqrt{MSE} = \sqrt{0.001001} = 0.0316385$
- $CES_Z = \log(f+1)$, where f = CES represented as a fraction of the reference mean
 So $CES_Z = \log(0.1+1) = \log(1.1) = 0.0413926$
- $n = 15$ for each site, thus $t_{\alpha} = 2.145$

- $t_{\beta} = \sqrt{\frac{15}{2(0.0316385/0.0413926)^2}} - 2.045 = 1.538$
- $t_{\beta} = 1.538$ corresponds to $\beta = 0.1486$
- Power = $1 - \beta = 0.8514$

The test had a moderate level of power (Power = 0.8514) to detect a difference of 10%, although the type II error ($\beta = 0.1486$) was not low enough to be equal to type I error ($\alpha = 0.05$), and the EEM program recommendation is that α should be set equal to β (risk to industry set equal to risk to the environment). Thus, preferably a higher α value should have been used for this ANCOVA before declaring non-significance, so that $\alpha = \beta$. In this particular case, the ANCOVA p-value of 0.831 was quite high, so the exposure vs. reference comparison would still have been declared non-significant, even if α had been set as high as $\beta = 0.1486$ ($p = 0.831 > 0.1486$). Rerunning the power analyses at higher α levels would result in lower β levels. So further post hoc power analysis would not be necessary in this case to be confident with declaring non-significance. Future monitoring efforts at this facility should use some combination of greater sample sizes and/or higher α values, so as to ensure sufficiently high power to detect the CES of interest. Thus, the study proposal for the next round of monitoring should include appropriate a priori power analyses.

Appendix 2: Graphical and Tabular Representation of Data

List of Figures:

- Figure A2-1:** Decisional flow chart outlining the various processes data should go through for fish and benthic effect endpoints and linking these to tabular and graphical examples present in this appendix
- Figure A2-2:** Box plots of descriptive statistics for age by fish species and sex
- Figure A2-3:** Analysis of Variance (ANOVA) results of mean age of fish taken from reference and exposure areas (mean and standard error)
- Figure A2-4:** Linear regression of fish liver weight at body weight as an example of effect summary for liver weight or gonad size
- Figure A2-5:** Descriptive statistics for benthic invertebrate total density using a control/impact design
- Figure A2-6:** Analysis of Variance (ANOVA) results of benthic invertebrate total density using a control/impact design
- Figure A2-7:** Plot of benthic invertebrate total density vs. distance from diffuser using a simple gradient design

List of Tables:

- Table A2-1:** Descriptive statistics for age by fish species and sex
- Table A2-2:** Analysis of Variance (ANOVA) results for fish age by species and sex
- Table A2-3:** Analysis of Covariance (ANCOVA) results for liver weight at body weight by sex and by species
- Table A2-4:** Fish result summary table
- Table A2-5:** Descriptive statistics for total benthic invertebrate density
- Table A2-6:** Analysis of Variance (ANOVA) results for benthic invertebrate total density
- Table A2-7:** Summary of all benthic invertebrate descriptive statistics
- Table A2-8:** Summary table of all benthic invertebrate results
- Table A2-9:** Overall summary of site effects

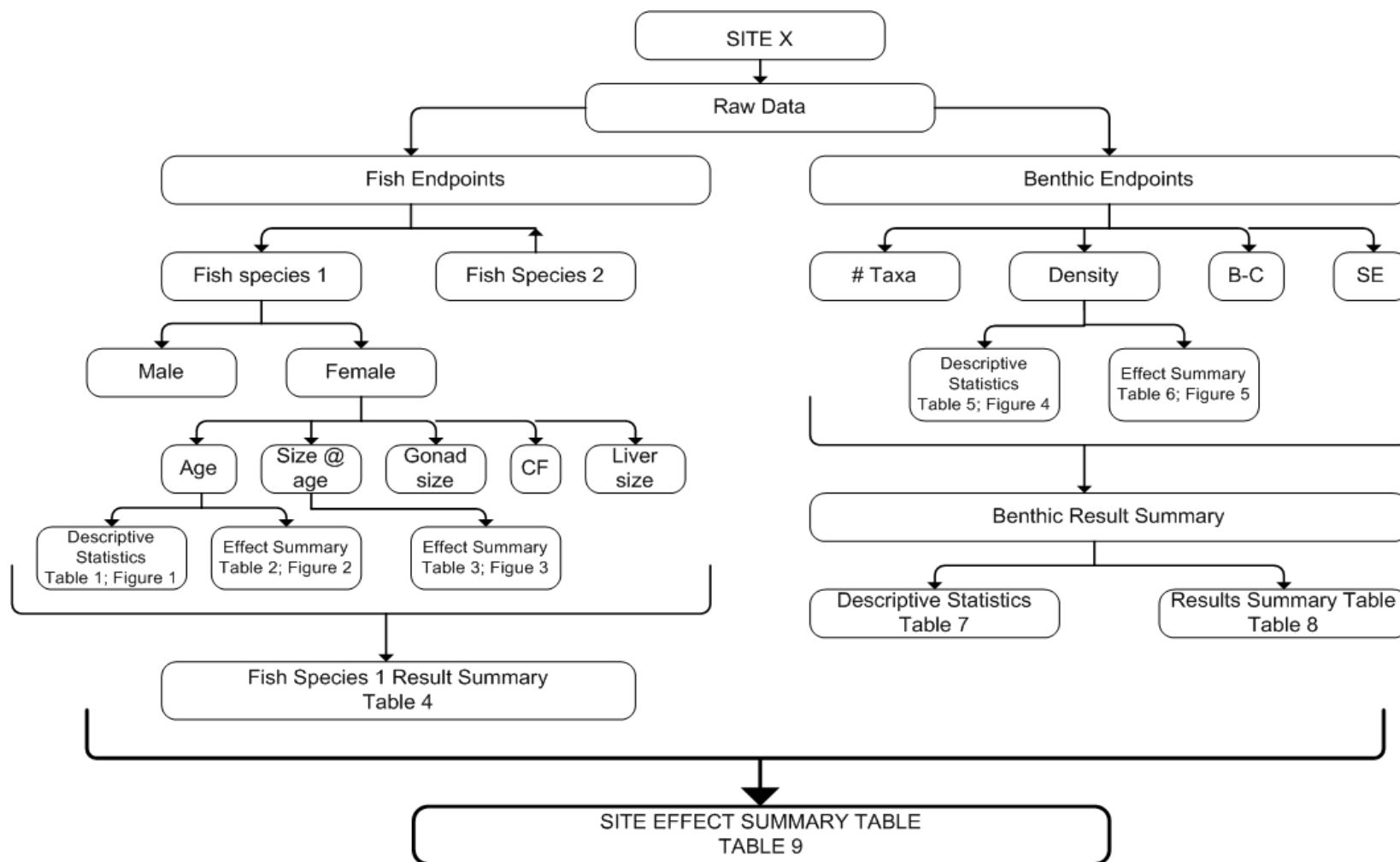


Figure A2-1: Decisional flow chart outlining the various processes data should go through for fish and benthic effect endpoints and linking these to tabular and graphical examples present in this appendix

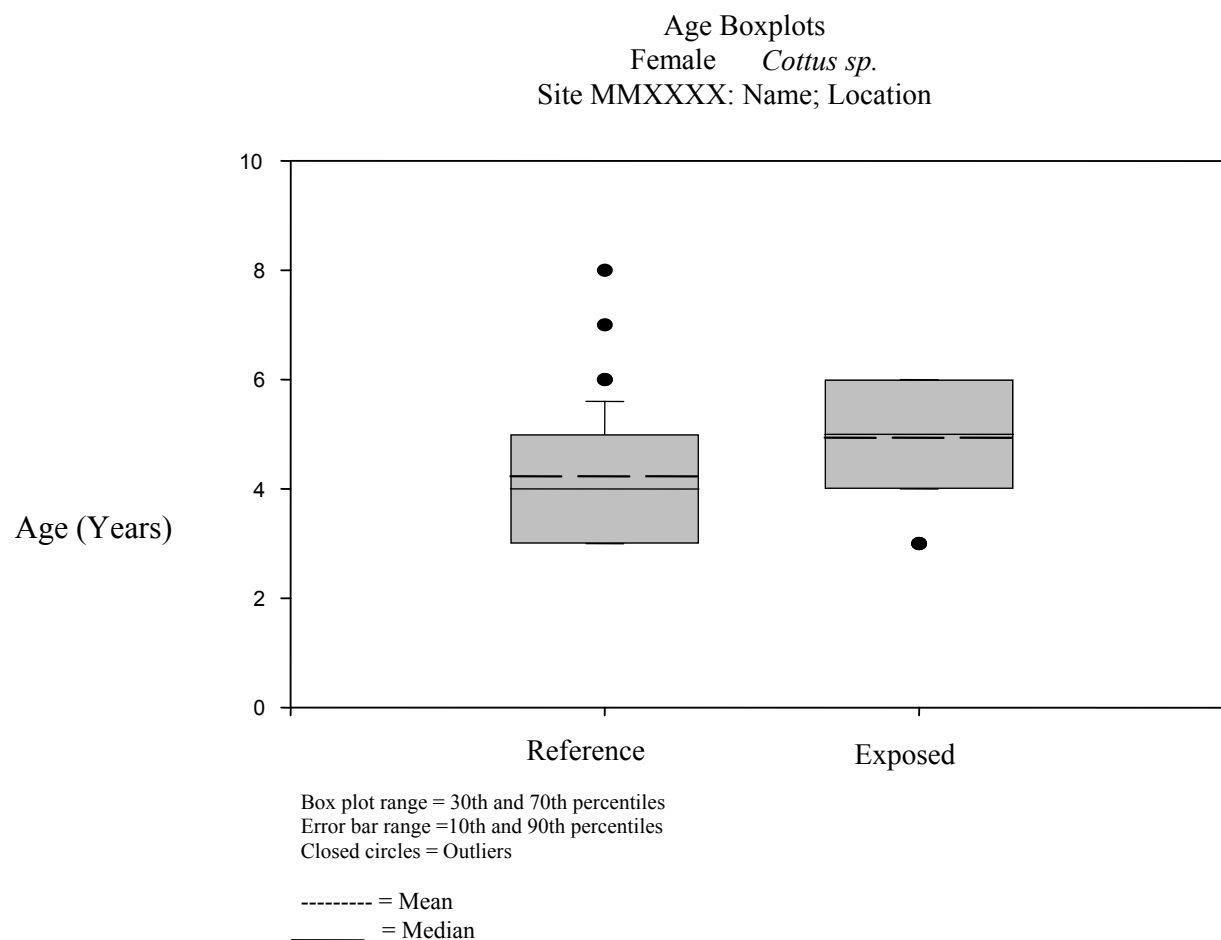


Figure A2-2: Box plots of descriptive statistics for age by fish species and sex

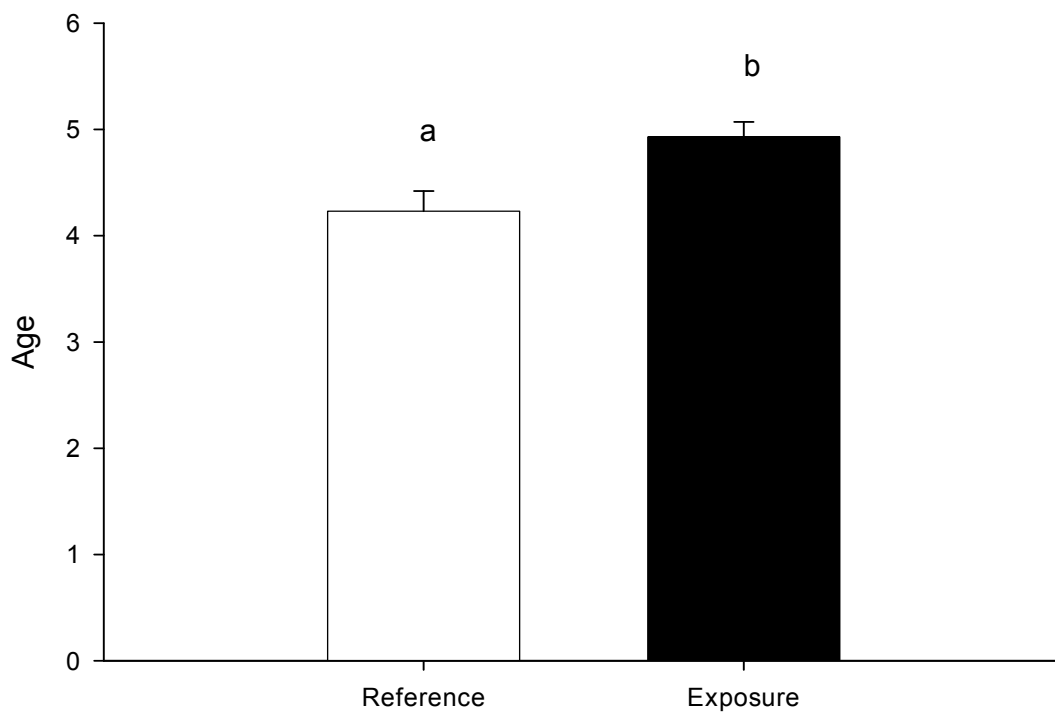


Figure A2-3: Analysis of Variance (ANOVA) results of mean age of fish taken from reference and exposure areas (mean and standard error)

Note: Bars with different letters are significantly different. The vertical bar is a mean and the horizontal variance bars represent the standard errors.

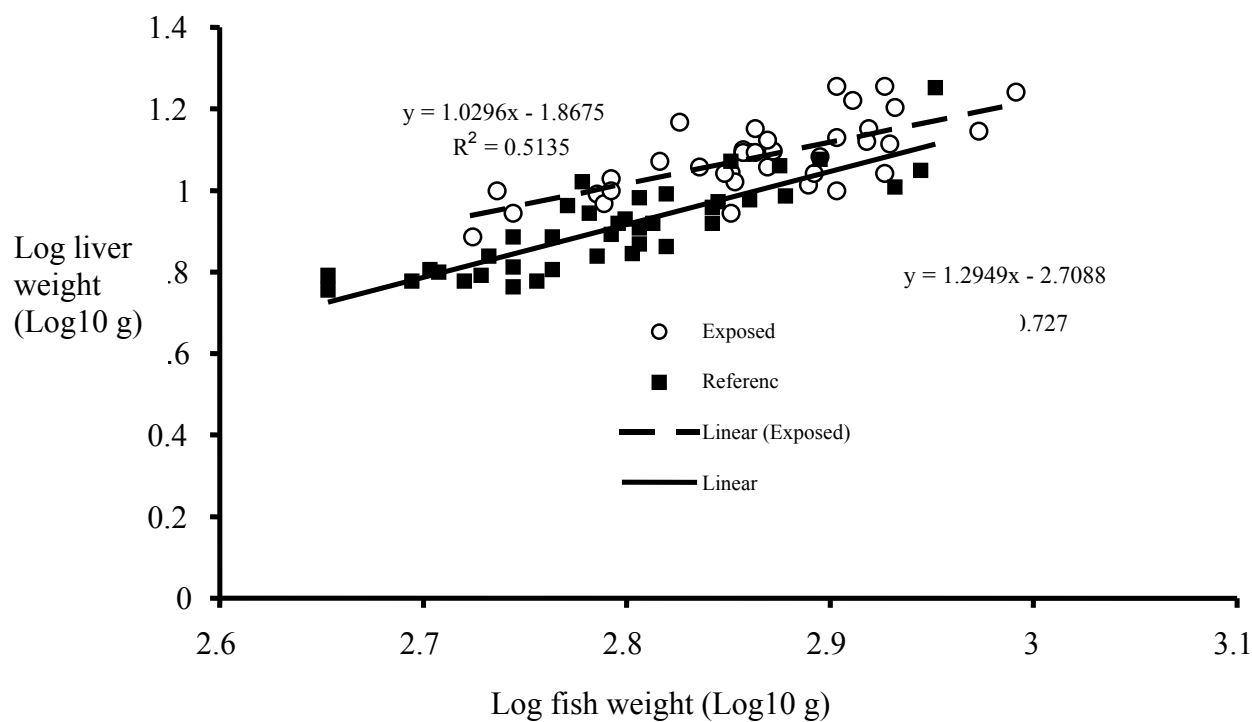


Figure A2-4: Linear regression of fish liver weight at body weight as an example of effect summary for liver weight or gonad size – male *Catostomus* sp.

Examples: Control/Impact Design

Total density benthic invertebrates
Site MMXXXX: Name; Location

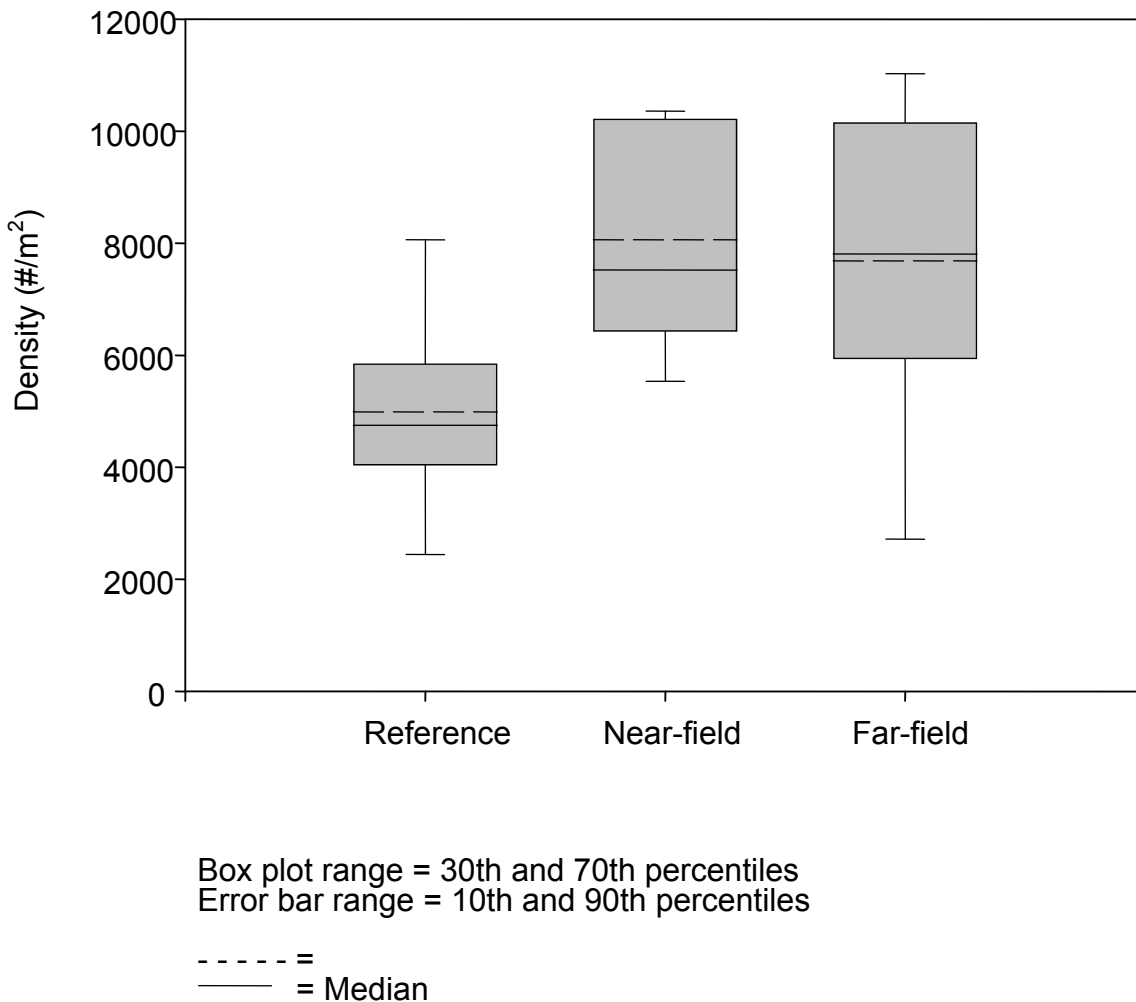


Figure A2-5: Descriptive statistics for benthic invertebrate total density using a control/impact design

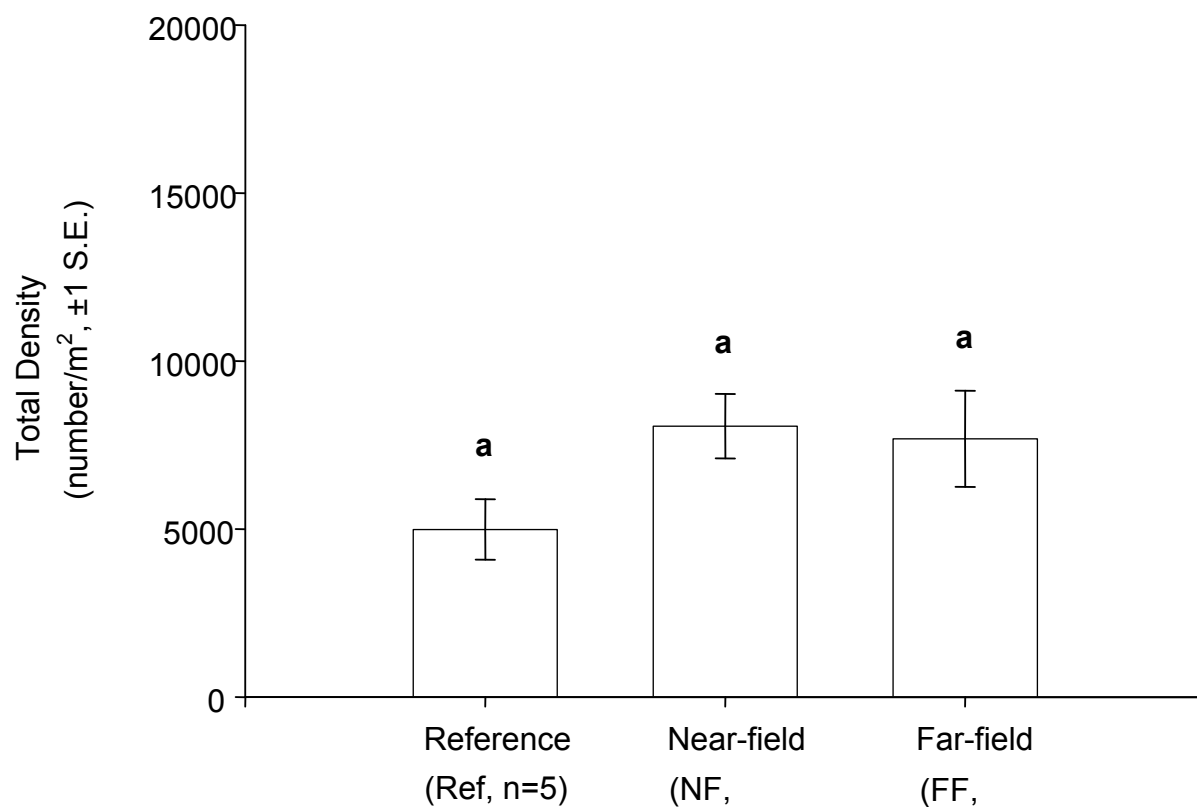


Figure A2-6: Analysis of Variance (ANOVA) results of benthic invertebrate total density using a control/impact design

Note: Bars with the same letters are not significantly different. Values reported are means and associated standard errors.

Example : Simple Gradient Design

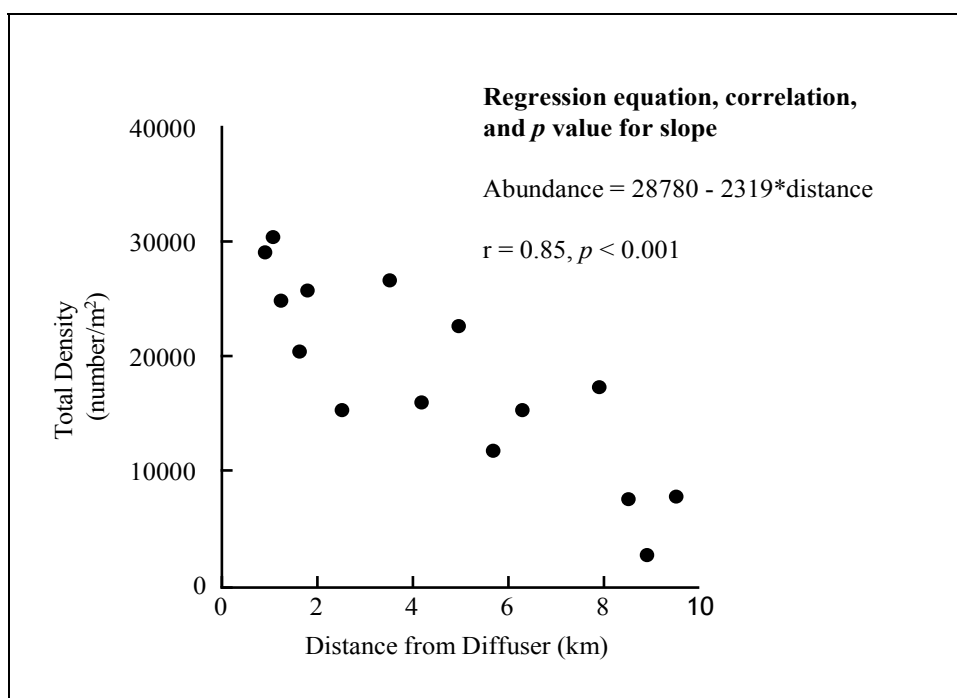


Figure A2-7: Plot of benthic invertebrate total density vs. distance from diffuser using a simple gradient design

Table A2-1: Descriptive statistics for age by fish species and sex**Site MMXXXX; Name; Location****Female *Cottus sp.* – Descriptive Statistics – Age**

Location	Mean	SD*	SE**	(n)	Max.	Min.
Reference	4.23	1.16	0.19	39	8.00	3.00
Exposure	4.93	0.93	0.14	46	6.00	3.00

* Standard deviation

** Standard error

Table A2-2: Analysis of Variance (ANOVA) results for fish age by species and sex**Site MMXXXX; Name; Location****Female *Cottus sp.* – ANOVA Results – Age**

Source of Variation	Sum of Squares (SS)	Degrees of freedom (df)	Mean Square (MS)	F-Ratio	p-value	sig. at p < 0.05
Between groups	0.072624	1	0.072624	11.25004	0.001202	Yes
Within groups	0.535802	83	0.006455			
Total	0.608426	84				

Table A2-3: Analysis of Covariance (ANCOVA) results for liver weight at body weight by sex and by species**Site: MMXXXX; Name; Location****ANCOVA Results: Liver Weight at Body Weight – Male *Catostomus* sp.**

Area	N	Log-transformed		R-Squared (R ²)	Slopes Different? (p-value)		Log-transformed		Means Different? (p-value)		Antilog LSM	Magnitude difference
		Slope	SD		(p-value)	sig.at p < 0.05	Least Squares Means (LSM)	SD	(p-value)	sig. at p < 0.05		
Reference	39	1.3	0.0547	0.727	-	-	0.95	0.0624	-	-	8.93	-
Exposure	38	1.03	0.0632	0.5135	0.212	no	1.04	0.0616	0.001	yes	10.96	23%

Table A2-4: Fish result summary table

Site MMXXXX; Name; Location: Fish Effect Summary							
Trophic Level	Species	Sex	Response	Effect Endpoint	Effect?	Direction	Magnitude
Fish	<i>Catostomus catostomus</i> (Longnose Sucker)	F	Survival	Age	NA		
			Energy Use	Weight-at-age	NA		
				Relative gonad weight	No		
			Energy Storage	Condition	Yes	ref < exp	7% ²
				Relative liver weight	Yes	ref < exp	21% ²
	<i>Catostomus catostomus</i> (Longnose Sucker)	M	Survival	Age	NA		
			Energy Use	Weight-at-age	NA		
				Relative gonad weight	NO		
			Energy Storage	Condition	Yes	ref < exp	6% ²
				Relative liver weight	Yes	ref < exp	23% ²
	<i>Cottus ricei</i> (Spoonhead Sculpin)	F	Survival	Age	Yes	ref < exp	8% ²
			Energy Use	Weight-at-age	Yes	ref < exp	52% ¹
				Relative gonad weight	Yes	ref < exp	57% ²
			Energy Storage	Condition	Yes	ref < exp	31% ¹
				Relative liver weight	Yes	ref < exp	62% ²
	<i>Cottus ricei</i> (Spoonhead Sculpin)	M	Survival	Age	Yes	ref < exp	8% ²
			Energy Use	Weight-at-age	Yes	ref < exp	106% ¹
				Relative gonad weight	Yes	ref < exp	11% ²
			Energy Storage	Condition	Yes	ref < exp	18% ²
				Relative liver weight	Yes	ref < exp	52% ²

¹ ANCOVA is done and the slopes are significantly different. See Appendix 1 for guidance on calculating magnitude of effect.

² Magnitude calculated by comparing the adjusted means between reference and exposed sites (if data were log-transformed; magnitude is calculated on the antilog of the adjusted means). In this case, the slopes are not significantly different and so the adjusted means can be compared directly. (The equation is: [(exposed adjusted mean – reference adjusted mean)/ reference adjusted mean] x 100).

Table A2-5: Descriptive statistics for total benthic invertebrate density (number of invertebrates/m²)**Site MMXXXX; Name; Location****Descriptive Statistics – Benthic Invertebrate Density**

Location	Mean	SD	SE	(n)	Max.	Min.
Reference	4986.85	2011.21	899.44	5	8062.02	2442.73
Near-field*	8062.73	2135.30	954.94	5	10360.31	5535.88
Far-field*	7685.04	3205.63	1433.60	5	11027.65	2717.00

* Near-field = high effluent exposure; Far-field = low effluent exposure

Table A2-6: Analysis of Variance (ANOVA) results for benthic invertebrate total density**Site MMXXXX; Name; Location****ANOVA Results – Benthic Invertebrate Total Density**

Source of Variation	SS	df	MS	F	p-value	Sig. at p < 0.05
Between Groups	2.81E+07	2	1.41E+07	2.236	0.15	NO
Within Groups	7.55E+07	12	6.29E+06			
Total	1.04E+08	14				

Table A2-7: Summary of all benthic invertebrate descriptive statistics

Site MMXXXX; Name; Location							
Effect Endpoint	Location	Mean	SD	SE	(n)	Max.	Min.
Taxa	Ref	19.60	1.52	0.68	5	21	18
	Near-field*	21.20	1.48	0.66	5	23	19
	(NF)						
	Far-field* (FF)	20.00	1.87	0.84	5	23	18
Density	Ref	4986.85	2011.21	899.44	5	8062.02	2442.73
	NF	8062.73	2135.30	954.94	5	10360.31	5535.88
	FF	7685.04	3205.63	1433.60	5	11027.65	2717.00
Simpson's Evenness	Ref	0.77	0.03	0.014	5	0.82	0.75
	NF	0.81	0.03	0.015	5	0.86	0.78
	FF	0.67	0.04	0.017	5	0.71	0.63
Bray-Curtis	Ref	0.24	0.11	0.05	5	0.42	0.14
	NF	0.37	0.10	0.05	5	0.48	0.24
	FF	0.44	0.09	0.04	5	0.55	0.34

* Near-field = high effluent exposure; Far-field = low effluent exposure

Table A2-8: Summary table of all benthic invertebrate results

Site MMXXXX; Name; Location: Benthic Invertebrate Effect Summary					
Trophic Level	Effect Endpoint	Effect?	Direction	Magnitude¹	
				%	SD
Benthos	Density	No			
	Number of taxa	No			
	Simpson's Evenness	Yes	ref > FF	13%	3.33
		Yes	NF > FF	17%	4.67
	Bray-Curtis	Yes	ref < FF	83%	1.82

¹ For a control impact design, magnitude of effect should be reported as the % difference from the reference area [(exposure mean – reference mean)/reference mean] x 100 and standardized for the SD of the reference area (exposure mean - reference mean) / reference SD

Table A2-9: Overall summary of site effects

Site MMXXXX; Name; Location: Site Effect Summary								
Trophic Level	Species	Sex	Response	Effect Endpoint	Effect?	Direction	Magnitude ³	
							%	SD
Fish	<i>Catostomus catostomus</i> (Longnose Sucker)	F	Survival	Age	NA			
			Energy use	Weight-at-age	NA			
				Relative gonad weight	No			
			Energy storage	Condition	Yes	ref < exp	7% ²	
				Relative liver weight	Yes	ref < exp	21% ²	
	<i>Catostomus catostomus</i> (Longnose Sucker)	M	Survival	Age	NA			
			Energy use	Weight-at-age	NA			
				Relative gonad weight	No			
			Energy storage	Condition	Yes	ref < exp	6% ²	
				Relative liver weight	Yes	ref < exp	23% ²	
	<i>Cottus ricei</i> (Spoonhead Sculpin)	F	Survival	Age	Yes	ref < exp	8% ²	
			Energy use	Weight-at-age	Yes	ref < exp	52% ¹	
				Relative gonad weight	Yes	ref < exp	57% ²	
			Energy storage	Condition	Yes	ref < exp	31% ¹	
				Relative liver weight	Yes	ref < exp	62% ²	

Site MMXXXX; Name; Location: Site Effect Summary								
Trophic Level	Species	Sex	Response	Effect Endpoint	Effect?	Direction	Magnitude ³	
							%	SD
	<i>Cottus ricei</i> (Spoonhead Sculpin)	M	Survival	Age	Yes	ref < exp	8% ²	
			Energy use	Weight-at-age	Yes	ref < exp	106% ¹	
				Relative gonad weight	Yes	ref < exp	11% ²	
			Energy storage	Condition	Yes	ref < exp	18% ²	
				Relative liver weight	Yes	ref < exp	52% ²	
Benthos				Density	No			
				Number of taxa	No			
				Simpson's Evenness	Yes	ref > FF	13%	3.33
					Yes	NF > FF	17%	4.67
				Bray-Curtis	Yes	ref < FF	83%	1.82

¹ ANCOVA is done and the slopes are significantly different. See Appendix 1 for guidance on calculating magnitude of effect.

² Magnitude calculated by comparing the adjusted means between reference and exposed sites (if data were log-transformed; magnitude is calculated on the antilog of the adjusted means). In this case, the slopes are not significantly different and so the adjusted means can be compared directly. (The equation is: [(exposed adjusted mean – reference adjusted mean)/ reference adjusted mean] x 100).

³ For benthic invertebrate community surveys following a control impact designs, magnitude of effect should be reported as the % difference from the reference area [(exposure mean – reference mean)/reference mean] x 100 and standardized for the SD of the reference area (exposure mean - reference mean) / reference SD

Appendix 3: Case study – ANCOVA and Power Analysis for Fish Survey

A case example is provided to demonstrate the application of some of the methods recommended above. The data were collected during a previous adult fish survey at a Canadian pulp mill. In this particular example, the mill was a bleached-kraft operation and discharged effluent into a lake receiving environment. The reference area was an adjacent bay of the lake exhibiting similar natural habitat characteristics as the near-field (high effluent exposure) area, and did not receive any allochthonous discharges. The sentinel fish species selected for the survey was White Sucker (*Catostomus commersoni*). The sample sizes approximated those recommended for the fish survey, with the exception of males at the near-field area:

	Near-field (high effluent exposure) area	Reference area
Males	12	22
Females	26	24

The data set included information on length, weight, liver weight, sex, gonad weight, and age of male and female adult sucker. Fecundity estimates were not available.

After the initial step of ensuring the data set was free of transcription errors, the mean and standard deviation of each variable were calculated per sex and sampling area (Table A3-1). Mathematical procedures were conducted separately for males and females. Normal probability plots were generated for each variable (per sex and area) to identify extreme outliers and to assess normality of the data. Examination of these plots did not indicate obvious extreme outliers with the exception of one male from the near-field area. Residual plots from the Analysis of Variance (ANOVA) / ANCOVA models can also be used to inspect the data.

Table A3-1: Mean, standard deviation (SD) and sample size (n) of measurements recorded on White Sucker (*Catostomus commersoni*) during the example survey

Sex	Area	Fork Length (cm)	Body Weight (g)	Gonad Weight (g)	Liver Weight (g)	Age (yr)
Female	Reference	44.4±2.4 (24)	1135.1±150.4 (24)	46.0±7.5 (24)	15.8±2.1 (24)	10.8±3.4 (24)
	Near-field*	41.5±3.0 (26)	1081.7±228.4 (26)	36.9±15.9 (26)	17.8±7.5 (16)	11.1±3.0 (26)
Male	Reference	41.4±1.8 (22)	975.9±90.1 (22)	66.4±11.1 (21)	10.8±1.6 (22)	10.5±3.0 (22)
	Near-field*	39.3±1.8 (12)	950.9±166.6 (11)	48.0±15.5 (12)	18.4±8.4 (12)	12.3±3.0 (12)

*Near-field = High effluent exposure

As previously outlined, most of the results were derived using ANCOVA. For the purposes of illustration, a detailed description is provided for the parameter size (length)-at-age for female White Sucker (one of the supporting endpoints for the fish survey). For these calculations, both length and age were $\log_{10}$ -transformed.

The first step is to conduct the preliminary test of equality of slopes. The model statement for this analysis of size-at-age is:

$$\log(\text{length}) = \text{constant} + \text{area} + \log(\text{age}) + \text{area} * \log(\text{age}),$$

where the interaction term $\text{area} * \log(\text{age})$ represents the test for equality of slopes of the area regression lines, and $\log(\text{age})$ is the covariate. From the ANCOVA table, it is evident that the interaction term, $\text{area} * \log(\text{age})$, is not significant ($P=0.376$) (Table A3-2a). This tells us that the slopes of the regression lines for each area can be treated as being approximately parallel. It also tells us that the interaction term can be dropped from the model, and we can proceed to the ANCOVA model:

$$\log(\text{length}) = \text{constant} + \text{area} + \log(\text{age}),$$

where *area* represents the test for differences in adjusted means. The mean square error (mean square error) from the resulting ANCOVA table will provide the estimate of variability (mean square error=0.00033) for length-at-age (Table A3-2b). While conducting the above analyses, the residuals from the preliminary and ANCOVA model can be saved for the purpose of assessing the assumptions of normality and homogeneity of variance.

Table A3-2: Size-at-age (log(length) vs log(age)) for female White Sucker using ANCOVA. The analysis includes: a) a preliminary test of the equality of slopes, and (b) the ANCOVA model table (test of adjusted means)

(a) Preliminary Analysis of Equality of Slopes

Source	Sum-of-Squares (SS)	Degrees of Freedom (df)	Mean-Square (MS)	F-Ratio	p-Value
Area	0.00086	1	0.00086	2.62639	0.11193
Log(age)	0.02126	1	0.02126	64.58474	< 0.0001
Area*Log(age)	0.00026	1	0.00026	0.79780	0.37640
Error	0.01514	46	0.00033		

(b) ANCOVA Model

Source	SS	df	MS	F-Ratio	p-value
Area	0.01240	1	0.01240	37.57576	< 0.0001
Log(age)	0.02167	1	0.02167	65.66667	< 0.0001
Error	0.01541	47	0.00033		

Calculation of Sample Size

To calculate sample size, the Z-value power equation described earlier can be used. As a reminder, the equation is:

$$n = 2 (Z_{\alpha} + Z_{\beta})^2 (SD/CES)^2 + 0.25Z_{\alpha}^2$$

The square root of the mean square error from the ANCOVA model substitutes for the SD in the power equation. The critical effect size (CES) refers to the effect or difference in the parameter one wishes to detect. For the purpose of this example and the remaining parameters of the case study, samples sizes were calculated for a CES of 5, 10, 20, 50 and 100% (i.e., differences between areas).

Many of the parameters calculated for the fish survey, are typically log-normally distributed and require log transformations. To calculate sample sizes, SD and CES should be expressed in logarithms. It should be noted, however, not to add 1 to values before logging for the purposes of the fish environmental effects monitoring (EEM) survey because it has undesirable effects on the calculated variances when changing measurement units. A difference in logarithms is equivalent to multiplying or dividing by some factor. For example, if the difference in log length between two areas is 0.301, then the fish from one area is twice the length ($\text{antilog } 0.301 = 2$) as fish from the other area. In the following table, CES has been expressed in logarithms with the corresponding antilog; these values of CES correspond roughly with those used for untransformed data:

Critical effect size (logarithm)	0.0212	0.0414	0.0792	0.176	0.301
Critical effect size (antilog)	1.05	1.10	1.20	1.50	2.00
% increase*	5	10	20	50	100
% decrease†	5	9	17	33	50

Therefore for length-at-age ($\log_{10}$ data):

- $SD = (\text{mean square error})^{0.5} = (0.00033)^{0.5} = 0.01817$
- $Z_{\alpha}(2)$ (2-tailed test) = 1.96
- Z_{β} (1-tailed test) = 1.282
- CES = 5% (see above table)

$$n = 2 (1.96 + 1.28)^2 (0.01817 / 0.0212)^2 + 0.25(1.96)^2$$

$$n = 16.4 \text{ (or, rounding up, } n=17\text{)}$$

Similarly for the remaining effect sizes, the estimated sample sizes (i.e., number of fish to be sampled per area) would be:

CES	5%	10%	20%	50%
<i>n</i>	17	5	3	2

The estimate of variability and sample size calculations for gonad weight, liver weight and condition for female and male sucker were calculated in the same fashion as described for length-at-age (Table A3-3). In all but one case, the slopes of the reference/near-field

* In exposure area vs reference

† In exposure area vs reference

regression lines were equal and the mean square errors from the ANCOVA model were used as the estimate of variability. For male White Sucker, the slopes of the regressions of log(weight) on log(length) (i.e., condition) were not equal among areas ($P=0.0068$). To investigate whether the one possible outlier (male, near-field) influenced the ANCOVA, it was rerun without this data point. In this case, the regressions were homogeneous between areas. This was partially a consequence of the low sample size (i.e., an increased influence of an outlier on the regression) and should be noted when reporting the data.

For mean age, the mean square error from the one-way ANOVA was used to estimate the variability (Table A3-3).

The final results of the sample size calculations (Table A3-3) indicate that the maximum numbers of fish needed to be collected from each area were approximately 703 males and 738 females to detect a 5% difference between areas (CES), 185 males and 194 females to detect a 10% difference, 52 males and 54 females to detect a 20% difference, and 12 males and 12 females to detect a 50% difference. Among all the parameters, mean age was the most variable and required the highest sample size to detect differences.

Table A3-3: Numbers of fish needed to detect significant differences in fish endpoints among areas using the model mean square error as the estimate of variability. Sample sizes were calculated for a range of CESs with power=0.90 and $\alpha=0.05$. All data were $\log_{10}$ -transformed – example survey

Endpoint	Sex	Model	Log	Estimated Sample Size (number of fish/area)			
				Mean Square Error	CES= 5%	CES= 10%	CES= 20% CES= 50%
Length-at-age	Male	ANCOVA	0.00014	8	3	2	2
	Female	ANCOVA	0.00033	17	5	3	2
Weight-at-age	Male	ANCOVA	0.00211	100	27	9	3
	Female	ANCOVA	0.00295	139	38	11	3
Condition	Male	ANCOVA ¹	N/A	-	-	-	-
	Female	ANCOVA	0.00100	48	14	5	2
Liver Weight	Male	ANCOVA	0.00994	466	123	35	8
	Female	ANCOVA	0.00626	294	78	22	6
Gonad Weight	Male	ANCOVA	0.00881	413	110	31	7
	Female	ANCOVA	0.01013	475	126	35	8
Mean Age	Male	ANOVA	0.01499	703	185	52	12
	Female	ANOVA	0.01574	738	194	54	12

¹ Preliminary analysis (test of slopes) conducted as first step to ANCOVA was significant (i.e., slopes not parallel).

Table of Contents

9. Alternative Monitoring Methods	9-1
9.1 Overview	9-1
9.2 Use of Mesocosms as an Alternative Monitoring Method	9-2
9.2.1 Background Information on Artificial Stream Development and Application	9-2
9.2.2 Applicability within the EEM Programs	9-27
9.2.3 Mesocosm Technology	9-27
9.2.3.1 Suitability of Mesocosm Design	9-27
9.2.3.2 Description and General Operation of Large Trailer Mesocosm System	9-28
9.2.3.3 Description and General Operation of Modular Mesocosm System	9-29
9.2.4 Considerations for Site Selection	9-30
9.2.5 Biological Monitoring Study Designs	9-31
9.2.5.1 Overview	9-31
9.2.5.2 Study Treatments	9-32
9.2.5.3 Replication	9-33
9.2.5.4 Sample Sizes and the Role of Effect Size	9-33
9.2.5.4.1 No Available Pre-existing (Historical) Data	9-33
9.2.5.4.2 Available Pre-existing (Historical) Data	9-34
9.2.6 Fish Monitoring – Effects Assessment	9-35
9.2.6.1 Study Design	9-35
9.2.6.1.1 Species Selection	9-35
9.2.6.1.2 Effect Indicators	9-37
9.2.6.1.3 Timing and Duration of Effluent Exposure	9-42
9.2.6.2 Fish Methods	9-43
9.2.6.2.1 General Set-up – Trailer Mesocosm System	9-43
9.2.6.2.2 General Set-up – Modular Mesocosm System	9-44
9.2.6.2.3 Maintenance and Monitoring	9-46
9.2.6.2.4 Sampling and Analysis of Fish	9-46
9.2.6.3 Data Assessment and Interpretation	9-46
9.2.6.3.1 General Requirements	9-46
9.2.6.3.2 Statistical Analysis of Parameters	9-47
9.2.6.3.3 Nested ANOVA Analyses	9-47
9.2.6.3.4 ANCOVA Analyses	9-48
9.2.6.3.5 Data QA/QC and Analysis	9-48
9.2.7 Benthic Invertebrate Community Monitoring - Effects Assessment	9-48
9.2.7.1 Study Design	9-48
9.2.7.1.1 Species Selection	9-48
9.2.7.1.2 Effect Endpoints	9-48
9.2.7.1.3 Timing and Duration of Effluent Exposure	9-48
9.2.7.2 Benthic Invertebrate Methods	9-49
9.2.7.2.1 Mesocosm Trailer – General Set-up	9-49
9.2.7.2.2 Modular Mesocosm – General Set-up	9-50
9.2.7.2.3 Maintenance and Monitoring	9-51

9.2.7.3	Sampling and Analysis of Benthic Invertebrates.....	9-51
9.2.7.4	Data Assessment and Interpretation.....	9-52
9.2.8	Supporting Measurements	9-53
9.2.8.1	Water Chemistry Parameters	9-53
9.2.8.2	Supporting Water Quality Parameters	9-53
9.3	Use of Caged Bivalves as an Alternative Monitoring Method	9-54
9.3.1	Introduction.....	9-54
9.3.2	Background.....	9-54
9.3.3	Approach.....	9-55
9.3.4	Species Selection	9-58
9.3.4.1	Most Commonly Used Taxa	9-58
9.3.4.2	Marine and Estuarine Bivalves	9-60
9.3.4.3	Freshwater Bivalves.....	9-61
9.3.4.4	Source	9-62
9.3.4.5	Species Identification.....	9-63
9.3.4.6	Size and Age	9-64
9.3.4.7	Number of Organisms.....	9-65
9.3.4.8	Handling and Holding Conditions	9-65
9.3.5	Study Design.....	9-65
9.3.5.1	Sampling Design.....	9-66
9.3.5.1.1	Area and Station Selection.....	9-67
9.3.5.2	Replication.....	9-67
9.3.5.2.1	Number of Cage Stations.....	9-67
9.3.5.2.2	Applying Replication to Study Design	9-68
9.3.5.3	Timing and Duration.....	9-68
9.3.5.4	Modifying or Confounding Factors and Supplementary Measures.....	9-69
9.3.6	Cage Designs	9-69
9.3.6.1	Mooring Systems	9-69
9.3.7	Methods for Test Initiation, Cage Deployment and Retrieval and Test Termination.....	9-72
9.3.7.1	Test Initiation.....	9-72
9.3.7.2	Cage Deployment.....	9-73
9.3.7.3	Cage Retrieval and Test Termination	9-73
9.3.8	Effect Indicators.....	9-74
9.3.8.1	Survival.....	9-75
9.3.8.2	Growth	9-75
9.3.8.3	Condition.....	9-76
9.3.8.4	Energy Measure	9-77
9.3.8.5	Reproductive Effort	9-78
9.3.9	Quality Assurance and Quality Control.....	9-81
9.3.10	Data Analyses	9-81
9.3.11	Mercury in Fish Tissue	9-82
9.3.12	Reporting.....	9-83
9.4	References.....	9-84

List of Tables

Table 9-1: Summary of artificial stream applications for assessing the effects of pulp and paper and mining effluents on aquatic ecosystems as required under Canadian environmental effects monitoring	9-4
Table 9-2: Fish mesocosm study effect indicators and endpoints and related statistical procedures	9-38
Table 9-3: Recommended response variables and suitable additional supporting information, and suggested statistical analysis for Fathead Minnow application in modular mesocosm systems	9-39
Table 9-4: Suggested taxa for use in caged bivalve studies for EEMs	9-59
Table 9-5: Differences noted between two species of mussels over 5-year study in the Burrard Inlet, Vancouver, British Columbia	9-61
Table 9-6: Differences noted between Unionoidea and Sphaeriidae	9-62
Table 9-7: Caged bivalve study effect indicators and endpoints and related statistical procedures	9-75
Table 9-8: An example of a field data sheet for recording survival and growth raw data	9-80

List of Figures

Figure 9-1: A) Large mesocosm system with streams situated on tables (Model I) used in the Athabasca River, Alberta. B) Large mobile mesocosm system with streams on 2 trailers (Model II) used in the Fraser River, British Columbia; the Saint John River, New Brunswick; and in Saint John Harbour, New Brunswick. C) Large mobile mesocosm system with streams on a single trailer (Model III) used in the Miramichi and Little rivers, New Brunswick; the Wapiti River, Alberta; and Junction Creek, Ontario.	9-18
Figure 9-2: A) Small microcosm system with streams situated on tables over mixing reservoirs used in the Thompson River, British Columbia. B) Modular mesocosm system with streams situated on tables over mixing reservoirs used in the Little River, New Brunswick; Junction Creek, Ontario; the Wabigoon River, Ontario; and Key Lake, Saskatchewan.	9-19
Figure 9-3: Schematic of large mesocosm trailer system (not to scale)	9-20
Figure 9-4: Photograph of a modular mesocosm parts diagram.	9-21
Figure 9-5: Modular mesocosm flow schematic	9-22
Figure 9-6: Multitrophic Fathead Minnow reproductive bioassay and feeding barrier	9-23
Figure 9-7: Site set-up for modular mesocosms	9-24
Figure 9-8: A) Factorial experimental design to investigate the importance of water vs. diet in responses of Fathead Minnow to metal mine effluent in modular mesocosms; B) Experimental design to investigate the influence of pH and natural organic matter (NOM) on Fathead Minnow responses after exposure to an MME mixture and a single metal in multitrophic modular mesocosms.	9-25
Figure 9-9: Factorial experimental design to investigate the effects of MME and historical sediment contamination in isolation and in combination on Fathead Minnow in modular mesocosms.	9-26
Figure 9-10: Mussel showing ripe mantle lobe	9-56
Figure 9-11: Reproductive cycle of Blue Mussels from British Columbia: A) Mantle energy stored in fall; B) Mantle reproductive content in spring	9-57

Figure 9-12: Mytilus spp. shell scars markings	9-64
Figure 9-13: Duplicate frame from a caged mussels exposure experiment.....	9-70
Figure 9-14: Modular mesocosm parts diagram	9-71
Figure 9-15: Modular mesocosm flow schematic.....	9-72
Figure 9-16: A) Mantle plug and tools necessary for its removal; B) Mantle plug after homogenization and ready for assessment.....	9-78

9. Alternative Monitoring Methods

9.1 Overview

At some mines, standard fish and benthic invertebrate community monitoring studies may not be appropriate. The reasons for this are site-specific, but the most common reasons are the presence of hazardous conditions (e.g., high flow velocity); unsuitable habitat for sampling; or the presence of confounding factors, such as other effluent discharges in the exposure area, that make it impossible to isolate any effects attributable to the effluent being monitored.

Where mines cannot design the fish or benthic invertebrate community surveys to resolve difficulties associated with confounding influences, they will provide a scientific rationale and justification, and propose cost-effective and technically feasible alternative monitoring methods within the study design. A number of alternative monitoring methods are recommended in this chapter.

Mines may choose other scientifically defensible methods, provided that the results can determine if the effluent is having effects on the fish population (growth, reproduction, condition and survival), fish tissue (mercury), or the benthic invertebrate community (benthic invertebrate density, taxa richness, the Simpson's Evenness Index and the Bray-Curtis Index). Currently, recommended alternatives to the fish survey monitoring method are mesocosm (artificial stream) and caged bivalve studies. For benthic invertebrate community surveys, the recommended alternative monitoring method is the mesocosm study.

Other alternatives to the fish and benthos field surveys may also exist. Mines can suggest other alternative methods in their study design. New alternative methods will be evaluated by the Authorization Officer, with the support of the Technical Advisory Committee and the Environmental Effects Monitoring (EEM) Science Committee. In reviewing the suggested alternative, some specific design elements will be considered as essential to meeting the objective of the program, including environmental relevance and interpretable results that are scientifically defensible and manageable.

The objective of section 9.2 is to provide guidance on the study design and implementation of mesocosm studies as an alternative EEM method for assessing the effects of metal mine effluent on benthic invertebrates and fish. This guidance document is intended to provide information on recommended standards of good scientific practice available to meet the outlined EEM requirements. In 2002, the first guidance document for the use of artificial stream systems (mesocosms) was released. Guidance was updated in 2011 to reflect ongoing research and development for improvement of this alternative EEM method.

The objective of section 9.3 is to provide technical guidance for conducting controlled experiments using caged bivalves suspended in the water column to test for effects associated with industrial discharges and to compare measurements between exposure and reference areas. Caged bivalves are also an alternative to the fish survey and may be considered for mines where the fish survey has been unsuccessful or impractical in past phases of EEM, or where there may be study design issues, such as confounding influences, or safety concerns.

9.2 Use of Mesocosms as an Alternative Monitoring Method

9.2.1 Background Information on Artificial Stream Development and Application

Artificial streams are recommended as monitoring alternatives because years of research and development have demonstrated that, with respect to effluent effects, they can produce good-quality data that fit within the required regulatory context (Table 9-1). Since 1991, field-based artificial stream system studies have been conducted for assessing the effects of point-source effluents on aquatic ecosystems. Field-based artificial stream studies relevant to EEM applications were conducted in Canada 14 times in 8 years between 1993 and 2008 (Table 9-1). All of these studies, and development of the alternative method, were conducted as collaborative partnerships between industry, government, academia and consultants. All funding for the research was acquired through mechanisms independent of the EEM Program. Applications of this method are presented in detail below to provide a thorough understanding of the work that has been done to date. A summarized version with references can be found in Table 9-1. These references should be consulted if similar types of studies and experimental designs are being considered.

1991-1996, Northern River Basins Study (NRBS), Alberta

One of the first mesocosm applications assessed the effects of pulp mill effluent (PME) on benthic invertebrate and periphytic algal communities in the Athabasca River (Table 9-1) (Culp and Podemski 1996; Culp et al. 1996; Podemski and Culp 1996; Podemski 1999; Culp et al. 2001). Artificial streams were used to distinguish the effects of nutrients in whole mill effluent from contaminants, on the basis of directional differences in biological response. Specifically, moderate nutrient enrichment would increase primary and secondary productivity, whereas contaminant effects would reduce growth and reproduction and eventually result in mortality (Culp and Podemski 1996; Podemski and Culp 1996; Culp and Lowell 1998; Culp et al. 2001). To achieve this objective, 3 treatments were tested in the spring of 1993: control Athabasca River water, 1% (volume/volume [v/v]) treated PME, and 1% (v/v) nutrients (nitrogen + phosphorus) at levels measured in the PME. The hypothesis was that exposure to both the PME and nutrient treatments would result in nutrient enhancement effects on the benthic food web and that the PME and nutrient treatments would not differ. This would suggest that the effects of PME at levels found in the Athabasca River were due to nutrient enrichment rather than contaminant toxicity.

A large non-mobile artificial stream system was used near the pulp mill at Hinton, Alberta. The system consisted of 16 circular tanks or streams (0.9 m² each) placed on tables (Model I, Figure 9-1A). River water was pumped into each stream at a controlled rate, and effluents and nutrients were added to the treatment streams as previously published by Culp and Podemski (1996) and Podemski (1999). A standardized benthic community, endemic to the Athabasca River, was created in each stream and exposed to PME for 28 days. At the end of the exposure period, algal biomass, growth of mayfly (Ephemeroptera: Siphonuridae, Baetidae) and stonefly (Plecoptera: Capniidae) nymphs, and insect abundance, increased in the treatment streams relative

to the reference treatment (Culp and Podemski 1996; Podemski and Culp 1996; Culp et al. 1996). In addition, these response variables did not differ between the 1% PME and the 1% nutrient treatments, supporting the hypothesis that the effects of PME on the benthic food web were attributable to nutrient enrichment.

1991-1998, Fraser River Action Plan (FRAP), British Columbia

The FRAP was conducted from 1991 to 1997 to determine the current state of health of the Fraser River Basin ecosystem, including assessment of 8 pulp and paper mill effluents (Gray and Tuominen 1998; McGreer and Belzer 1998).

Thompson River, Kamloops, British Columbia (1993-1994)

The Thompson River showed signs of nutrient enrichment due to the discharge of PME at the City of Kamloops. This problem has been investigated since the early 1970s, when excessive accumulations of periphytic algae occurred in the river downstream of the pulp mill (Federal-Provincial Thompson River Task Force 1976). Bothwell and Daley (1981), Bothwell (1985), Bothwell et al. (1992) and Bothwell and Culp (1993) illustrated how periphytic algae growth was enriched by bioavailable phosphorus discharged in PME.

Artificial streams were used to separate the interacting effects of nutrients and contaminants in PME on algae and benthic invertebrates (Table 9-1). The approach differed from the NRBS studies in that a dose-response design was employed with the expectation of observing nutrient effects at low effluent concentrations and contaminant effects at higher concentrations. In 1993 and 1994, periphytic algae and chironomids were exposed to a dilution series of PME (0.25-10% [v/v]) (Dubé and Culp 1996; Culp and Lowell 1998). Smaller artificial streams were used for testing the effects of the PME on single insect species (Lowell et al. 1995, 1996) and simplified benthic food webs (Dubé and Culp 1996). The single-species approach focused the assessment of effects on key sentinel taxa, to improve our understanding of species-specific responses (Culp et al. 2000b).

The artificial stream system was set up on the banks of the Thompson River at Kamloops just upstream of the effluent outfall. The system included a water distribution system, treatment reservoirs for mixing the respective effluent dilutions with a continuous supply of river water, and small circular 0.33-L streams (45 cm² planar area) (Dubé 1995; Lowell et al. 1995) (Figure 9-2A). Algae and chironomid larvae (Diptera: Orthocladinae) from a reference area were placed into the streams, and changes in algae and chironomid biomass were measured after 2-3 weeks of effluent exposure (Dubé and Culp 1996). Dubé and Culp (1996) reported that algal biomass (chlorophyll *a*) increased in all effluent concentrations due to nutrient enrichment. Total chironomid biomass and individual weight were also enriched at low effluent concentrations (< 5%). At higher concentrations (5% and 10%) chironomid biomass decreased, possibly due to contaminant effects.

Table 9-1: Summary of artificial stream applications for assessing the effects of pulp and paper and mining effluents on aquatic ecosystems as required under Canadian environmental effects monitoring

Year	Program ¹	Effluent Type ²	Research Objective	Location	Artificial Stream System	References
1993-1994	NRBS	PME	To determine the effect of nutrients and contaminants in PME on periphytic algae and benthic invertebrate communities	Athabasca River, Hinton, AB	Model I: Large non-mobile mesocosm. Benthos. Field study.	Podemski (1999) Culp and Podemski (1996) Podemski and Culp (1996) Culp et al. (1996) Culp et al. (2001)
1993-1994	FRAP	PME	To assess the effects of increasing concentrations of PME (0.25–10%) on periphyton and chironomids (Diptera: Orthocladiinae)	Thompson River, Kamloops, BC	Small microcosm system. Benthos. Field study.	Dubé (1995) Dubé and Culp (1996) Dubé et al. (1997) Lowell et al. (2000)
1993	FRAP	PME	To determine the food-dependent effects of PME on the mayfly species Blue-winged Olive (Ephemeroptera: <i>Baetis tricaudatus</i>)	Thompson River, Kamloops, BC	Small microcosm system. Benthos. Field study.	Lowell et al. (1995) Lowell et al. (1996) Culp et al. (1996) Lowell et al. (2000)
1994	FRAP	PME	To determine the effects of PME (1% and 3%) on periphyton and benthic invertebrate communities	Fraser River, Prince George, BC	Model II: Large mobile mesocosm. Benthos. Field study.	Culp and Cash (1995) Culp et al. (1996) Culp et al. (2000a)
1997	Pulp and Paper EEM	PME	To assess the effects of PME (3%) on a small-bodied fish, Mummichog (<i>Fundulus heteroclitus</i>), in a marine environment	Saint John Harbour, Saint John, NB	Model II: Large mobile mesocosm. Fish. Field study.	Cash et al. (2003)
1997-1998	Pulp and Paper EEM	PME	To determine the effects of final PME, in-mill process streams, and a mill process change on a small-bodied fish, Mummichog, in an estuarine environment	Saint John River, Saint John, NB	Model II: Large mobile mesocosm. Fish. Field study.	Dubé (2000) Dubé and MacLachy (2000a) Cash et al. (2003)

Year	Program ¹	Effluent Type ²	Research Objective	Location	Artificial Stream System	References
1999	TSRI	PME	To determine the effects of primary- and secondary-treated PME on a small-bodied fish, Mummichog, in an estuarine environment	Miramichi River, Miramichi, NB	Model III: Large mobile mesocosm. Fish. Field study.	Dubé et al. (2002)
2000	TSRI	MME	To determine the effects of MME (20%, 80%) on juvenile Atlantic Salmon (<i>Salmo salar</i>)	Little River, Brunswick Mines, Miramichi, NB	Model III: Large mobile mesocosm. Fish. Field study.	Dubé et al. (2005)
2001	TSRI	PME MSE	To evaluate the individual and combined impacts of MSE and PME on Longnose Dace (<i>Rhinichthys cataractae</i>)	Wapiti River, AB	Model III: Large mobile mesocosm. Fish. Field study.	Dubé et al. (2004)
2001	TSRI	PME MSE	To evaluate the individual and combined impacts of MSE and PME on the benthic food web	Wapiti River, Grande Prairie, AB	Model III: Large mobile mesocosm. Benthos. Field study.	Culp et al. (2004)
2001-2002	Industry	MME	To assess effects of treated MMEs from three mines discharging to Junction Creek, Sudbury, on Creek Chub (<i>Semotilus atromaculatus</i>) and Pearl Dace (<i>Semotilus margarita</i>)	Junction Creek, Sudbury, ON	Model III: Large mobile mesocosm. Fish. Field study.	Dubé et al. (2006)
2002	Industry	MME	To evaluate the effects of MME (45%) on the partial life cycle of the chironomid <i>Chironomus tentans</i>	Junction Creek, Sudbury, ON	Modular mesocosm system. Benthos. Field study.	Hruska and Dubé (2004)
2003	NSERC	MME	Comparison of a partial–life-cycle bioassay in artificial streams to a standard beaker bioassay, to assess effects of	Junction Creek, Sudbury, ON	Modular mesocosm system. Benthos. Lab study.	Hruska and Dubé (2005)

Year	Program ¹	Effluent Type ²	Research Objective	Location	Artificial Stream System	References
			MME (45%) on the chironomid <i>C. tentans</i>			
2003	NSERC/ Industry	PME	To determine effects of final PME (1%, 100%) and various process streams on the partial life cycle of Fathead Minnow (<i>Pimephales promelas</i>) under environmentally realistic conditions (i.e., ambient water and effluent quality)	Terrace Bay, ON	Bioassay trailer. Fish. Field study.	Rickwood et al. (2006a, 2006b)
2004	NSERC/ Industry	MME	To develop a self-sustaining multitrophic bioassay, using <i>C. tentans</i> and Fathead Minnow (partial life cycle) to comparatively assess effects of water-borne vs. food- and water-borne exposure to MME (45%) on Fathead Minnow reproduction	Junction Creek, Sudbury, ON	Modular mesocosm system. Multitrophic. Lab study.	Rickwood et al. (2006c)
2005	NSERC/ Industry	MME	To develop a self-sustaining multitrophic bioassay, using <i>C. tentans</i> and Fathead Minnow (partial life cycle) to comparatively assess effects of water-borne vs. food- and water-borne exposure to MME (45%) on Fathead Minnow reproduction	Junction Creek, Sudbury, ON	Modular mesocosm system. Multitrophic. Field study.	Rickwood et al. (2008)
2006	NSERC/ Industry	PME	To assess effects of PME (20%, 40%, 60%) on Fathead Minnow (partial life cycle)	Wabigoon River, Dryden, ON	Modular mesocosm system. Fish. Field study.	Pollock et al. (2009)

Year	Program ¹	Effluent Type ²	Research Objective	Location	Artificial Stream System	References
2007	NSERC/ Industry	MME and potential causative metal	To comparatively evaluate response patterns of Fathead Minnow (partial life cycle) to an MME mixture (100%, 25%, 5%) vs. selenium as selenate, using a multitrophic mesocosm bioassay	Unknown Lake, Key Lake, SK	Modular mesocosm system. Fish. Lab study	Pollock et al. (unpublished)
2008	NSERC/ Industry	MME	To comparatively evaluate effluent (current discharge: 25%) vs. sediment (historical contamination) exposure pathways on Fathead Minnow (partial life cycle)	Unknown Lake, Key Lake, SK	Modular mesocosm system. Fish. Field study.	Driessnack et al. (unpublished)
2008	NSERC/ Industry	MME	To assess the effects of three different MME discharges on Fathead Minnow (partial life cycle).	Junction Creek, Sudbury, ON	Modular mesocosm system. Multitrophic. Field study.	Ramilo et al. (unpublished)

¹NRBS: Northern River Basins Study; FRAP: Fraser River Action Plan; EEM: Environmental Effects Monitoring; TSRI: Toxic Substances Research Initiative; NSERC: Natural Sciences and Engineering Research Council of Canada.

²PME: pulp mill effluent; MSE: municipal sewage effluent; MME: metal mine effluent.

In 1993, Lowell et al. (1995, 1996) conducted small-scale artificial stream experiments on the Thompson River in concert with those of Dubé and Culp (1996). Using the mayfly species Blue-winged Olive, the effects of PME (1% and 10% v/v) on survival, growth, moulting and morphological development were investigated under 2 feeding regimes (low and high). Effluent exposure significantly stimulated growth and development, with 20–50% increases in dry body weight relative to controls. Although moulting frequency increased with moderate effluent exposure (1%), higher exposure (10%) reduced moulting frequency, suggesting a contaminant-mediated mechanism (Lowell et al. 1996). These artificial stream results using mayflies as the sentinel species were consistent with the chironomid exposure experiments conducted by Dubé and Culp (1996), which showed an enrichment response at low PME concentrations and the appearance of inhibitory effects at higher concentrations.

In addition to consistency among artificial stream experiments, these results were consistent with field survey results (Culp and Lowell 1998). Long-term trend analysis showed that several families of stoneflies (Plecoptera), caddisflies (Trichoptera), and mayflies (Ephemeroptera) were more abundant in the years when the mill output of suspended solids and phosphorus was higher (Lowell et al. 1996, 2000). Field monitoring by Dubé et al. (1997) also showed that temporal and spatial patterns in water-column phosphorus, periphyton biomass and chironomid biomass (Diptera: Orthocladiinae) were consistent under normal mill operating conditions. The effects of the mill on the Thompson River benthic food web were restricted to nutrient enrichment. However, Dubé (1995) also observed that toxic effects of mill-related contaminants decreased chironomid densities in the Thompson River at low effluent exposure (far-field) sites in 1992 when the mill's secondary effluent treatment system shut down.

Fraser River, Prince George, British Columbia (1994)

The effects of PME on benthic food webs using artificial streams were also examined in the Fraser River at Prince George, British Columbia, which received effluent from 4 pulp mills located within a 100-km stretch of the river (Culp and Lowell 1998). In 1994, benthic communities were exposed to 1% and 3% concentrations (v/v) of PME for 35 days to determine if nutrient enrichment effects occurred at low PME concentrations and toxic effects manifested at higher concentrations (Table 9-1) (Culp and Cash 1995; Culp et al. 2000a). The number of measured response variables increased in this study, and included bacterial number, periphyton biomass, composition, accumulation of target PME contaminants, and benthic invertebrate community structure. Community-level responses, in addition to species-specific responses, were measured to increase the ecological relevance of the study (Culp et al. 2000b).

The design of the large artificial stream system was modified to improve its flexibility, transportability and cost-effectiveness. The streams and tables were secured onto 2 mobile flatbed trailers (Culp et al. 1996) (Model II, Figure 9-1B). In addition, each trailer was constructed with enclosed laboratory space for effluent and water header tanks, pump storage, and space for sample processing. Design and operation of the streams, including benthic community inoculation, flow rates and sampling protocols, were as previously described by Culp and Cash (1995), Culp and Lowell (1998), and Culp et al. (2000a).

Results from the Fraser River studies supported those from both the Thompson River and NRBS studies, illustrating that the effects of PME on the benthic food web were caused by nutrient enrichment. Culp et al. (2000a) reported that bacterial numbers, periphyton biomass and the biomass of dominant insect taxa (i.e., chironomids and stoneflies) increased with effluent exposure. Interestingly, although a dose-response relationship was observed for pulp mill contaminants (i.e., resin acids and chlorinated phenolics) measured in periphyton, these tissue burdens did not translate into a decrease in algal growth or a change in species richness. Results were also consistent with laboratory and field studies building a weight of evidence on the effects of PME on riverine benthos response patterns in the Fraser River (Culp et al. 2000a).

1997-1998, Industrial EEM Pilot Studies, New Brunswick

Studies conducted during the NRBS and FRAP illustrated the utility of employing artificial streams for assessing the effects of PME on bacteria, periphyton and benthic invertebrate communities. The systems provided a mechanism to measure responses of endemic biota to controlled effluent concentrations under ambient, environmentally relevant conditions of light, temperature and water quality (Culp et al. 1996). These demonstrated qualities also made artificial stream application an attractive alternative for assessing the effects of PME on fish (Courtenay et al. 1998; Parker and Smith 1997). Three industrial EEM pilot studies were conducted in marine and estuarine environments in 1997 and 1998 to develop artificial stream techniques for assessing PME effects on fish.

Saint John Harbour, Saint John, New Brunswick (1997)

The first pilot study was conducted in Saint John Harbour, New Brunswick, using the large mobile mesocosm system (Model II, Figure 9-1B) to assess the effects of a secondary-treated thermo-mechanical PME on a saltwater killifish, the Mummichog (Table 9-1) (Cash et al. 2003; Dubé et al. 2002). The mill discharged to a complex marine environment characterized by extreme tidal fluxes, historical sediment contamination, and the presence of other effluents (e.g., treated and untreated sewage, storm water, another pulp mill effluent, and oil refinery effluent).

The artificial stream system was situated on shore at the end of a breakwater. Receiving water, unexposed to PME, was pumped into each stream during each tidal exchange as described by Cash et al. (2003) and Dubé et al. (2002). Two treatment conditions were created: control receiving water and 3% effluent (v/v). The 3% effluent concentration represented the concentration found over the largest spatial extent in the receiving waters as determined by plume delineation studies. Effluent was dosed for 28 days into each 3% treatment stream in conjunction with receiving water exchanges simulating exposure conditions of the sentinel species (Mummichog) that remain in tidal pools during ebb and low tide (Kneib 1986). Juvenile fish (120 fish per treatment) and adult fish (60 per sex per treatment) were allocated to the control and to the 3% effluent treatments, and were fed daily using frozen brine shrimp at a rate of 3% total stream biomass. Mummichog was selected as the sentinel species because it is well-studied, endemic to Saint John Harbour, a suitable size to place into the streams, and sexually dimorphic for ease in controlling sex ratios (Kneib and Stiven 1978; Atz 1986; Scott and Scott 1988). In addition, juvenile growth rates are high enough to detect effluent-related effects over the exposure period at ambient study temperatures (Kneib and Stiven 1978).

Response variables included effect endpoints congruous with the EEM wild fish survey (i.e., growth, gonad and liver size, condition factor) as well as additional physiological supporting endpoints (mixed-function oxygenase [MFO] induction, reproductive hormone levels) (Cash et al. 2003; Dubé et al. 2002).

This study provided information on mill-related effects for an endemic fish species. The rate of survival was close to 100% in all treatments, and effluent exposure did not affect growth or MFO activity (Cash et al. 2003). However, effluent exposure did significantly reduce gonad and liver size in males and increased production of some sex steroids in both sexes.

Saint John River, Saint John, New Brunswick (1997-1998)

In 1997 and 1998, Dubé (2000) used artificial streams to determine the effects of a PME on Mummichog in the Saint John River and to evaluate changes in final effluent quality associated with a mill process change (Table 9-1). This same large artificial stream system (Model II, Figure 9-1B) and sentinel species were used as above. This study differed with respect to the scope of the hypotheses tested, the type of mill process investigated (bleached-kraft chemical pulping process) and the type of receiving environment studied (estuarine).

In 1997, before the mill process change, adult Mummichog were exposed to final mill effluent (1% v/v) for 27 days (Dubé and MacLatchy 2000a). In 1998, after the mill process change, both adult and juvenile Mummichog were exposed to 3 concentrations of PME (0.5%, 1.0% and 5.0% v/v) for 30 days and 60 days. The large artificial stream system was situated on the Saint John River beyond the zone of effluent influence. Reference water was pumped continuously into each stream to simulate site-specific exposure conditions. Response variables included juvenile growth, adult organ size (liver, gonad), condition, MFO induction, and reproductive hormone levels.

In both studies, the survival rate was > 95% and fish in all treatments increased in biomass throughout the exposure period, showing an adequate feeding rate (Dubé 2000). Exposure to final effluent at 1% did not affect adult organ size (gonad or liver) in either study. However, to illustrate the responsiveness of Mummichog to PME and to support conclusions that adult fish were largely unaffected by exposure to environmentally relevant concentrations of PME at this mill, Dubé (2000) employed a dose-response study design in 1998. Exposure to a 5% concentration of PME for 60 days resulted in significant increases in liver size in both sexes and significant decreases in both the length and weight of juvenile fish (Cash et al. 2003; Dubé et al. 2002).

The artificial stream system was also used in this study to evaluate the effects of a mill process change on final effluent quality (Dubé and MacLatchy 2000a, Dubé et al. 2002). Changes in liver size, gonad size and condition in adult Mummichog were not observed between 1997 and 1998. However, patterns in reproductive hormone levels differed between years, showing significant depressions in plasma testosterone in both males and females in 1997 but not in 1998. Further investigation using toxicity tests (Dubé and MacLatchy 2000b) and laboratory exposures of Mummichog to mill process effluents (Dubé and MacLatchy 2001) in a weight-of-evidence approach confirmed that the process change removed acute toxicity of the final effluent and

significantly reduced sublethal toxicity, including reducing reproductive effects on a local fish species.

1999-2001, Toxic Substances Research Initiative (TSRI)

Artificial stream development occurred in 4 main areas over this period: further development and optimization of technology design, use with other effluents (metal mining), use with other fish species, and use in cumulative effects bioassessment programs with multiple effluents (Table 9-1).

Miramichi River, Miramichi, New Brunswick (1999)

This study evaluated the effects of primary and secondary bleached-kraft PME (1% v/v) on Mummichog after 23 days of exposure, using a redesigned, large artificial stream system (Table 9-1) (Dubé et al. 2002). The system consisted of 16 circular tanks (0.42 m²) on a single trailer for improved transportability (Model III, Figure 9-1C). Improved control over effluent dilution and dissolved oxygen levels was also attained, by redesigning the plumbing and adding an air-lift system (Cash et al. 2003). AMEC Earth & Environmental Ltd. (previously Washburn & Gillis Associates Ltd.) constructed and currently owns the system.

Adult survival was high in all treatments (> 90%) and effluents did not affect length, weight, condition, liver somatic index (LSI) or gonadosomatic index (GSI) after 23 days of effluent exposure. However, both sexes of Mummichog exposed to secondary-treated effluent showed significant, 5-fold depression in plasma testosterone concentrations compared to the control fish. These concentrations were also significantly depressed relative to levels measured in fish exposed to a 1% primary-treated effluent. These results suggest that secondary treatment of some bleached kraft pulp mill effluent may not remove the compounds responsible for depression of reproductive hormones in some fish.

Little River, Bathurst, New Brunswick (2000)

In this study, artificial stream techniques were applied to assess the effects of an MME. In 2000, artificial stream studies were conducted by Culp et al. (unpublished) to evaluate the effects of an MME on benthic invertebrate and algae communities. Dubé et al. (2005) concurrently evaluated MME effects on juvenile Atlantic Salmon through water-borne exposures as well as through exposure in a naturally cultured multitrophic-level food web (algae + benthic invertebrates + fish). Studies were conducted at a mine near Bathurst, New Brunswick. In the first study, the large (Model III) artificial stream system (Figure 9-1C) was used to assess the effects of 20% and 80% (v/v) MME on salmon. The treatment levels for this study were selected to represent current effluent discharge (80%) into the Little River, New Brunswick, and predicted discharge levels upon mine closure (20%). The experiments consisted of 37 days of exposure, and response variables included growth, liver size, condition, metal tissue burdens, and stress variables including levels of muscle glycogen (Dubé et al. 2005). In the second set of studies, the modular stream system (Figure 9-2B) was used to measure benthic invertebrate responses to 20% and 80% MME after 24 days of exposure. Response variables included changes in total invertebrate density, taxon richness, Simpson's Diversity Index, Bray-Curtis Index and insect emergence

(Culp et al. unpublished). In the third set of experiments, the modular stream system (Figure 9-2B) was also used to expose a self-sustaining multitrophic-level food web to 20% and 80% concentrations of MME for 26 days (Dubé et al. 2005). In these multitrophic-level studies, young-of-the-year Slimy Sculpin (*Cottus cognatus*) were placed into streams that had been inoculated with algae and benthic invertebrate communities from a reference river. This permitted assessment of MME effects on fish using a more environmentally realistic pathway of contaminant exposure (i.e., through the food web as opposed to using an unexposed food source).

Wapiti River, Grande Prairie, Alberta (2001)

Mesocosms were used to separate out the confounding effects of a secondary-treated bleached-kraft PME from an MSE on survival, growth, condition and reproduction in adult and juvenile Longnose Dace (Dubé et al. 2004). Longnose Dace were exposed to the following treatments for 42 days: reference river water, PME (3%), PME (10%), MSE (1%), and MSE (1%) + PME (3%). The objective of the dose-response exposure to PME was to examine the response pattern to PME in isolation under low and high concentrations. The MSE and mixture treatments were representative of conditions upstream (MSE 1%) and downstream (MSE 1% + PME 3%) of the PME discharge in the Wapiti River. Results showed that 10% PME slightly reduced juvenile condition and altered some reproductive hormones in adults. Exposure to 3% PME slightly increased juvenile condition, suggesting nutrient enrichment at lower PME concentrations. No effects on survival, growth, liver size, gonad size, or stage of gonadal development were observed with PME exposure. MSE affected reproductive response variables such as male gonad size, female fecundity and some hormone levels in males and females. Hormonal changes after exposure to 10% PME were similar in magnitude to changes measured after exposure to 1% MSE. This study specifically examined the effects of water-borne exposure to PME and MSE on a forage fish after 42 days in a field-based mesocosm.

Culp et al. (2004) examined the cumulative effects of PME and MSE on benthic invertebrate and algal communities. Four treatments were established, as in the above study (i.e., control, 1% MSE, 3% PME, 1% MSE + 3% PME). Replicate benthic food webs were established across all treatments by inoculating each mesocosm stream with substratum, the associated microbes and algae, and invertebrates that were obtained from a reference area. Adult insects were collected from emergence traps placed over each stream every 2-3 days (Figure 9-2B), while benthic invertebrates and algal biomass were sampled at the end of the experiment. The results indicate that both MSE and PME were a significant source of nutrients to the river. MSE appeared to be a primary source of nitrogen, while PME appeared to be an important source of phosphorus and carbon. Algal biomass increased with effluent exposure and was more strongly related to nitrogen than to phosphorus or carbon. Insect emergence data suggested a synergistic rather than additive effect of exposure to the 2 complex effluents (Culp et al. 2004).

2001-2008, Academic/Industry Partnership Applications

Junction Creek, Sudbury, Ontario (2001, 2002)

Junction Creek in Sudbury, Ontario, historically exposed to sediment contamination from decades of mining, receives 3 treated mine effluents, a municipal wastewater effluent, and several other nonpoint-source impacts. In 2001 and 2002, effects of treated MMEs from 3 different mining operations discharging to Junction Creek on 2 fish species—Creek Chub and Pearl Dace—were assessed (Dubé et al. 2006). Treatments tested for 35 to 41 days included reference water, MME #1 (30%), MME #2 (20%), and MME #3 (45%). In 2001, effects on chub included reduced survival (not statistically significant) and depressed testosterone levels. In 2002, chub and dace survival were reduced to less than 60% in MMEs #1 and #3. In addition, the total body weights of male and female dace were reduced after exposure to these same effluents. In 2001 and 2002, responses were most common to MMEs #1 and #3, with consistent increases in nickel, rubidium, strontium, iron, lithium, thallium, and selenium observed across treatment waters and body tissues. These studies identified changes in response variables for fish endemic to Junction Creek and after exposure to mine discharges independent of historical sediment contamination.

After several years of investigation of effluent effects on fish in water-borne exposures, there was a need to further develop the mesocosm systems for trophic-transfer applications. This was based on the fact that dietary pathways of exposure to contaminants are more environmentally relevant than exposures through the water alone (although the latter is certainly most common in aquatic toxicological research). In addition, results from the national EEM assessments for the pulp and paper EEM program were indicating that reproductive effects of effluents on fish were a dominant national response pattern. This suggested that development of the fish mesocosm method should focus on dietary exposure pathways as well as a more thorough evaluation of reproductive response variables.

In 2002, we developed an in situ life-cycle bioassay with the chironomid *C. tentans* in the modular artificial streams, to evaluate the effects of an MME under ambient environmental conditions in Junction Creek, Ontario (Hruska and Dubé 2004). The chironomids were exposed throughout their life cycle to MME #3, which is the average effluent concentration measured in the creek. *C. tentans* in the effluent treatment exhibited reduced survival, total emergence, hatching success and increased time to emergence. This research showed how a life-cycle bioassay could be used in situ to assess MME effects on a benthic invertebrate. In addition, valuable information was obtained on *C. tentans* growth rates, hatchability and survival in mesocosms, which is information required for improved development of culture-based multitrophic-level mesocosm systems.

In 2003, development of the *C. tentans* mesocosm continued for assessment of MME (Hruska and Dubé 2005). The utility of this test was compared to an existing standard beaker life-cycle bioassay under laboratory conditions. *C. tentans* larvae were exposed to 45% (v/v) treated MME #3 from day 11 through hatching of the second generation. Response patterns were consistent between the 2 bioassays for hatching success and time to emergence but inconsistent for other variables. Significant effects were obtained for growth, survival, number of adults emerged, and number of eggs per egg case in the artificial stream bioassay but not in the beaker bioassay. Conversely, significant effects on sex ratio and number of egg cases per female were observed in the beaker bioassay but not in the artificial stream bioassay. These differences are believed to be a consequence of the number of organisms per replicate used in each bioassay,

which results in a difference in statistical power. As a result, higher coefficients of variation and effects sizes were observed in the beaker bioassay relative to the artificial stream bioassay for almost all variables. These results provided evidence that the mesocosm approach was an effective tool for evaluating the effects of MME on life-cycle variables in *C. tentans*. It is recognized that the EEM program focuses on benthic invertebrate community structure and not individual benthic species. However, these studies were necessary to set the scientific basis for developing a culture-controlled multitrophic mesocosm, as well as to provide mesocosm options that might be of value when programs move into investigation-of-cause phases and may require more detailed information—especially in cases where benthic communities are dominated by chironomids.

The *C. tentans* mesocosm approach was valid to serve as the self-sustaining food base for a fish mesocosm. This would increase the relevance of the fish mesocosm to more natural exposure conditions wherein fish are exposed to effluents through both water and diet. This was a critical improvement, as many metals are known to affect fish through dietary pathways. Another improvement that was required for the fish mesocosm was to increase the relevance and significance of the response variables investigated.

Terrace Bay, Ontario (2003)

Exposure of fish to a contaminant through a partial-life-cycle experiment provides the opportunity to examine the direct effects of effluent on reproduction in adults as well as effects on offspring. In addition, standard EEM effect endpoints (condition, relative liver size, relative gonad size) can also be investigated. Fathead Minnow is a toxicological workhorse used to assess and screen contaminants worldwide for endocrine-disrupting substances. Short-term (7-day), medium-term (21-day) and long-term (full life cycle) tests have been developed for Fathead Minnow. These tests provide an opportunity to directly assess effects on actual reproductive performance (number of eggs, size of eggs, number of spawning events) as well as more indirect measures such as gonad size. However, almost all of the studies in the literature using Fathead Minnow are water-borne exposures that allow for toxicant screening, but at the expense of greater environmental realism. The first objective was to assess if the 21-day Fathead Minnow test could be implemented in the field with natural reference as dilution water for PME assessment, holding water temperature and photoperiod constant. The second objective was to link the chironomid *C. tentans* life-cycle bioassay with the partial-life-cycle bioassay of the Fathead Minnow to develop the multitrophic mesocosm system.

In 2003, a 21-day Fathead Minnow test was implemented at a pulp mill in Terrace Bay, where reproductive effects on wild fish have been documented. The first objective was to determine the effects of PME on Fathead Minnow at 1% and 100% concentrations (Rickwood et al. 2006a). The second objective was to use the Fathead Minnow test to identify waste stream sources within the mill that affect reproductive indicators (Rickwood et al. 2006b). Various process streams were selected, characterized with respect to effluent chemistry and acute toxicity, and a subset were tested on-site with the bioassay. An enclosed mobile bioassay trailer (photo not shown) was set up on-site at a bleached-kraft mill for 60 days, allowing supply of both ambient water (Lake Superior, Canada) and final PME. This was not an outdoor, exposed mesocosm system, as the interest was in evaluating if the Fathead Minnow 21-day bioassay could be used with ambient

reference water and holding other factors (temperature and photoperiod) constant. The results demonstrated a stimulatory response pattern at 1% PME (e.g., increased egg production, cumulative spawning events) compared to the controls. In the 100% PME treatment, spawning was delayed, resulting in fewer eggs produced in the first 2 weeks of exposure. Exposure to 100% PME also resulted in ovipositor development in males and development of male secondary sex characteristics in females. The results for the second objective showed that both the combined mill effluent (before secondary treatment) and the combined alkaline stream (CALK) caused decreased spawning events (~ 55% for both streams) and decreased egg production (28% and 74%, respectively), and the CALK stream resulted in significant male ovipositor development. By comparing response patterns, the CALK stream was identified as a source of the compounds affecting reproductive indicators in Fathead Minnow at this mill.

Junction Creek, Sudbury, Ontario (2004-2005)

Development of the life-cycle bioassay in mesocosms with the chironomid *C. tentans*, and the 21-day partial-life-cycle bioassay with Fathead Minnow using natural reference water as dilution water, established the foundation for a multitrophic mesocosm bioassay. The objective was to develop a self-sustaining trophic-transfer bioassay, using *C. tentans* and Fathead Minnow, that made it possible to assess the effects of water-borne (Fathead Minnow only) and food- and water-borne (trophic-transfer) exposure to MME. The reproductive performance of Fathead Minnow was assessed for 21 days under controlled laboratory conditions to obtain baseline data on various parameters, including egg production and hatching success (Rickwood et al. 2006c). Exposure to MME #3 (see above) was then conducted for a further 21 days in the laboratory. It was evident that reproductive output in both the water-only and the trophic-transfer system was reduced compared to controls. It was only in the trophic-transfer system that a significant reduction in larval hatching and an increase in deformities occurred after exposure to the MME. This would suggest that contaminated food was an exposure pathway for effects on offspring.

The multitrophic mesocosm was then taken out into the field for application in 2005 (Rickwood et al. 2008). The objectives were to assess (1) the effects of a mine effluent and municipal wastewater mixture on Fathead Minnow reproduction in an on-site artificial stream and (2) the importance of food (*C. tentans*) as a source of exposure using a trophic-transfer system. Exposures to the effluent mixture through the water significantly reduced egg production and spawning events. Exposure through food and water using the trophic-transfer system significantly increased egg production and spawning events. Embryos produced in the trophic-transfer system showed a similar hatching success, but also showed an increased incidence and severity of deformities after exposure to the mixture. It was concluded that the effects of the effluent mixture on Fathead Minnow were more apparent in water-borne exposures. Exposure through food and water may have reduced effluent toxicity, possibly due to increased nutrients and organic matter that may have reduced metal bioavailability.

Wabigoon River, Dryden, Ontario (2006)

This study investigated the link between PME and endocrine disruption in an attempt to explain the presence of intersex fish in the Wabigoon River, Ontario (Table 9-1; Pollock et al. 2009). A field survey of the Wabigoon River near Dryden, Ontario, in the fall of 2000 found intersexed

Walleye (*Sander vitreus vitreus*) with significantly altered hormone levels and reduced gonad size. The Wabigoon River receives discharge from a bleached-kraft pulp and paper mill and MSE. It also has historical wood-fibre mats contributing to extended periods of low dissolved oxygen under low-flow drought conditions. A partial-life-cycle test was conducted exposing Fathead Minnow to reference water and to 20%, 40% and 60% PME in field mesocosms. A field survey of Walleye in the Wabigoon River was also conducted. Testosterone decreased in males with increasing effluent concentration, and vitellogenin induction occurred in males exposed to 60% PME. These results did not reflect the magnitude of endocrine disruption seen in the wild fish survey. Several hypotheses have been proposed to explain these discrepancies. Specifically, evidence from published studies indicated that either hypoxia or MSE, alone or in combination with PME, may explain the discrepancy between the field experiment and the wild fish survey. Later studies at this site have examined the effects of low dissolved oxygen levels on Fathead Minnow as well as the interactive effects of low dissolved oxygen (6.0 milligrams per litre [mg/L] no-effect level) and PME (40% no-effect level) (Dubé unpublished).

2007-2009, Ongoing Mesocosm Development

The development of fish mesocosm applications has been consistent and ongoing, with improvements in the methodology with each application. The assessment of effects on small-bodied fish endemic to the system of interest, or assessment using a partial-life-cycle Fathead Minnow test in outdoor mesocosms (water-borne or trophic-transfer experimental designs), is now fairly straightforward. Future investigations are now focusing more on applications in the investigation of cause for the metal mining EEM program.

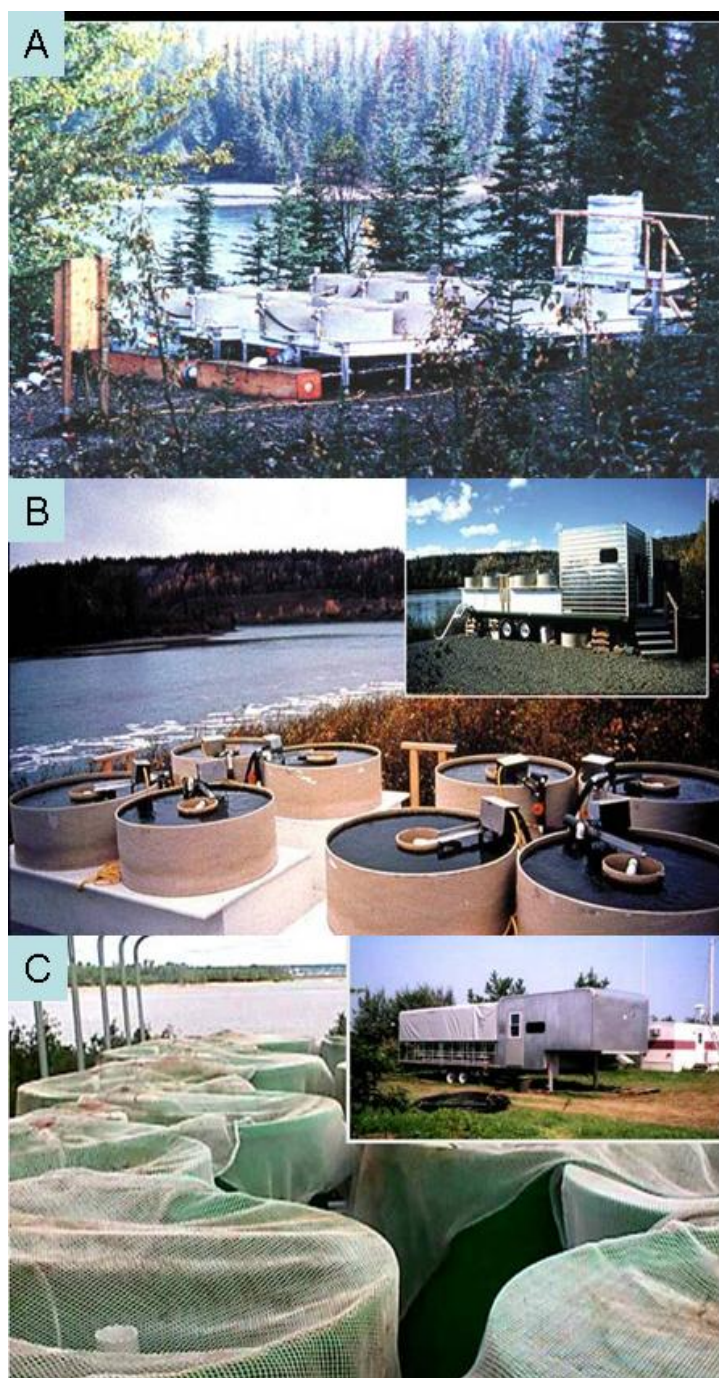
Junction Creek, Sudbury, Ontario (2008-2011)

Due to the complexities of aquatic ecosystems, the understanding of how metals and metal mixtures affect river food webs is limited. Furthermore, the assessment of effects is only the first step toward the mitigation of those effects and, ultimately, sustainable development. Understanding the causes of the effect (e.g., causative metals) and the factors that modify toxicity is the next step toward the investigation of solutions. Ongoing mesocosm research in Sudbury, Ontario, will *i*) confirm responses of Fathead Minnow to MME on-site using self-sustaining multitrophic-level bioassays; *ii*) contrast and compare minnow response patterns to whole effluent mixtures relative to effluent-equivalent doses of single metals of potential concern (copper [Cu], selenium [Se] and thallium [Tl]); *iii*) ascertain the relative importance of water and diet as the pathway of exposure causing toxicity of metals to Fathead Minnow; and *iv*) explore factors with the potential to modify toxicity (pH/alkalinity and natural organic matter, and diet quality and quantity) of effluent mixtures and dominant single metals (Cu, Se and Tl) to Fathead Minnow (Dubé et al. unpublished).

Key Lake, Saskatchewan (2007-2011)

This study is being conducted over 4 years, also using a combination of field-based and laboratory-based mesocosm studies (Driessnack et al., unpublished, Dubé et al. unpublished). The objective of the laboratory mesocosm study is to assess changes in an aquatic food chain (multitrophic mesocosm), including reproductive output of Fathead Minnow due to exposure to a

uranium effluent. In addition, a comparative evaluation of Fathead Minnow response patterns to the effluent mixture vs. Se as selenate will be conducted. This experimental design isolated the contribution of Se from that of the effluent mixture. Results illustrated that the response patterns in egg production could not be explained by Se in the comparison between treatments and relative to controls. The objective of the field mesocosm study was to determine the relative and cumulative contribution of water-borne (current) vs. sediment-borne (historical) Se contamination on the reproductive success and survival of breeding Fathead Minnow and their offspring. Results showed that effects on fathead minnow were exclusively effluent mediated with insignificant contributions from contaminated sediments. The sediments tested in the study were of sand composition as it represented the largest sediment type in the Key Lake drainage. Further work is required to determine the significance of organic sediments to responses and in the context of their spatial and temporal distribution at the site.



© M. Dubé

Figure 9-1: A) Large mesocosm system with streams situated on tables (Model I) used in the Athabasca River, Alberta. B) Large mobile mesocosm system with streams on 2 trailers (Model II) used in the Fraser River, British Columbia; the Saint John River, New Brunswick; and in Saint John Harbour, New Brunswick. C) Large mobile mesocosm system with streams on a single trailer (Model III) used in the Miramichi and Little rivers, New Brunswick; the Wapiti River, Alberta; and Junction Creek, Ontario.

A)

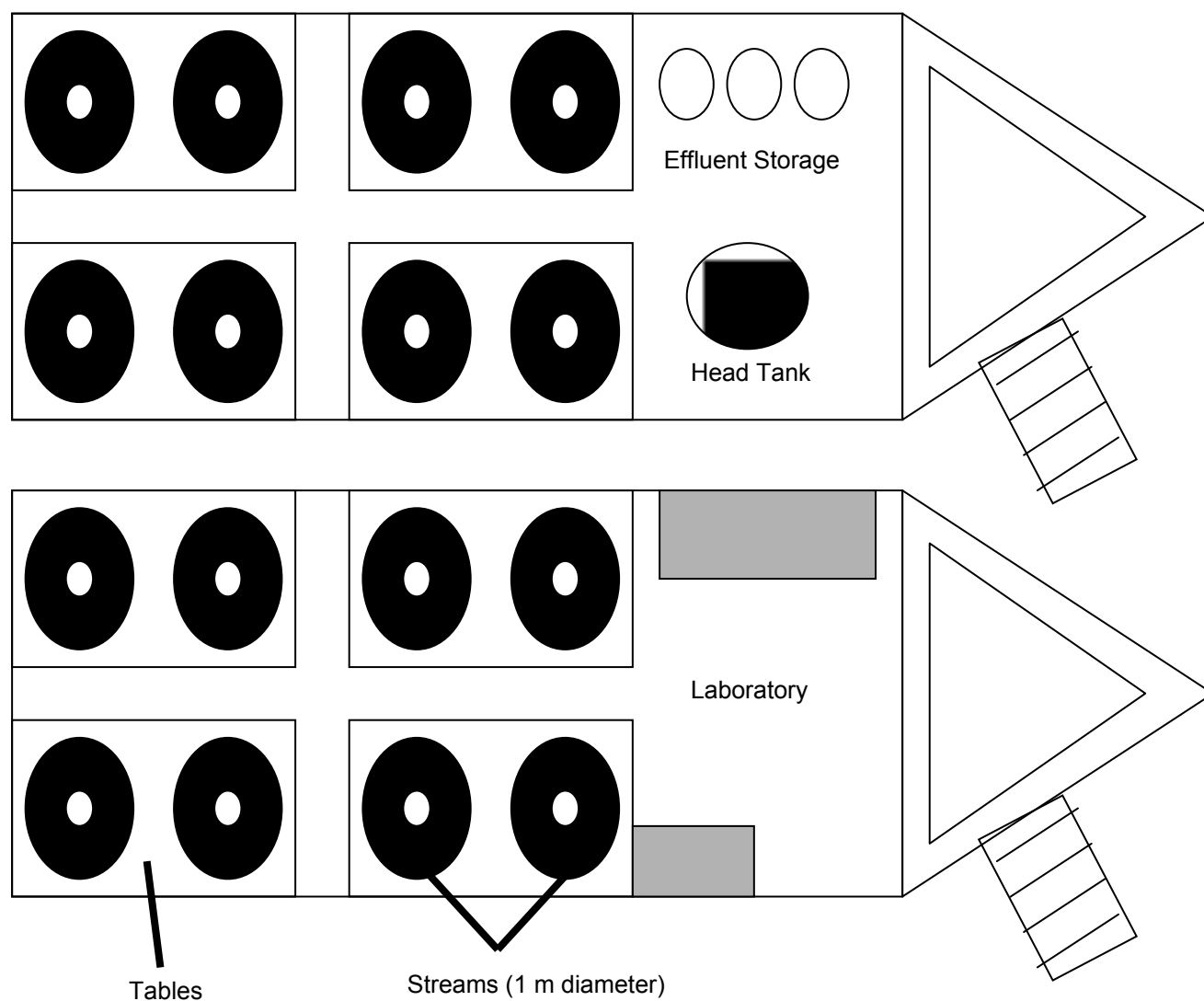


B)



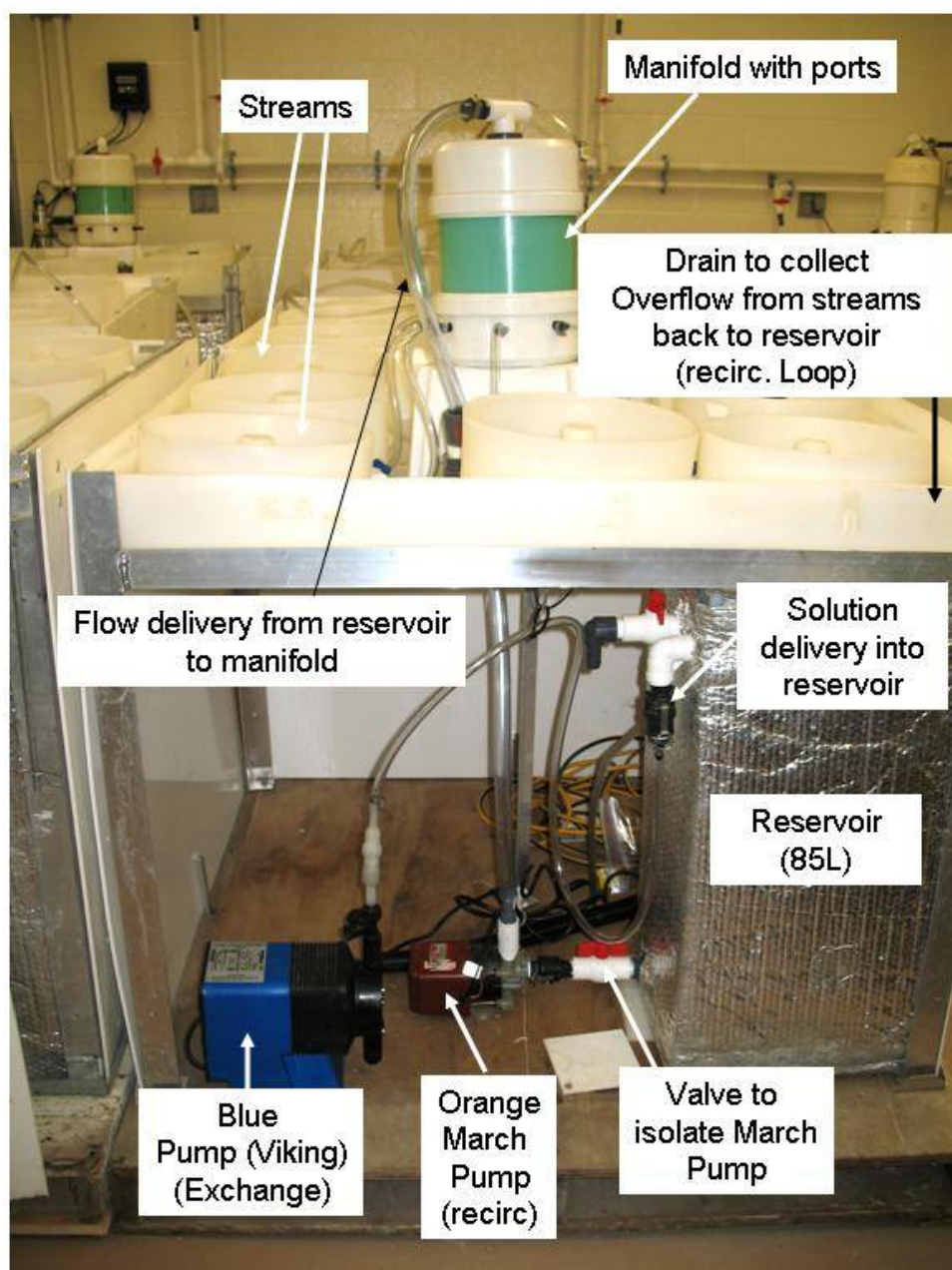
© M. Dubé

Figure 9-2: A) Small microcosm system with streams situated on tables over mixing reservoirs used in the Thompson River, British Columbia. B) Modular mesocosm system with streams situated on tables over mixing reservoirs used in the Little River, New Brunswick; Junction Creek, Ontario; the Wabigoon River, Ontario; and Key Lake, Saskatchewan.



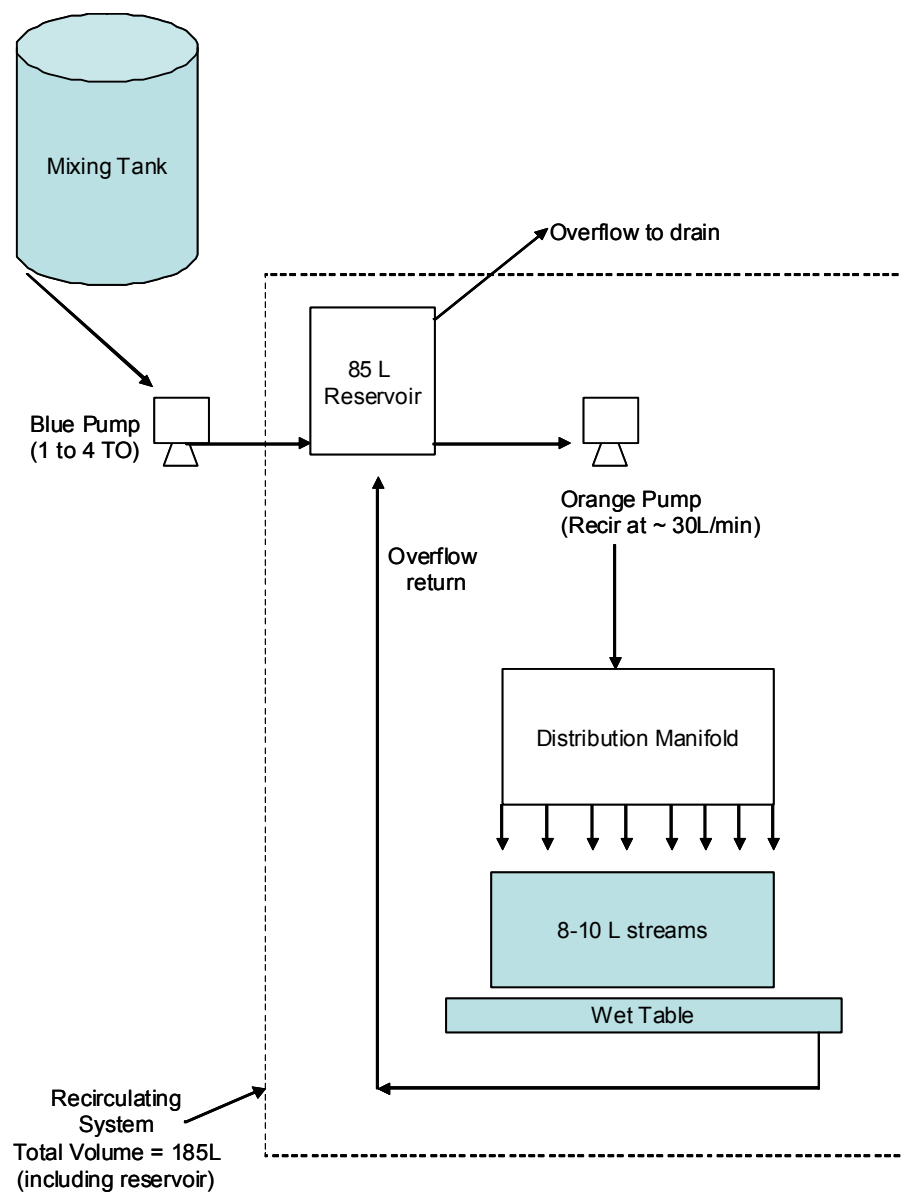
© M. Dubé

Figure 9-3: Schematic of large mesocosm trailer system (not to scale)



© M. Dubé

Figure 9-4: Photograph of a modular mesocosm parts diagram.



© M. Dubé

Figure 9-5: Modular mesocosm flow schematic.



© M. Dubé

Figure 9-6: Multitrophic Fathead Minnow reproductive bioassay and feeding barrier.



© M. Dubé

Figure 9-7: Site set-up for modular mesocosms

A)

		WATER	
		Reference	Effluent
FOOD	Reference	Fish held in reference water and fed chironomids raised in reference water	Fish held in effluent and fed chironomids raised in reference water
	Effluent	Fish held in reference water and fed chironomids raised in effluent	Fish held in effluent and fed chironomids raised in effluent
		Multi-trophic	Multi-trophic

B)

	Ref pH & NOM	↑ pH	↑ NOM
Ref			
Effluent			
Metal			

© M. Dubé

Figure 9-8: A) Factorial experimental design to investigate the importance of water vs. diet in responses of Fathead Minnow to metal mine effluent in modular mesocosms; B) Experimental design to investigate the influence of pH and natural organic matter (NOM) on Fathead Minnow responses after exposure to an MME mixture and a single metal in multitrophic modular mesocosms.

		WATER	
		REFERENCE (Reference Lake)	CONTAMINATED (25% Effluent; conc. at Nearfield Zone)
SEDIMENT	REFERENCE (Reference Lake)	<i>Complete Reference (RWRS)</i>	<i>Assess contribution of effluent alone (EWRS)</i>
	CONTAMINATED (Contaminated Lake)	<i>Assess contribution of sediment alone (RWCS)</i>	<i>Complete Exposure. (EWCS)</i>

© M. Dubé

Figure 9-9: Factorial experimental design to investigate the effects of MME and historical sediment contamination in isolation and in combination on Fathead Minnow in modular mesocosms. Legend: RWRS: Reference Water and Reference Sediment; EWRS: Effluent Water and Reference Sediment; RWCS: Reference Water and Contaminated Sediment; EWCS: Effluent Water and Contaminated Sediment

9.2.2 Applicability within the EEM Programs

It should be emphasized that while mesocosms are a recommended monitoring alternative, their use should only be considered when field surveys cannot be designed to unequivocally answer the hypothesis, or simply cannot be conducted. Such situations include confounded receiving environments or where unsafe sampling conditions exist. Examples of confounded receiving environments include areas with the presence of historical effects, the absence of suitable reference areas for comparison to exposure areas, the presence of other effluents, and changes in relevant habitat types that cannot be factored out in the design of a field survey. Mesocosms can also be used for assessment of magnitude (dilution series) and for investigation of cause. The above-described case studies illustrate the different types of questions and experimental manipulations that can be applied to the different phases of the EEM program.

It may be possible to use the fish from a fish survey mesocosm study for the fish usability component of EEM as well, provided that the effect endpoint has the potential for responding over exposure periods typical of mesocosm studies. In this case, the effect endpoint and statistical procedures would be as outlined in Chapter 3.

There are advantages and disadvantages to using mesocosms, and these should be weighed against the advantages and disadvantages of other monitoring alternatives before the final selection of an approach.

9.2.3 Mesocosm Technology

9.2.3.1 Suitability of Mesocosm Design

To maintain data quality and ensure consistent application of mesocosm technologies in EEM programs, a review of the physical design of the system used and the study design, including standard operating procedures, is required. This section outlines physical design guidance and experimental design recommendations for both a large trailer mesocosm system as well as for the smaller modular design.

Regardless of the system used, general operation is the same. Replicated tanks or streams hold the biota (fish, benthic invertebrates), algae, and/or substrate of interest. A total of 5 to 8 replicate streams per treatment is considered adequate. The systems are not static but either completely flow-through (trailer mesocosm) or partially recirculating (modular mesocosm). Ideally, reference water and dilution water are collected from a reference site and delivered to a head tank. Treated effluent is collected daily, or no less frequently than weekly, and stored on-site in a head tank. Head tank liquids can be heated or chilled depending on the circumstances. Mixing tanks are used to mix reference water and effluent to the desired test concentration. Mixed water or “treatment water” is delivered from each mixing tank to the mesocosm streams at a flow rate to achieve a target turnover time (minimum of one turnover or complete stream volume exchange every 24 hours and up to 6 turnovers per 24 hours). The experiment is typically run for 30 to 65 days, with daily and weekly measurements taken of physical and chemical variables, flows, fish mortality, and reproduction (if eggs are collected

daily, for example). At the end of the exposure period, fish are examined and supporting measures are taken.

9.2.3.2 Description and General Operation of Large Trailer Mesocosm System

The large trailer mesocosm system, developed by the National Water Research Institute of Environment Canada, has been used to assess PME and MME effects on benthic food webs (benthic invertebrates and algae) (Culp and Podemski 1996; Culp et al. 1996, 2000a, 2001; Cash et al. 2003) and small-bodied fish (Dubé 2000; Dubé and MacLachy 2000a; Cash et al. 2003; Dubé et al. 2002). This system has been used for benthic invertebrate community assessments as well as water-borne exposures with fish. The system can still be used, although most applications now use the modular mesocosm system described in the next section.

The large trailer stream system consists of 16 circular tanks or streams, with a surface area of 0.9 m², placed in pairs on tables that are 74 cm high (figures 9-1 and 9-3). Water from a reference area is pumped into a 378-L polyethylene head tank placed on a platform that is 1.2 m high, and gravity-fed through a system of pipes to the streams. Intake pumps such as a land-based pump (e.g., commercially available irrigation or pool pumps) or a submersible pump (0.5-HP Hydro-Matic Model # SPD50-H) can be used, depending upon volume requirements, pumping distance, and the head required. Pumping reference water from a mine water intake is also a possibility. Gate valves control water flow to individual streams and allow flow rate calibration for each stream. Water is delivered to each stream at a rate of 2 L/min, resulting in a total water requirement of 32 L/min for the 16-stream system.

Water depth in the streams is maintained at 26.9 ± 0.1 cm ($\bar{x} \pm 1$ SE) by an overflow drain that returns all wastewater to the river. The overflow drains are screened to prevent fish loss and limit emigration of insects from the streams. Each stream contains 227 L of water, resulting in a hydraulic residence time of approximately 2 hours. By increasing water residence time within the streams to 4 hours, the volume of effluents and reference water required during a study is minimized. Final determination of residence time for a particular system depends upon the size of the tank used and species requirements for oxygen and temperature. If mesocosm studies are conducted in the late autumn, the head tank and water delivery lines can be wrapped with heat tape and insulated to allow the system to be operated when freezing temperatures may occur (-5°C). Shade cloth can be used to reduce solar heating in the summer.

The streams on the large mesocosm system are tanks, 107 cm in diameter, constructed of polyester fibreglass. Streams are placed on 8 tables that are 74 cm high, 2 to a table. The water outflow pipe passes through a standpipe and drains into pipes beneath. These drain pipes connect to a general outflow pipe for discharge to the river downstream of the water intake point.

For invertebrate applications, current velocity in each stream can be created using a propeller system (Podemski 1999). Water velocity in each stream near the water column midpoint is normally maintained at 20 cm/sec, although site-specific velocities should be the determining factor. Other current-generation mechanisms can be used, provided they produce the water velocities. In fish studies, current can also be generated using the propeller system, although

spray bars attached to the water delivery system for each stream have also been used for species where water velocity is not a critical requirement.

Effluent is collected daily or every second day and stored in polyethylene containers. The point of effluent collection depends upon the study design. For the first monitoring study at a metal mine, for example, the final effluent that is representative of what is being discharged to the aquatic environment is the target source. Effluent treatments are delivered independently and continuously to individual streams by peristaltic pumps (Masterflex® L/S Nema-type 13 wash-down controllers and cartridge pump heads). Effluent flow rates depend upon the site-specific, environmentally relevant concentrations to be tested. For example, if plume delineation studies in the field have determined effluent concentrations to be 1%, then an effluent flow rate into each stream is set at 20 ml/min (1% of 2 L/min). If there are 16 tanks on the trailer, which are allocated between 2 treatments (8 streams for control; 8 streams for 1% exposure), then the following effluent volume is required on a daily basis: 20 ml/minute x 60 minutes/hour x 24 hour/day = ~ 29 L/day/stream x 8 streams = ~ 231 L/day. This volume would fit into a small polyethylene container.

After water and effluent flows have been calibrated for each system, the individual tanks or streams are seeded with natural substrates, algae, benthic invertebrates and/or fish species endemic to the receiving environment being studied. Inoculation of biotic populations is described below for fish and invertebrates. Further details on the construction of this trailer system can be found in the literature referenced herein.

9.2.3.3 Description and General Operation of Modular Mesocosm System

Each modular mesocosm unit or table consists of a shipping pallet, metal frame, wet table, up to 8 replicate circular polyethylene streams, a reservoir for holding the exposure solution, a manifold for equal flow distribution to the replicate streams, a blue Viking pump for flow delivery from the mixing tank to the reservoir, and an orange March pump for flow recirculation within the unit from the reservoir to the manifold/stream complex (figures 9-4 and 9-5).

Each mesocosm holds a total capacity of 185 L of water (85-L reservoir, 82.4 L = 10.3 L/stream x 8 streams, ~ 15 L in manifold/hoses/pumps) and requires 3.0 amps to run without heating, cooling or aerating the water. Each replicate circular stream has a volume of 10.3 L. There is a central non-functional standpipe in each stream. The circular high-density polyethylene streams sit on top of a table that drains into an 85-L dilution reservoir. Each stream requires a cover to keep fish in and other things (e.g., birds) out. A window screen or Nytex mesh cover is used, and secured with bungee cords. Each table has 8 polyethylene streams of 10.3 L each that are custom-made and moulded. The reservoirs are 85-L polyethylene plastic totes that are also custom-made and moulded. The 8-port manifold system was custom-designed to allow for equal flow distribution to each stream without requiring 8 separate pumps. There are 2 types/sizes of tubing in the mesocosm system to deliver water to and from the manifold and to the streams (internal diameter [ID] 3/8" and ID 3/4").

An electronic metering pump (blue Viking/Pulsatron pump, Series E 240 GPD LEH 75A-PHC3-XXX) controls the water/effluent turnover times in the reservoir. The March "Series 3" Seal-less

Magnetic Drive Centrifugal Pump (orange pump) controls the movement of water/effluent from the reservoir to the streams. Pumps are the most expensive part of this system and the most critical component of the study. An incorrect flow rate means incorrect dilutions to the test organisms.

Each mesocosm unit or table represents a treatment with 8 replicate streams per treatment. Note that it can and has been argued that the systems are pseudo-replicated at the level of the reservoir. This point has been successfully argued and defended through the peer-review publication process. For fish mesocosm application, streams can be fitted with feeding barriers for multitrophic studies. This mesh barrier will allow a benthic invertebrate culture to develop under treatment conditions, while controlling access for the fish above it (Figure 9-6).

Researchers have experimented with many different forms of flow delivery, including in-line mixing pumps. Based on experience, and in the interest of keeping costs manageable, mixing the treatment solutions in a mix tank and delivery using a pump to each mesocosm reservoir has become the preferred option over that of in-line mixing of water + contaminant. Water is delivered from a mixing tank to a mesocosm reservoir using a blue Viking pump at a rate of 1-4 turnovers every 24 hours (Figure 9-5). There is an overflow drain at the back of each reservoir and a baffle inside each reservoir to prevent short-circuiting of the inflow to the overflow drain.

The March pump is the recirculating pump. Flow moves from the March pump through the manifold, into the streams, overflows from each stream into the wet table, and then drains back into the reservoir through the hole in the table. The manifold must be level and free of air bubbles to operate effectively. All tubing must be the same length for the manifold to be able to pressurize equally and deliver water at an even flow rate to each of the streams. Water enters the stream along the length of water inlet tubes threaded through the stream wall. Because of the angle at which the water enters the stream, a slight circular current is created. Water fills the streams and then drains over the top, collects briefly on the tabletop, and is rerouted back down to the reservoir. Reservoirs are insulated using silver insulation sheets wrapped around the outside of the tank.

9.2.4 Considerations for Site Selection

Irrespective of the mesocosm used, there are some basic site requirements. The system is typically located at a reference area for access to reference water. In a freshwater riverine situation, reference water is pumped from upstream of the effluent outfall into the system and discharged back to the river downstream from the point of intake. This is the most straightforward scenario for mesocosm use. Distance to the reference water source is also a consideration for site selection, as extensive pumping requirements with respect to distance, or elevation (head), can exceed pump specifications for the rate required. Other site requirements include adequate space, site access, electrical power and security.

Access to power is one of the major site-selection requirements. The facility typically provides this, via a power line that the facility's electricians install into the mesocosm's power panel on either the trailer or the power pallet designed specifically for the modular mesocosm system. In remote areas, if power is not available, generators have been used to power the system, although

this approach is not recommended. Due to maintenance and supervision needs, generator use is often not cost-effective.

During site set-up, the process consists of selecting the site, unloading equipment, and placing the head tanks and mixing tanks at the end opposite the power pallet (in the case of the modular system). Modular mesocosms are typically placed under a tent for shade, to reduce particulate deposition (dust), and for security purposes. The head tanks, mixing tanks and power pallet do not go under the tent (Figure 9-7). The modular mesocosm or trailer is placed in a north-south direction perpendicular to sunset/sunrise. First, all equipment is placed, then power is connected, mesocosm tables are levelled, tubing is run from head and mixing tanks to streams, all tanks are filled, pumps are turned on and calibrated, tents are set up (for modular mesocosms), and then animals are added.

It is critical to calculate the electrical requirements of the system for the experimental design selected and to assess the site to determine adequate electrical availability. Power reliability at industrial sites is an important consideration.

9.2.5 Biological Monitoring Study Designs

9.2.5.1 Overview

The objective of a study design is to outline what mesocosm and associated laboratory work is needed to complete the biological monitoring portion of the EEM study. Many of the study components are similar to those outlined in the various chapters of this document.

A study design is submitted to the Authorization Officer at least 6 months prior to the commencement of sampling for biological monitoring studies. The study design will include:

- a site characterization;
- a description of how the studies respecting the fish population and fish tissue will be conducted and how these studies will provide the information necessary to determine if the effluent has an effect on the fish population and fish tissue;
- the fish species, sampling areas, and sample size selected;
- A detailed timetable for conducting the mesocosm study;
- a description of how the study respecting the benthic invertebrate community will be conducted and how this study will provide the information necessary to determine if the effluent has an effect on the benthic invertebrate community;
- the dates and times that the samples will be collected for the biological monitoring;
- the field and laboratory methods selected;
- a description of the quality assurance and quality control (QA/QC) measures that will be implemented to ensure validity of the data that is collected; and
- a summary of the results of any biological monitoring studies that were submitted previously.

Other recommended details of the study design may include:

- defining the goals and expectations of the EEM study;
- determining the overall approach, including stating the rationale for choosing an alternative, which may be based on previous monitoring results;
- establishing statistical design criteria: development of hypotheses, selection of statistical methods, determination of data needs (statistical significance and power analysis);
- developing operating plans and procedures: sampling procedures, laboratory analysis procedures, QA/QC procedures, data storage and retrieval, data analysis; and
- describing a plan for data interpretation and program evaluation.

9.2.5.2 Study Treatments

In mesocosm studies, the effect of effluent on fish is evaluated by comparing effect indicators (growth, reproduction, condition and survival) between fish held under control conditions (reference water) to those held in effluent. In mesocosm studies using benthic invertebrates, an effect is determined by comparing effect indicators (total benthic invertebrate density, taxa richness, evenness index and similarity index) between invertebrates in reference vs. exposure streams. It is crucial that mesocosm studies be designed to maximize the possibility of detecting effects if they are present. This includes selecting appropriate treatments, level of replication, sentinel species, response variables, and conducting the studies at the proper time of year.

Mesocosm studies provide for controlled experimental manipulation, with the added benefit of environmental relevance (natural water quality, photoperiod, water and air temperature). The flexibility in experimental design is one of the most significant and creative advantages of using mesocosms. Several study designs can be employed, depending upon site-specific requirements and the phase of the monitoring program (i.e., magnitude and geographic extent, and investigation of cause). In the simplest case, 2 treatments (control vs. effluent exposed) are compared. It is recommended that the environmentally relevant concentration of effluent be based upon a plume delineation study conducted during site characterization (Chapter 2); thus it should represent the effluent concentration in the high effluent exposure (near-field) area after complete mixing.

In some cases, additional treatments may be desired. Dose-response study designs where additional and higher effluent concentrations are used are helpful to confirm biota responsiveness and the absence of effects (see Dubé and MacLachy 2000a; Culp et al. 2000a). Additional treatments may also be desired if more than one mine discharges into the same sampling area. Each discharge can represent an exposure treatment in the experimental design and the effects of each effluent can be examined in isolation or in combination. For example, if a mine discharges 3 effluents into the same receiver, effluent effects can be evaluated using the following treatments: control, effluent 1 (environmentally relevant concentration), effluent 2, effluent 3, effluent 1 + 2 + 3 (environmentally relevant concentration; see Table 9-1).

More recently, modular mesocosms have been used to assess water vs. dietary pathways of exposure for Fathead Minnow exposed to MME. This design was of value for investigating different metals in effluents and their causal contribution to fish response patterns. Fish were held in water (reference or effluent) and fed with the chironomid *C. tentans* cultured under either control water or effluent exposure conditions (Figure 9-8A). These results were compared to

those of concurrent multitrophic mesocosm treatments. Studies have also been conducted to evaluate the influence of different water chemistry variables in ameliorating the toxicity of a metal and MME on Fathead Minnows by focusing on the effects of increased pH and natural organic matter using multitrophic mesocosms (Figure 9-8B).

9.2.5.3 Replication

In modular mesocosm systems with benthos or fish, typically 5 to 8 replicates are used. In the trailer mesocosm system employing a control/impact design with 2 treatments (reference vs. exposure), 8 replicates will be used as there are 16 tanks on the mesocosm trailer. If additional treatments are preferred, then 5 replicate streams per treatment should be the minimum number of replicates for EEM studies, as this should provide adequate power for assessing effects when streams are the level of replication and when no previous monitoring data are available. This approach is consistent with the recommended method for determining the number of sampling stations, using statistical power in field survey designs (Chapters 3 and 4).

The unit of replication is less clear when the system is used to measure individual-based response variables in fish. The basis of the decision lies in the quantification of the importance of a tank or stream effect (i.e., an effect of one tank relative to another within the same treatment). In many laboratory studies where fish are held in aquaria, an assumption is made that there is no biological reason for tank differences within a treatment, and thus all individual fish measurements are pooled within a treatment (i.e., the variability attributable to tank effect is often ignored). For mesocosm studies it is suggested that a tank effect may exist, and the variation attributed to that effect requires some consideration in the statistical design. This is especially the case for longer-term exposures (e.g., 60 days). A nested analysis of variance (ANOVA) could be used in this example where the tank effect is nested within the fixed-effect factor of effluent treatment. If a tank effect exists, the level of replication should be at the stream level. If tank effects do not exist, fish can be selected as the unit of replication and pooled across streams within a treatment, resulting in significant increases in the number of replicates. For regression-based analyses (analysis of covariance [ANCOVA]), any tank differences would likely emerge as outliers in the analysis.

9.2.5.4 Sample Sizes and the Role of Effect Size

9.2.5.4.1 No Available Pre-existing (Historical) Data

Trailer Mesocosm System: In the trailer mesocosm, there is increased capacity to hold fish for longer periods than in the modular mesocosm system. For a standard fish population field survey, where there are no background monitoring data, the minimum sample size recommended is 20 sexually mature males and 20 sexually mature females of 2 fish species collected from each of the reference and exposure areas. If small-bodied fish species are chosen as one or both of the fish species, an additional 20 sexually immature fish should also be collected. The rationale for using 20 fish of each sex is that there is little change in the 95% confidence limits with increasing sample size beyond 20 fish. In trailer mesocosms, 15-20 sexually mature males and 15-20 sexually mature females of one small-bodied sentinel species are added to each tank. In addition, 20-30 juvenile fish are allocated to the same tank. These are recommended sample sizes

that can be increased or decreased depending upon the species selected, the statistical power of the study design, and fish variability. If tanks within a treatment are selected as the unit of replication, 20 fish per sex per tank provides a mean with good precision (i.e., there is little change in the 95% confidence limits with increasing sample size beyond 20 fish). If tank effects do not exist and error terms are pooled so that fish become the unit of replication, 15-20 fish per sex per tank provides much higher replication per treatment. For example, a study design of control vs. 1% effluent with 8 streams per treatment, and 20 males, 20 females and 20 juveniles per tank, can result in 160 males, 160 females, and 160 juveniles per treatment if fish are the unit of replication.

Modular Mesocosm System: In modular mesocosm studies using benthic invertebrates or fish, sample size represents the number of streams replicated per treatment, as streams are the unit of replication. In a control-impact design, 8 streams are replicated per treatment. Each mesocosm table represents a treatment and thus this simplest design would only require 2 mesocosm tables for assessment. If additional treatments are required, a minimum of 5 replicates or samples are required for each treatment to address power requirements without pre-existing data (see below). Breeding pairs or trios can be used in each stream depending upon the objectives of the study. In the protocol developed by Ankley et al. (2001), a ratio of 2 males to 4 females (2M:4F) is used per replicate. However, this breeding ratio is too high for the mesocosm streams and certainly too many fish to sustain in a self-sustaining multitrophic mesocosm test over 21 days of exposure. Thus, the use of pair or trio (1M:2F) breeding is recommended. If breeding trios with Fathead Minnow are used in a modular mesocosm design, data from both females are assessed with a measure of central tendency keeping the level of replication at the stream level.

It is important to consider the objectives of the fish study when selecting which mesocosm system to use. While the trailer system provides for greater numbers of fish in each replicate, exposures are water-borne only and effect endpoints measured are condition, survival, organ size in adults and growth in juveniles, the latter as a substitute for size-at-age. The modular mesocosms have lower fish numbers per replicate but allow for multitrophic and/or water-borne exposures, and allow for investigations over a partial life cycle from breeding to offspring production over time in a repeat spawner such as Fathead Minnow. Reproductive variables, such as cumulative numbers of eggs produced per female per day, can be replicated over time, and distributions in egg production can be assessed for each treatment using distribution-based statistical tests such as the Kolmogorov-Smirnov test.

9.2.5.4.2 Available Pre-existing (Historical) Data

If data are available, the sample sizes to measure a certain effect size in a parameter with a targeted level of statistical power can be calculated, because sample variability is known. Larger numbers of samples are needed if parameters are highly variable, if detection of small differences between reference and exposure streams (small effect size) is desired, or if a high level of power is required. To determine the sample sizes using pre-existing data, the effect size and the statistical power level that is acceptable for the decision-making process needs to be determined. The purpose of defining an effect size and power level is to determine when the sampling program is collecting adequate information to provide decision support. Sample sizes can be calculated using methods described in Chapters 3 and 4. Appendix 1 of Environment

Canada (1997) provides a detailed discussion of power relationships, effect sizes and the benefits of reducing variability in terms of increasing power. It is recommended that α and β be set equally at 0.10 for power calculations.

9.2.6 Fish Monitoring – Effects Assessment

9.2.6.1 Study Design

9.2.6.1.1 Species Selection

The most important factors when selecting fish species for mesocosm studies are environmental relevance, abundance for collection, size, availability of adults and juveniles, sexual dimorphism, spawning period, and sensitivity to effluent. The recommended method for assessing effluent effects in mesocosm studies conducted in the trailer mesocosm system is by monitoring pre-spawning adults (sexually mature fish) and juveniles of one small-bodied fish species (e.g., darters, minnows, sculpins) that is relevant to the receiving environment. Monitoring of adults provides for assessment of effluent effects on survival, reproduction (energy use) and condition (energy storage). Monitoring of juveniles provides for assessment of effluent effects on survival, energy use (i.e., growth as a size-at-age substitute) and energy storage (i.e., liver size and condition).

Species selection for mesocosm studies is restricted by size requirements (numbers of fish per tank decrease with increasing fish size) and sexual dimorphism (if males and females cannot be externally sexed, increased numbers should be considered to ensure adequate sample sizes). As such, mesocosms are best suited for use with small-bodied fish species. Sampled species should also be suitable for measuring the recommended variables. To date, fish species used in mesocosm studies include Mummichog, juvenile Atlantic Salmon, Slimy Sculpin, Creek Chub, Longnose Dace and Fathead Minnow (Table 9-1).

The advantages and disadvantages of using small-bodied species in field surveys have been described by Gibbons et al. (1998a, 1998b). A small-bodied fish can be considered as a fish species that has a maximum length of 150 mm. Information on maximum growth of fish species can be found in the scientific literature, including Scott (1967), Scott and Crossman (1973), Fritz et al. (1975), Roberts (1988), Nelson and Paetz (1992), Jenkins and Burkhead (1993), Coad (1995), and Leblanc and Couillard (1995). Small-bodied fish species are usually more abundant, easier to capture, and more sedentary than larger-bodied fish species. Smaller home ranges are desirable, as it increases probability and consistency of effluent exposure compared to larger, more mobile, possibly migratory species.

There can be disadvantages to using small-bodied fish. Often, less is known about their basic biology, particularly their spawning habits, making it difficult to determine the best sample areas, times and methods. However, due to the significant amounts of data collected on small-bodied fish as part of EEM programs as well as many Canadian research studies, knowledge of small-bodied fish life-history strategies and basic biology has increased significantly (Munkittrick, University of New Brunswick, unpublished). Some species are multiple spawners (i.e., they produce more than one clutch of mature ova every year; see Heins and Rabito 1986;

Burt et al. 1988; Paine 1990). This can be a disadvantage for field surveys because reproductive effort in these species is difficult to estimate from a single sample. Reproductive tissue can be turned over almost completely between clutches (i.e., most of the mass of ova in the ovary will be spawned and then a new clutch of mature ova will be developed). The number of clutches produced during the spawning season becomes the important reproductive variable and is difficult to estimate for an individual female in the field, even with frequent sampling. However, multiple spawners such as the Mummichog have been used successfully in mesocosm studies to evaluate effluent effects on EEM effect endpoints, including changes in gonad size (Dubé and MacLatchy 2000a; Cash et al. 2003; Dubé et al. 2001). An advantage of using controlled-exposure mesocosm studies is that the state of fractional spawners can be monitored throughout the exposure period.

The latest advances in mesocosm technology and application using the modular system has evolved around the use of Fathead Minnow in a 21- to 30-day partial-life-cycle exposure experiment (Table 9-1). Mesocosm applications using the trailer system have been successful in assessing effluent effects. However, exposures are water-borne, which lessens environmental relevance. Water-borne exposure is standard practice in fish toxicological assessments with single contaminants, but given the nature of complex mixtures, dietary pathways of exposure should be considered. For example, numerous studies have investigated the importance of the trophic-transfer of metals (Ni et al. 2000; Chen et al. 2000; Mason et al. 2000; Xu and Wang 2002) in aquatic environments. Including dietary exposure in the mesocosm approach would therefore be an improvement. The ability to directly assess reproductive output (number of eggs produced, number of spawning events and offspring survival, hatching success, deformities) after exposure to effluent, in addition to the standard EEM effect endpoints, would also be desirable for more causal investigations.

Fathead Minnows have been used extensively as a toxicological workhorse in laboratory investigations, as they represent an ecologically significant part of the Cyprinidae family, they have been extensively tested, and a large database of knowledge exists regarding their culture and life cycles (Panter et al. 2002; Ankley et al. 2001; Jensen et al. 2001). They are also used in risk assessment and government/industry monitoring studies on an international scale (US EPA 1982, 1996, 1999, 2002; OECD 2001; Shaw et al. 1995a, 1995b). In addition, Fathead Minnows are small (average length of 6 cm and width of 1 cm), fractional, substrate spawners that, under specific conditions, can easily be manipulated in captivity to produce clutches of 50-150 eggs every 3-5 days. They are also an environmentally relevant species, as they are abundant in freshwater systems across Canada. Thus, the development of a mesocosm approach using Fathead Minnows to measure EEM effect endpoints and provide focus for direct evaluation of how reproductive output is affected by effluent mixtures would be highly useful in cases where field surveys cannot be conducted. Furthermore, self-sustaining (no external food source) mesocosms wherein fish and their diet are co-cultured, resulting in fish exposure through both dietary and water-borne pathways (i.e., a multitrophic mesocosm), would also be highly relevant.

Ankley et al. (2001) developed a short-term bioassay using Fathead Minnow that assesses reproduction as well as aspects of early development in a shorter time frame than traditional life-cycle assays. The time frame of the partial-life-cycle test (21 days) made it a suitable candidate for mesocosm use. A number of investigations have used the 21-day bioassay to

monitor effects of estrogenic (Harries et al. 2000; Ankley et al. 2001; Sohoni et al. 2001), androgenic (Ankley et al. 2003) and anti-androgenic (Jensen et al. 2004) compounds. These studies have predominantly focused on single contaminants that do not represent the complexity of effluents. A limited number of investigations into the effects of industrial effluents using the 21-day bioassay have been conducted (Martel et al. 2003; Parrott 2005). However, these studies have used differences in methodology (e.g., number of independent replicates, number and type of variables measured) and were conducted in the laboratory under water-borne exposure conditions.

Development of the Fathead Minnow 21-day bioassay for use in EEM mesocosm studies first required in situ use with natural receiving water, yet controlled temperatures and photoperiods (Rickwood et al. 2006a, 2006b). As the chironomid *C. tentans* life-cycle bioassay had been previously established for use in the modular mesocosm system (Hruska and Dubé 2004, 2005), the next step was to combine these tests into a multitrophic mesocosm test for use in the modular mesocosm system. This was first done in the lab with controlled water temperature (Rickwood et al. 2006c) and then moved to the field, with full testing under ambient temperatures and photoperiods and using ambient reference water (Pollock et al. 2009; Rickwood et al. 2008). The Fathead Minnows used in these studies have been laboratory cultured and transported to the site to control for fish quality, age, state of reproductive development and exposure (or lack thereof).

9.2.6.1.2 Effect Indicators

Trailer Mesocosm System: The effect indicators measured in the trailer mesocosm study using fish are the same as those measured in a field survey. Effects on fish growth, reproduction, condition and survival are determined in answer to the question, “Has the fish population been modified by effluent?” Survival, growth and reproduction (energy use), and condition (energy storage) are measured to detect any effluent-related effect on the fish population. All of these measurements can be taken from fish exposed to effluents in mesocosm studies (Table 9-2).

Growth is the change in size (weight or length) with time or age. Mesocosm studies over a longer exposure period (45-60 days) can assess effluent effects on juvenile growth, specifically changes in total body weight and length relative to control fish. In mesocosms, growth is measured in YOY (young of the year) or juvenile fish from the start of the study to the end, a duration of 6-8 weeks depending upon the temperature during the study. The determination of growth effects over a short exposure period requires that studies be conducted during warm water temperatures and with a sentinel species that has a high growth rate. Kneib and Stiven (1978) have shown that juvenile growth rates of Mummichog, for example, are very high post-hatching (an increase of 15 mm in 2 months), which makes them ideal candidates for assessing the effects of effluent on growth. Other studies using YOY Slimy Sculpin have shown that growth effects can be measured after only 26 days of effluent exposure (Dubé et al. 2005). Although length and weight measurements are taken in adults, growth is unlikely during the short duration of the mesocosm studies.

Table 9-2: Fish mesocosm study effect indicators and endpoints and related statistical procedures

Effect Indicators	Effect and Supporting Endpoints	Statistical Procedure
Growth	• *Change in size (weight and length) with time (end measurements compared to starting measurements) • Juvenile growth (change in length over time) • Juvenile growth (change in whole animal wet weight over time) • Adult body weight (whole) • Adult length	ANOVA
Reproduction	• *Relative gonad size (gonad weight against body weight [adults]) • Relative gonad size (gonad weight against length [adults]) • Fecundity (number of eggs/female against body weight, length and/or age); n/a for juvenile fish and males • Relative egg size (mean egg weight against body weight)	ANCOVA
Condition	Juvenile and Adult Condition • *Body weight relative to length (k)^x • *Relative liver size (liver weight against body weight) • Liver weight against length	ANCOVA
Survival	Juvenile and Adult Survival • *Percentage surviving at the end of the exposure period	ANOVA

© M. Dubé

* Mesocosm effect endpoints used for determining effects as designated by statistically significant differences between exposure and reference streams. Other supporting endpoints can be used to support analyses.

Table 9-3: Recommended response variables and suitable additional supporting information, and suggested statistical analysis for Fathead Minnow application in modular mesocosm systems

Type of Response	Response Variable	Dependent Variable (Y)	Independent Variable (X)	Covariate (X)	Statistics (Single Factor)	Statistics (Two Factors)
Stage of experiment: <u>Pre-exposure (for fish that met criteria)</u>						
Energy Use (Adults)	Mean total egg production (number)	Mean total number of eggs per breeding group	Treatment	n/a	ANOVA (one-way)	ANOVA (two-way)
	Mean egg production (number)	Mean number of eggs per female per day	Treatment	n/a	ANOVA (one-way)	ANOVA (two-way)
	Cumulative eggs per female per day (number)	Cumulative eggs per female	Day	n/a	Kolmogorov-Smirnov	Kolmogorov-Smirnov
	Mean total spawning events (number)	Mean total number of spawning events per breeding group	Treatment	n/a	Chi-Square	Chi-Square
Offspring	Fertilization success (%)	Number of eggs fertilized/ number of eggs laid x 100	Treatment	n/a	ANOVA (one-way)	ANOVA (two-way)
Stage of experiment: <u>Exposure</u>						
Adult Survival	Mean adult survival (%)	Number of adults surviving at end/Number of adults at start x 100	Treatment	n/a	ANOVA (one-way)	ANOVA (two-way)
Energy Storage (Adults)	Condition (g/cm)	Total body weight (log)	n/a	Length (log)	ANCOVA	ANCOVA (two-way)
	Relative liver size (g)	Liver weight (log)	n/a	Total body weight (log)	ANCOVA	ANCOVA (two-way)
		Liver weight (log)	n/a	Length (log)	ANCOVA	ANCOVA (two-way)
	Relative egg size (µm)	Mean egg size (log)	n/a	Total body weight (log)	ANCOVA	ANCOVA (two-way)
		Mean egg size (log)	n/a	Length (log)	ANCOVA	ANCOVA (two-way)
Energy Storage (Adults) (If ANCOVAs cannot be done due to lack of spread on x)	Mean condition factor (%)	Total body weight/ (length) ³ * 100	Treatment	n/a	ANOVA (one-way)	ANOVA (two-way)
	LSI (%)	Liver weight/body weight * 100	Treatment	n/a	ANOVA (one-way)	ANOVA (two-way)

Type of Response	Response Variable	Dependent Variable (Y)	Independent Variable (X)	Covariate (X)	Statistics (Single Factor)	Statistics (Two Factors)
axis)	GSI (%)	Gonad weight/body weight * 100	Treatment	n/a	ANOVA (one-way)	ANOVA (two-way)
Energy Use (Adults)	Total body weight (g)	n/a	Treatment	n/a	ANOVA (one-way)	ANOVA (two-way)
	Fork length (cm)	n/a	Treatment	n/a	ANOVA (one-way)	ANOVA (two-way)
	Relative gonad size (g)	Gonad weight (log)	n/a	Total body weight (log)	ANCOVA	ANCOVA (two-way)
		Gonad weight (log)	n/a	Length (log)	ANCOVA	ANCOVA (two-way)
	Cumulative eggs/breeding group/day (number)	Cumulative eggs/breeding group	Day	n/a	Kolmogorov-Smirnov	Kolmogorov-Smirnov
	Mean total egg production/day (number)	Mean total number of eggs/female/day	Treatment	n/a	ANOVA (one-way)	ANOVA (two-way)
	Mean egg production/day (number)	Mean number of eggs/female/day	Treatment	n/a	ANOVA (one-way)	ANOVA (two-way)
	Cumulative spawning events/breeding group/day (number)	Cumulative total number of spawning events/breeding group	Day	n/a	Kolmogorov-Smirnov	Kolmogorov-Smirnov
	Mean total spawning events/day (number)	Mean total number of spawning events/female/day	Treatment	n/a	Chi-Square	Chi-Square
	Relative fecundity (number)	Number of eggs/female (log)	n/a	Total body weight (log)	ANCOVA	ANCOVA
		Number of eggs/female (log)	n/a	Length (log)	ANCOVA	ANCOVA
Other Adult Reproductive Responses	Development of secondary sexual characteristics	Presence/absence of tubercles; banding; fin dot; fat pad	Treatment	n/a	Chi-Square	Chi-Square
		Rank of ovipositor size	Treatment	n/a	Chi-Square	Chi-Square
	Reproductive hormones	Male/female testosterone (ng/g)	Treatment	n/a	ANOVA (one-way)	ANOVA (two-way)
		Female estrogen (ng/g)	Treatment	n/a	ANOVA (one-way)	ANOVA (two-way)

Type of Response	Response Variable	Dependent Variable (Y)	Independent Variable (X)	Covariate (X)	Statistics (Single Factor)	Statistics (Two Factors)
		Male vitellogenin (ng/g)	Treatment	n/a	ANOVA (one-way)	ANOVA (two-way)
	Gonadal histology	Number and stage of cells	Treatment	n/a	Chi-Square	Chi-Square
Offspring	Mean hatching success (%)	Number of eggs hatched/ number of eggs fertilized x 100	Treatment	n/a	ANOVA (one-way)	ANOVA (two-way)
	Mean total deformities (%)	Total number of deformities at hatch/total number of larvae x 100	Treatment	n/a	ANOVA (one-way)	ANOVA (two-way)
	Mean fertilization success (%)	Number of eggs fertilized/ number of eggs laid x 100	Treatment	n/a	ANOVA (one-way)	ANOVA (two-way)
	Mean larval survival (%)	Number of larvae survived after 5 days/total number of larvae x 100	Treatment	n/a	ANOVA (one-way)	ANOVA (two-way)
	Mean days to hatch	Mean number of days to hatch	Treatment	n/a	ANOVA (one-way)	ANOVA (two-way)
Water Chemistry	In situ measurements	Various	Treatment	n/a	ANOVA (one-way)	ANOVA (two-way)
	Lab analysis	Various	Treatment	n/a	ANOVA (one-way)	ANOVA (two-way)
Fish Tissue Body Burden	Lab analysis	Various	Treatment	n/a	ANOVA (one-way)	ANOVA (two-way)
Fish Tissue - Gonads	Lab analysis	Various	Treatment	n/a	ANOVA (one-way)	ANOVA (two-way)
Fish Tissue - Gills	Lab analysis	Various	Treatment	n/a	ANOVA (one-way)	ANOVA (two-way)
Fish Tissue - Liver	Lab analysis	Various	Treatment	n/a	ANOVA (one-way)	ANOVA (two-way)
Chironomid Tissue	Lab analysis	Various	Treatment	n/a	ANOVA (one-way)	ANOVA (two-way)
Algal Tissue	Lab analysis	Various	Treatment	n/a	ANOVA (one-way)	ANOVA (two-way)
Sediment Analysis	Lab Analysis	Various	Treatment	n/a	ANOVA (one-way)	ANOVA (two-way)

© M. Dubé

Legend: LSI: Liver Somatic Index; GSI: Gonadosomatic Index.

Reproduction is expressed as reproductive effort, fecundity, egg size or gonad weight relative to body size. To date, mesocosm studies have examined changes in gonad size to assess effluent effects on reproductive function. However, measures of fecundity and egg size are easy to measure, if an appropriate sampling time is chosen. Ideally, exposure studies should commence 6-8 weeks prior to the spawning season in order to assess effluent effects on gonad weight.

During the mesocosm study, the physical state of the fish is also assessed. A visual estimation of physical malformations and lesions on the body surface, including eroded, frayed or hemorrhagic fins, parasites, or other physical deformations, is required.

Modular Mesocosm System: The response variables measured in the modular system, which has been used primarily with Fathead Minnow, are summarized along with suggested statistical analysis procedures in Table 9-3. The variables shaded in grey are those recommended for use. Those variables that are not highlighted are variables that have been measured in different study designs and could be included to provide additional support. It is important to note that use of the standardized Ankley et al. (2001) 21-day reproductive assay has been modified for improved application under the EEM program and specifically for use in mesocosms as an alternative to the fish survey. The methods are described below and remain to be validated and standardized by the EEM Science Committee. Mines proposing this alternative approach should expect to evaluate the response variables to be measured and should include additional supporting variables to adapt and improve this method for use in their EEM program.

9.2.6.1.3 Timing and Duration of Effluent Exposure

The timing of the mesocosm study is important and is dependent upon the test species. Water temperatures affect the experimental design. If a study is being conducted in the spring when the water is cold, exposure times may be longer, especially if a response such as growth is being measured. Temperatures below -5°C are prohibitive to mesocosm operation due to freezing of water lines.

Mesocosm studies should strive to balance duration with cost-efficiency. Exposures of 21-45 days are common for invertebrate community studies in either mesocosm system. Exposures of 21 days are common if the Fathead Minnow partial life cycle is being used as the sentinel species in the modular mesocosm system. If the trailer system is being used for a water-borne exposure, 30 days is common during summer and fall months, and has been used to measure changes in adult organ size (liver size, gonad size) and growth rates of YOY Slimy Sculpin as a result of effluent exposure.

All mesocosm studies should be conducted during normal industrial operations. Effluents should be representative of normal operating conditions. Mesocosm studies should not be conducted when effluent has not been discharged for long periods or during wastewater treatment upsets. Ensure any planned shutdowns are identified well in advance and studies are not affected.

Trailer Mesocosm System: Timing of the mesocosm study is important and dependent upon the spawning cycle of the fish species. See Chapter 3 for additional information on spawning cycles

and seasonal sampling. Ideally, for spring spawners, studies will be conducted 6-8 weeks before spawning commences. For early spring spawners where it is impossible to study for this length of time prior to the spawning season, studies should be conducted as late in the year as possible to allow for gonadal senescence and recrudescence. For fall spawners, a spring or summer survey is appropriate. However, this may not apply to fish in which ova mature rapidly. For example, some late-spring-spawning minnows should be studied in early spring, rather than in fall, when ova may still be immature. Water temperatures are also a consideration for timing of studies where juvenile growth rates are higher during the summer months.

Modular Mesocosm System: The modular system, when used with Fathead Minnow, is on-site for a 10-14 day pre-exposure (length depends upon when fish meet appropriate baseline selection criteria) and a 21- to 30-day exposure period. Studies should be conducted from May to late September when water and air temperatures are between 15°C and 28°C.

9.2.6.2 Fish Methods

9.2.6.2.1 General Set-up – Trailer Mesocosm System

Once the trailer mesocosm system has been transported to the study site and reference water and power have been connected, substrate can be added to the streams and water flow rates can be calibrated. Depending upon the site-specific substrate characteristics, washed and crushed gravel, or sand, can be placed into each tank to a depth of 5 cm. In previous studies, large washed rocks were also added to each tank to serve as refuge for the fish.

Fish are collected from a reference area that is not exposed to the effluent being studied. Obviously, collection requires non-lethal sampling techniques including minnow traps, trap nets or electro-fishing with barrier nets (Portt et al. 2006). Fish are usually sorted in the field to ensure adequate numbers of juveniles and of adults of each sex (if sexually dimorphic), and to collect fish of similar size-classes. If the species chosen is not sexually dimorphic, an assumption is made that sex ratios in the field are equal and the fish should be randomly allocated to the tanks. Fish are transported back to the study site in containers with covers and under adequate aeration. If reference areas are not available, hatchery-reared fish can also be used if these fish are relevant to the study area.

At the study site, and preferably within the same day, fish are measured for length and weight, and randomly assigned to each stream. The precision of these measurements is as described in Chapter 3, with increased precision when using small-bodied species. The optimal numbers of fish per stream are 20 adults of each sex and 20-30 juveniles. To ensure random allocation of fish to streams, 16 aerated buckets with reference water are set up. Fish are measured and placed, one by one, until each bucket contains 1 fish. This procedure is repeated until all buckets have 5 fish each. Buckets are then randomly allocated to the streams and the procedure is repeated until target sample sizes are attained in the streams. It is essential that this procedure be followed to ensure that the largest fish, for example, are not all allocated to the first few streams.

There are 2 possible fish allocation scenarios. Both adults and juveniles can be allocated to the same tanks if juvenile growth rates will be based solely on starting and ending length and weight

measurements, and adults are not cannibalistic. If juveniles will be sampled during the course of the study for growth (perhaps juveniles are tagged and repeated measurements are conducted), allocating adults and juveniles to different tanks (or different areas within a tank if these areas can be physically separated) would be recommended to minimize adult capture stress. Mesocosm streams are randomly assigned to each study treatment (control, effluent) and fish are randomly allocated to each stream. This ensures a randomized distribution to minimize stream differences due to factors such as stream position on the trailer.

After inoculation, fish are acclimated to artificial stream conditions for a minimum of 72 hours prior to effluent addition or until fish are feeding. Streams should be covered with netting to minimize fish loss due to escape or predation by birds. During the exposure period, which extends from 30 to 60 days, fish should be fed trout pellets at a rate of 4-6% total body weight per tank per day. The size of the pellets is dependent upon the size of the species used. In previous studies, trout pellets were crushed prior to feeding.

9.2.6.2.2 General Set-up – Modular Mesocosm System

In the modular mesocosm system, each system consists of a table (one per treatment) holding 5-8 replicate (preferred), 10.3-L, circular, high-density polyethylene streams. The replicate streams sit on the table, which drains into an 85-L dilution reservoir as previously described. If the trophic-transfer system is being used, each stream consists of a sediment (pre-cleaned silica sand) culture of the chironomid *C. tentans*, a feeding barrier, spawning tile and one breeding group of Fathead Minnow (Figure 9-6) (Rickwood et al. 2006c, 2008). The feeding barriers are used to control Fathead Minnow access to the *C. tentans* that have established in the substrate during the pre-exposure portion of the study. Each circular barrier contains a 1/8" (0.32 cm) square mesh screen with a pie-shaped opening of 1/10th of the stream area (approximately 71 cm²) that is turned every second day to dispense the appropriate amount of food (1 g *C. tentans*/day) (see below on culturing appropriate densities of *C. tentans*).

Pre-exposure Design: Six-month-old, naive Fathead Minnow are obtained from a reputable culture laboratory. Fish are typically transported to site by ground or air within 24 hours in oxygenated bags with water inside coolers. Upon arrival, containers are aerated and allowed sufficient time for acclimation to ambient temperatures. Reference water is slowly added to acclimate. Once acclimatized, fish should be placed into an appropriate holding tank to become reproductively stimulated prior to the pre-exposure breeding trial. This stimulation period should last for approximately 3 days.

The Fathead Minnow modular mesocosm method requires a pre-exposure and exposure trial. The pre-breeding trial normally consists of double the number of required breeding groups, which are bred in independent replicates in the absence of effluent to establish baseline reproductive performance. At the beginning of the breeding trial, total body weight (g), fork length (mm) and secondary sex characteristics are recorded. Secondary sex characteristics include banding, nuptial tubercles, dorsal pad and fin dot in males and ovipositor size in females (Parrot and Wood 2001). Female Fathead Minnow are size-matched, if possible, to within $\pm 25\%$ of the male body length (Pollock et al. 2008). Each breeding pair is fed 0.5 g of frozen bloodworms twice daily throughout the pre-exposure period. Each day prior to feeding and recording of water quality, breeding tiles in each stream are checked for egg deposition. If

breeding has occurred, eggs will be gently rolled off of the spawning tile into petri dishes with reference water and photographed for digital counting. At the end of the breeding trial, breeding groups will be selected for the exposure phase of the experiment. Breeding groups (pairs or trios, depending upon the study) are selected on the basis that there is 100% survival of all adults, that eggs are present in each replicate at least once in the immediately preceding 7 days, and that > 80% fertilization of eggs has occurred (OECD 2006; US EPA 2007). Breeding groups at both extremes (superior breeders and breeders with very few eggs) should be excluded from the study. The selected groups should be distributed throughout the mesocosm treatments and streams so as to minimize variance between the treatments. Statistical analyses are performed prior to final selection to ensure that there are no significant differences among and between treatments before effluent exposure commences. The unit of replication is $n = 5-8$ per treatment.

Trophic-transfer system: The trophic-transfer system and associated sediment cultures of the chironomid *C. tentans* are set up during the pre-exposure period. Target invertebrate densities in each stream are based on an optimal daily feeding amount of 1 g/breeding pair/day (Rickwood et al. 2006a, 2006b). Seven-day-old *C. tentans* larvae are shipped to the site once per week for 3 weeks to establish the cultures in the mesocosm streams before adding the fish. The larvae are obtained from a reputable culture supplier. *C. tentans* will have been exposed to the effluents for at least 7 days prior to the introduction of the fish, to ensure dietary exposure. In addition, this ensures that *C. tentans* will be at various life stages (egg, larvae, pupae, adult) within the streams to maintain a healthy breeding cycle. The number of invertebrates required to sustain the fish over 21 days in each stream is calculated based on the average number of *C. tentans* that emerge from one egg sac (~ 300) and by determining the number of third and fourth instars with a combined weight of 1 g (~ 50 3rd and 4th instars). Once it is known how many *C. tentans* weigh 1 g, the total number of *C. tentans* needed for the entire 21-day exposure period will be calculated ($50 \text{ } C. \text{ tentans/g} \times 21 \text{ days} = 1050 \text{ } C. \text{ tentans}$ or 350 7-day-old larvae/stream/week for 3 weeks).

Once the larvae have arrived at the site, they are acclimated using treatment water in their individual containers, by adding 25% of the treatment water to each container 4 times in 12 hours. The larvae are acclimated to the ambient temperature by slowly lowering the water temperature in a water bath no faster than 1°C every 1-2 hours. Once the *C. tentans* are acclimated to the ambient temperature and effluent water, they are distributed among the artificial streams and fed Tetramin™ slurry (100 g Tetramin™ flakes to 1000 ml reference water) at a rate of 10 ml in the first week, 20 ml in the second, and 30 ml in the third and subsequent weeks, 3 times/week. It is recommended that 3 sediment cores (core sampler area approximately 9 cm²) be taken from each stream at the end of the pre-exposure period before fish placement so that invertebrate densities can be calculated.

Exposure Design: Once Fathead Minnow are allocated to the treatment streams, treatment solutions are delivered to the mesocosms as previously described, and the exposure period commences.

9.2.6.2.3 Maintenance and Monitoring

Daily maintenance for the duration of the study includes calibration of flows to each reservoir to ensure target effluent dilutions are attained, cleaning of screens, and feeding of fish. Screens are placed over the water outlet pipes on each stream to prevent fish loss. These screens may require daily cleaning.

Daily monitoring requirements for each stream include recording of any fish mortality and monitoring of physical water variables, including temperature, dissolved oxygen, conductivity and pH. These measurements can be taken using YSI Inc. instruments or continuous recording equipment such as hydrolabs or thermisters. Dissolved oxygen is of critical importance, especially in studies using effluents high in organic content (e.g., pulp mill effluents). Dissolved oxygen levels should be maintained above 60% (minimum), and preferably above 80%, in all streams, using aeration if required.

For the modular mesocosm system, daily observations and measurements can include egg production, hatching success and larval survival (Table 9-3). Breeding tiles are checked daily at the same time and before feeding or water quality measurements are taken. Eggs are removed from the tiles and photographed. A consistent sub-sample of eggs from each productive brood can be collected for future analysis and total egg production corrected for this removal. Remaining eggs are then rolled into an egg cup and placed in the appropriate aerated culture tubs with treatment water. Twenty-four hours after spawning, the eggs are photographed again to check fertilization success. They are then left undisturbed until all eggs are either hatched or dead (~ 4-5 days). Once larvae have hatched, they are preserved in 10% formalin for latter enumeration and examination for deformities where the latter is a desired parameter for supporting assessments.

9.2.6.2.4 Sampling and Analysis of Fish

At the end of the exposure period, fish are anaesthetized and fork length (mm), whole body weight (g) and secondary sexual characteristics are recorded. Fish are then euthanized by spinal severance, and gonads, liver and eviscerated (carcass) weights are recorded. All streams should be sampled on the same day and all fish from each stream should be sampled before progressing to the next stream.

The measurements required and level of precision are the same as those outlined in Chapter 3. Sample preparation, laboratory analysis and QA/QC procedures for mesocosm studies are the same as those required for field fish surveys.

9.2.6.3 Data Assessment and Interpretation

9.2.6.3.1 General Requirements

Similar to field studies, in mesocosm studies, data assessment and interpretation follow each monitoring or assessment phase. In data assessment and interpretation, the following questions are answered:

- Is there an effect?
- Is the effect mine-related?
- Is the magnitude and extent of the effect known?
- Is the mine-related cause of the effect known?

An overview of data analysis and interpretation for mesocosm studies is presented here.

9.2.6.3.2 Statistical Analysis of Parameters

To determine whether there is an effect on fish exposed to effluents in mesocosm streams, statistical analyses of the data are conducted of the biological variables, as suggested in Table 9-2 for the trailer mesocosm system and Table 9-3 for the modular mesocosm system with Fathead Minnow. Table 9-3 lists typical analytical approaches for two types of experimental designs: single-factor ANOVA (e.g., effect of effluent) or two-factor ANOVA (e.g., effect of effluent and contaminated sediment).

Sex differences in growth rate, body weight, condition factor, gonad size and liver size are common due to differences in overall energetic requirements between male and female fish. Therefore, for all parameters, sexes should be treated separately when estimating variability. Immature or juvenile fish should also be treated separately. The analyses that are suggested in tables 9-2 and 9-3 are preferred. However, other analyses for specific parameters can be conducted depending upon the variability of the data set. For example, in mesocosm studies, individuals of a similar size-class are selected for placement in the streams, resulting in reduced variability in parameters compared to those measured in field surveys. Thus the range over which regressions such as ANCOVAs are conducted for mesocosm data is narrow enough that ANOVAs on original organ weights (liver size, gonad size) and ratio metrics (LSI, GSI) or condition factors are justified in this instance.

Statistical considerations specific to mesocosm data sets are stated below. Additional reference can be found in the literature cited in Table 9-1.

9.2.6.3.3 Nested ANOVA Analyses

For any variable measured once on a whole replicate stream, the statistical design is a simple t-test, comparing exposed and control treatments (control/impact design). The mesocosms or streams are replicates for the treatments. However, all biological and biochemical variables are measured on replicates at a lower level—either individual fish or composites of several fish within streams. A nested ANOVA can be used to analyze these variables. If the variance among replicate streams within a treatment (Error I) is large relative to the variance among fish or composite samples within streams (Error II), then the nested ANOVA is effectively a t-test comparing treatments, with the stream means as replicate observations. However, if Error I is not large relative to Error II (e.g., $p > 0.25$), then the two error terms can be pooled to increase the power of the test comparing treatments (i.e., individual fish or composite samples rather than streams are used as replicates).

9.2.6.3.4 ANCOVA Analyses

Most parameters are normally estimated using ANCOVA, as discussed in Chapter 8. This is often unnecessary in mesocosm studies, as mentioned above, unless limited fish availability prevented the standardization of size and age.

9.2.6.3.5 Data QA/QC and Analysis

The importance of ensuring data quality cannot be overstated. There are basic requirements for study design, consistency of methods and measurements, and definition of protocols and procedures; these are outlined in detail in Chapter 8.

9.2.7 Benthic Invertebrate Community Monitoring - Effects Assessment

9.2.7.1 Study Design

9.2.7.1.1 Species Selection

The benthic community that is established in the mesocosm streams is representative of that found in the reference field areas. Benthic samples are collected from the river and the entire community is inoculated into each mesocosm stream. Culp et al. (1996, 2001), Podemski (1999) and Culp et al. (2000a) have conducted studies that compared the community structure of benthic invertebrates in the mesocosm streams to that of field communities at reference areas. No ecologically significant differences in structure were observed, illustrating the effectiveness of the inoculation procedures and the suitability of mesocosms for testing effluent effects on environmentally relevant communities of benthic invertebrates.

9.2.7.1.2 Effect Endpoints

All the effect endpoints (total benthic invertebrate density, taxon richness, Simpson's Evenness Index, similarity index [Bray-Curtis]) used to assess effluent effects on benthic invertebrates in EEM field surveys can be measured in mesocosm studies. Diversity, taxon density, proportion or presence/absence are recommended to allow for the interpretation of effects (Chapter 4). These effect endpoints are summary metrics selected to encompass the range of responses that may result from effluent, including changes in productivity, species composition and biodiversity. Many other benthic invertebrate descriptive metrics are available in the literature (for a review see Resh and McElravy 1993) and may be used, if applicable, on a site-specific basis to aid in the interpretation of effects determined with the effect endpoints listed above.

9.2.7.1.3 Timing and Duration of Effluent Exposure

Mesocosm studies using benthic invertebrates can be conducted at any time from early spring to late fall. The primary limiting factor is temperature, with air temperatures below -5°C prohibitive to mesocosm operation due to freezing of water lines. Studies should be conducted during periods when field communities are under maximal effluent exposure for improved environmental realism. Studies should also be conducted at the time of year when the benthic

invertebrate diversity is highest and water temperatures are conducive to growth. If historical data exist, it would be useful to examine the data and, if appropriate, conduct the study during similar periods so that the studies can be compared. Subsequent monitoring should also be conducted during similar periods of the year to be comparable.

The duration of the mesocosm studies for benthic invertebrate programs is typically 30-45 days, including a 7- to 12-day inoculation period for primary producers and a 25- to 30-day effluent exposure period. Effluents are collected daily or every second day during the studies and should be representative of normal industrial operating conditions.

9.2.7.2 Benthic Invertebrate Methods

9.2.7.2.1 Mesocosm Trailer – General Set-up

General set-up requirements specified in sections 9.2.3.2 and 9.2.3.3 also apply to benthic invertebrate studies. The recommended total flow of reference water or of dilution water and effluent into each stream is 2 L/minute. This flow rate results in a 2-hour volume turnover time for each stream (stream volume 227 L). Stream volume turnover rates of 4 hours have been used to reduce effluent requirements. However, longer turnover rates are not recommended due to the effects on stream temperature and dissolved oxygen. For invertebrate applications, current velocity in each stream can be created by a belt-driven propeller system. Current in each stream is established based on site-specific conditions but is normally set between 10 and 20 cm/second.

Prior to placement of substrate into the streams, five sampling bags (0.1 m², 500-µm mesh) are installed on the stream bottom. These bags are lifted at the end of the experiment, resulting in 5 sub-samples per stream. A standardized benthic environment is created in each stream to simulate the dominant environmentally relevant habitat type. To date, mesocosm studies using benthic invertebrates have focused on riverine habitats where riffle substrate is dominant. The bottom of each stream and the sampling bags are covered with an 8-cm layer of washed gravel (stones of 1-2 cm in diameter). The gravel is then left to colonize for a 7- to 12-day period to allow sufficient algal and microbial growth. Only water delivery to the mesocosm streams occurs (2 L/minute) during this colonization period. The duration of the colonization period is site-specific and depends upon temperature and colonization of algae from the river into the system.

Following the algal colonization period, existing benthic communities are transplanted to the stream mesocosms. The transplantation protocol is determined on a site-specific basis. The following is an example of inoculation protocol for a riverine, riffle habitat with a large cobble/gravel-dominated substratum (Podemski 1999; Podemski and Culp 1996; Culp and Cash 1995; Culp et al. 1996) (Table 9-1):

1. Large cobbles (surface area ~ 535 cm²) are randomly selected from the river at the reference area for placement into the mesocosm streams with their associated periphyton and invertebrate biota. These cobbles provide additional substrate and also stock the streams with a natural community of periphyton and benthic invertebrates.
2. During collection, the stones are enclosed with a 0.1-m² U-net (500-µm mesh) (Scrimgeour

et al. 1993), carefully lifted from the stream bed, and placed into a container with river water. In addition, the gravel substratum beneath the stones is gently disturbed (to a depth of 5 cm) to collect any invertebrates under and around the base of the stone. The container is carefully transported to the mesocosm so as not to dislodge the periphyton and invertebrates associated with the stone. The cobbles are randomly placed in the artificial streams, oriented to the water flow in a manner similar to their original orientation in the natural environment. This process continues until the appropriate density (e.g., 10 large cobbles/artificial stream) of large cobbles is reached, based on natural substratum composition. Other samplers may be used to inoculate the streams with benthic communities, as outlined in Chapter 4. The objective, however, is to establish ambient densities in the mesocosm streams. To determine the number of river samples to place in the streams, the mesocosm stream area (0.9 m²) is divided by the area of the selected sampler. This random allocation of sub-samples from pooled invertebrate-community samples ensures that the initial invertebrate composition is similar among streams and limits the amount of variability in community composition among mesocosms that can be attributed to the pattern of species introduction (Wrona et al. 1982).

3. An additional series of benthic samples (biota only) are collected from the reference area and are pooled into a common container to estimate initial composition and density. Sub-samples are then removed and randomly inoculated into each artificial stream until the density of invertebrates approximates ambient densities.

Standardization of inoculation protocols is important because the sequence through which species from a common-source pool are added to mesocosms can produce large differences in community structure; these dissimilarities are unrelated to the intended study treatments (Drake et al. 1996). The test community includes multiple trophic levels constructed from random samples taken from the study river. Consequently, measurements of community-level variables, such as species composition, community production/respiration, and decomposition, are possible. These community-level variables integrate both the direct effects of stressors and, importantly, indirect ecological effects that cannot be simulated in single-species or simple food-chain systems (e.g., competition-mediated shifts in community structure, or biomagnification) (Carlisle 2000).

Following biotic inoculation, the mesocosm system is left to acclimatize for 24-48 hours (water flow only) prior to commencement of effluent delivery to the treatment streams.

9.2.7.2.2 Modular Mesocosm – General Set-up

The modular system that was described above for fish is also commonly used with benthic invertebrates. The mesocosm consists of wet tables (one per treatment) upon which partial flow-through streams are placed, and below which a reservoir containing treatment water for circulation to the streams is located. River water is pumped to a head tank, then distributed to each wet table reservoir by pumps (Culp et al. 2004). Water and effluent are pumped through distribution manifolds to the replicate artificial streams. Treated effluents are delivered to the mesocosm system daily or by truck every 2-3 days. The hydraulic residence time of each table reservoir is typically 1 hour for benthic invertebrate studies, while residence time in the circular artificial streams is about 4-5 minutes. This turnover time can vary depending upon effluent and

reference water availability. Water velocity in the modular streams can be produced with paddle wheels that generate velocities of 11-12 cm/second; these velocities are typical of the substrate-water interface in rivers. Insect emergence traps are placed over each stream, and the wet tables are covered by a shade canopy that reduces light levels by approximately 60% to better simulate light levels at the river substratum.

The artificial streams are designed to simulate typical riffle communities of reference areas. Benthic food webs are established across all treatments and replicates by inoculating each stream with substratum extracted from a reference area not influenced by effluents. The stream bed substrate is handled carefully so that the associated microbes and algal biota remain intact. Using these techniques, algal growth in all streams is sufficient for invertebrate inoculation in less than 7 days. Similar benthic invertebrate communities are established in all stream mesocosms by inoculating each stream with biota from the reference area upstream. The area sampled establishes initial invertebrate densities of ~ 1.2-1.4 times ambient levels in the mesocosms to adjust for the possibility of initial handling mortality. Invertebrate communities are allowed to acclimate to the experimental conditions for 24 hours before the effluent dose is applied.

9.2.7.2.3 Maintenance and Monitoring

Trailer Mesocosm System: Maintenance and monitoring is as described for the fish in section 9.2.6.2.3.

Modular Mesocosm System: Daily maintenance of the systems includes regular calibration of all pumps and delivery systems to ensure target delivery volumes and current velocity are achieved. In addition, drain screens and the algae on inner stream walls are frequently brushed to prevent fouling of the streams.

Weekly grab samples of effluent, reference water and the treatments are collected and analyzed for general chemistry, nutrients and metals. Adult insects are collected from emergence traps each day with an aspirator and preserved in 80% ethanol for later identification to the family level.

9.2.7.3 Sampling and Analysis of Benthic Invertebrates

For benthic invertebrate sampling, streams are also sampled in a random selection to minimize differences due to the time of sampling among replicate streams within the same treatment.

Trailer Mesocosm System: Invertebrate samples are collected by lifting the sampling bags within each stream, which results in 5 sub-samples per stream. Sub-samples are then washed through a 500- μ m mesh sieve. Field sieving is required immediately after sample retrieval and before preservation. The recommendation for sieve and/or mesh size for all freshwater mesocosm applications is 500 μ m. In freshwater, macroinvertebrates are defined as those retained by mesh sizes of 200-500 μ m (Slack et al. 1973; Weber 1973; Wiederholm 1980; Suess 1982), although immature life stages of some taxa may be smaller and some adult life stages may be larger. Note that these mesh sizes are applicable to all equipment used in the field and laboratory (i.e., both the Nytex mesh on the benthic samplers and sieving apparatus). In some

site-specific circumstances it may be desirable for the samples to be screened for smaller organisms by using a smaller sieve size. Some examples of situations where the use of a smaller mesh size (less than 500 μm) may be appropriate include the following:

- 1) for comparative purposes if historical benthic surveys for the system under investigation utilized smaller mesh sizes; or
- 2) if sampling needs to be conducted, for logistical reasons, at times when organisms are very small; Rees (1984) and Barber and Kevern (1974) provide information on seasonal effects of mesh size.

In these aforementioned cases, it is highly recommended that a stack of screens be used that minimally have the mandatory sieve sizes, and then any other smaller sizes that are appropriate. This procedure simultaneously allows site-specific concerns to be addressed and fulfills the EEM objectives by allowing for national or regional comparisons to be conducted on the standardized mesh sizes.

Modular Mesocosm System: Benthic invertebrates are collected at the end of the experiment by washing the entire contents of each stream through a 250- μm sieve and preserving the samples in a 10% formalin solution. In the laboratory, benthic invertebrate samples are sorted under 12x magnification, identified to the family level, and enumerated.

All summary statistics and descriptive metrics should be calculated and reported at the family level for submission in the EEM interpretative reports (see section 4.6.2 of Chapter 4). Organisms that cannot be identified to the desired level of taxonomic precision should be reported as a separate category in the fundamental data set. It is recommended that investigators use taxonomic keys appropriate to the geographic region of study; a detailed list of taxonomic references for various groups of freshwater organisms is provided in Chapter 4.

9.2.7.4 Data Assessment and Interpretation

Use of mesocosms as a monitoring alternative for assessing effluent effects on the benthic invertebrate community largely follows the guidance of Chapter 8 regarding data assessment and interpretation in field surveys, as the effect endpoints measured to assess effects are the same. The only difference in data assessment is that replication per treatment in mesocosm studies is $n = 8$ for control/impact designs rather than the $n = 5$ as recommended for field surveys.

During the effects assessment, a significant difference between reference and exposure areas in any of the following effect endpoints is to be interpreted as an effect on the benthic invertebrate community: total benthic invertebrate density, taxa richness, evenness Index, and similarity index (MMER Schedule 5, section 1).

Diversity, taxon density, proportion or presence/absence are recommended to aid in the interpretation of effects. Calculation of these metrics is described in detail in the benthic invertebrate community survey (Chapter 4). Details on recommended statistical analyses, data QA/QC, and reporting requirements are as outlined for field surveys, with the qualification that a

mesocosm treatment level is equivalent to an exposure area and a replicate artificial stream is equivalent to a field replicate station.

9.2.8 Supporting Measurements

9.2.8.1 Water Chemistry Parameters

In mesocosm studies, it is essential that stream differences in physical water chemistry be minimized as much as possible to ensure effluent-related effects are not confounded. Streams are monitored daily for temperature, dissolved oxygen, pH, and input flow for water and effluent. Current velocity should be measured at the start and end of each study, especially for benthic invertebrate studies. The distribution of water velocities in the streams is characterized using one of many brands of current meter. In previous studies, mean velocity in the mesocosms (above stones) ($x = 0.26 \pm 0.01$ m/second, $n = 150$) was similar to water velocity measured above stones at a similar water depth in the field (water velocity $x = 0.26 \pm 0.01$ m/second, water depth $x = 24.8 \pm 0.72$ cm, $n = 30$) while the study was being conducted (Podemski 1999).

Water temperature can be measured using continuous temperature data loggers placed in a stream and the head tank. Temperatures in the head tank reflect the temperature of incoming river water. In previous studies, a comparison of data from these 2 thermographic locations indicated that the 2-hour hydraulic residence time in the streams resulted in slight heating or cooling of water in the streams depending upon ambient air temperatures (Culp and Podemski 1996). For example, over a 3-day period, the streams were cooler at night and warmer during the day as compared to the incoming river water. The maximum instantaneous difference between water temperature in the river and the streams was $< 5^{\circ}\text{C}$.

9.2.8.2 Supporting Water Quality Parameters

In addition to assessment of daily changes in physical water chemistry among all of the mesocosm streams, water samples should also be collected on a less frequent basis for analysis of general chemical parameters. It is recommended that samples be collected from the reference water head tank, the full-strength effluent, and each mesocosm stream upon completion of the study. Samples should be analyzed for parameters as outlined in Chapter 5. Often this information proves invaluable to confirming effluent dilutions in the treatment streams. For example, in freshwater systems exposed to pulp mill effluents, sodium (Na) is a relatively conservative ion that appears in high concentrations. By comparing sodium ion (Na^{+}) concentrations in the reference and exposure streams, the concentration of effluent in the streams can be verified.

9.3 Use of Caged Bivalves as an Alternative Monitoring Method

9.3.1 Introduction

A description of methods for caged bivalve studies is provided in this section, which includes detailed guidance on:

- background and general approach of caged bivalve studies;
- species selection;
- study design;
- variables to be measured;
- methods for implementing the study;
- data analyses;
- tissue concentrations; and
- reporting requirements.

9.3.2 Background

In October 2000, the National EEM Science Committee recommended the use of caged bivalve studies as a scientifically defensible alternative approach to a wild fish survey if it is not practical or technically feasible to conduct the fish survey. The Metal Mining Fish Subgroup also recommended the use of caged bivalves as an alternative method during the metal mining EEM multi-stakeholder consultation process. For more information see Courtenay et al. (1998), Andrews and Parker (1999), and Applied Biomonitoring (2000).

In November 2000, the American Society for Testing and Materials (ASTM) approved a method for conducting environmental studies using caged bivalves (Salazar and Salazar 2000). The ASTM method serves as the basis for this technical guidance on conducting caged bivalve studies. A number of other studies using caged and wild bivalves were also considered in developing this specific guidance for applying this approach within the framework of the EEM program. The format of this guidance follows that used for the fish survey. Bivalves, such as oysters and mussels, have been used in Mussel Watch programs since the mid-1970s to monitor trends in chemical contamination and assess the effects of human activities on coastal and estuarine areas. Mussel Watch programs began in the United States (Goldberg et al. 1978) and have since become international in scope (Jernelov 1996). The following are some of the reasons why bivalves are suitable test species:

- bivalves are relatively non-mobile, such that exposure to contaminants is assured and is representative of the exposure area;
- bivalves are abundant in many marine, estuarine and freshwater environments, and are relatively easy to handle and sample year-round;
- the biology of many shellfish species is well known and considerable research has been conducted regarding effects on shellfish of exposure to various anthropogenic and natural environmental stressors;

- several bivalve species have been shown to readily accumulate many chemicals from a variety of pathways (water, sediment, food) and show sublethal effects associated with exposure;
- bivalve growth is relatively easy to measure and has been shown to be as sensitive or more sensitive than mortality in other standard test species such as *Daphnia*, Fathead Minnow and Rainbow Trout (see Salazar and Salazar 2000); and
- bivalves are an important fisheries resource, with both the Atlantic and Pacific regions having commercially valuable shellfish aquaculture industries as well as commercial and recreational shellfish harvests.

9.3.3 Approach

The caged bivalve approach provides a number of advantages to the investigator in conducting a monitoring program (Crane et al. 2007). These include experimental control and realism, use of organisms naturally found in the study environment, and known exposure period. By using caged bivalves rather than resident populations, the variability in biological measurements can be reduced by using individuals of similar size and environmental history, thereby increasing the discriminating power of the test (Crane et al. 2007; Salazar and Salazar 1995). A considerable number of caged bivalve studies have been conducted in Canada and the United States, as well as other countries (see Salazar in Stewart and Malley 1997; St-Jean et al. 2003, 2005; Crane et al. 2007).

The effect indicators for caged bivalve studies in EEM are survival, growth, condition, reproduction and energy storage. A tissue analysis (mercury) may be required, and caged bivalves can be used to meet this requirement. Other chemicals or metals may be used to assess bioaccumulation to aid in interpreting results or for use during investigation of cause.

One of the difficulties associated with caged bivalve exposures in the EEM program is related to the difficulties in comparing between responses obtained through the adult fish survey and the caged bivalve exposures. The difficulty lies with the following assumptions:

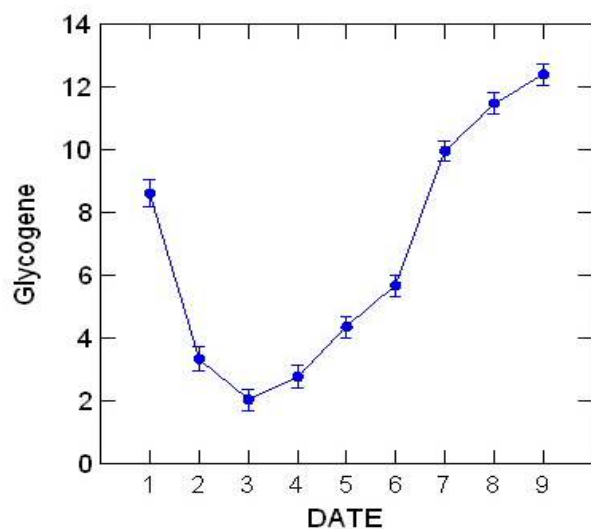
- 1- Mussels used in caged bivalve studies generally originate from clean areas or reference sites, while fish in the adult fish survey are generally long-term residents; therefore, their responses cannot be expected to be the same.
- 2- At most sites in Canada, the reproductive cycle of mussels (Blue Mussels [*Mytilus edulis*]) spans a minimum of 9 months, whereas gametes produced in the spring are derived from energy (mostly glycogen) accumulated in the fall; therefore, a 60- or 90-day exposure in the spring will have difficulties capturing effluent effects on reproductive effort.
- 3- In the mussels, the same organ is used for energy storage and reproduction. In the fall, the mantle (Figure 9-10) is mostly composed of energy (glycogen), which will be used to develop the eggs in the spring. Figure 9-11 represents the cycle between energy and eggs in a population of Blue Mussels from British Columbia.



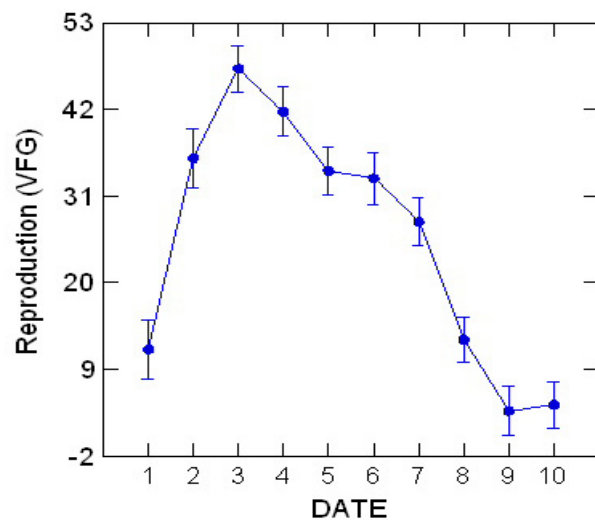
© S. St-Jean

Figure 9-10: Mussel showing ripe mantle lobe

A)



B)



© S. St-Jean

Figure 9-11: Reproductive cycle of Blue Mussels from British Columbia: A) Mantle energy stored in fall; B) Mantle reproductive content in spring

Notes: Numbers on the axis represent months: from February (1) to November (10); A) Glycogen is expressed as mg/g; B) Reproduction is expressed as volumetric fraction of gametes VFG.

Therefore, in order to maximize caged bivalve results and comparability to the adult fish survey, and also facilitate the interpretation of results in terms of reproduction and energy, exposure of adults should occur from the onset of energy accumulation in the fall until the release of gametes in the spring.

9.3.4 Species Selection

Many bivalve species have been used for assessing chemical bioavailability or effects in marine, estuarine and freshwater environments. Ideally, species that have wide geographic distributions should be used so that test results can be compared across studies. Species selection for caged bivalve studies should be made carefully and should consider the biology of the species and local conditions, such as:

- Are conditions at the exposure and reference areas similar to the natural habitat of the species in terms of tolerance limits for natural factors such as temperature, salinity, dissolved oxygen and pH? Is the species naturally present in the area under evaluation?
- Is there documentation indicating that the species can accumulate and/or be sensitive to the contaminants of concern?
- Is the life history of the species well known in terms of spawning cycle and life-stage requirements?
- Does the species have threatened or endangered status?
- Is an abundant supply of the species readily available?
- Is the species easy to handle in the field?

For individual mussels, care should be taken in regards to the following:

- Are specimens' shells abnormally thick?
- Does the shell have signs of worms? (holes in the shell are often a telltale sign)
- Is the shell cracked?
- Does the mussel have a slow-closing valve reflex?

9.3.4.1 Most Commonly Used Taxa

The species most commonly used in field bioassays in Canadian waters and considered relevant for use in the EEM programs are described below and listed in Table 9-4. The temperature and salinity tolerance limits for each species are provided, along with information on age at maturity, spawning periods and general distribution within Canada. Other species, such as marine clams (*Mya arenaria*, *Macoma balthica*) and scallops, may also be suitable for some applications. However, the environmental requirements and sensitivity of alternative bivalve test species should be established before they are used in an EEM program. For species-specific needs, Salazar and Salazar (2000) described the salinity, temperature and general distribution of several species of bivalve in Canada.

Table 9-4: Suggested taxa for use in caged bivalve studies for EEMs

Species and Reference	Temperature Range (°C)	Salinity Range (parts per thousand)	Reproductive Information	General Distribution in Canada
Marine and Estuarine Bivalves				
<i>Mytilus edulis</i> (Blue Mussel) (Freeman et al. 1994; Grout and Levings 2000; Mucklow 1996; Stewart 1994; Salazar and Salazar 2000; Newell 1989; Toro et al. 2002)	-1.5 to 25	5 to 33	Most energy utilized for spawning at length greater than 3.5 cm, or roughly 2.5-4 years old. Generally an abrupt spawning on the East Coast: no more than 3 weeks, between mid June and mid July. But spawning may vary among populations; some low-level spawning throughout year, mostly in area of anthropogenic influence or repeat spawners; first in early summer, second in the fall, mostly on the West Coast.	Atlantic coast
<i>Mytilus trossulus</i> (Bay Mussel, or Foolish Mussel) (Freeman et al. 1994; Salazar and Salazar 2000; Skidmore and Chew 1985; Toro et al. 2002)	0 to 29	4 to 33	Most energy utilized for spawning at lengths greater than 3.5 cm. Spawning generally spans 12-13 weeks from June to September.	Atlantic and Pacific coasts
<i>Crassostrea gigas</i> (Pacific Oyster) (Waldock et al. 1996)	4 to 24	25 to 35	Spawning July to August	Pacific coast
Freshwater Bivalves				
<i>Elliptio complanata</i> (Eastern Elliptio) (Beckvar et al. 2000; Day et al. 1990; Hinch et al. 1989; McMahon 1991; Metcalfe-Smith et al. 1996)	0 to 30	0 to 3	Age at maturity 6-12 years. Spawning occurs mostly June to July; some May to September.	Eastern Canada
<i>Pyganodon (Anodonta) grandis</i> (Common Floater Clam) (Clarke 1973; Couillard et al. 1995a, 1995b; Malley et al. 1996)	0 to 30	0 to 3	Spawns mostly April to May, some to August	Interior and eastern Canada
<i>Anodonta kennerlyi</i> (Western Floater Clam) (Clarke 1981; Stewart and Malley 1997; Williams et al. 1993)	0 to 30	Freshwater	Spawning begins in early August.	Alberta and British Columbia
Sphaeriid clams (e.g., <i>Musculium securis</i> , <i>Sphaerium rhomboideum</i> , <i>Sphaerium striatinum</i>) (Hornbach et al. 1982; Mackie 1978a, 1978b; Mackie and Flippance 1983; Mackie et al. 1974; Stephenson and Mackie 1981.)	10 to 25* (*optimal growth range)	Freshwater	Life cycle generally 1 year; life histories of many species are well documented; reproductive effort can be quantified.	Widely distributed in Canada

Format modified from Salazar and Salazar 2000

9.3.4.2 Marine and Estuarine Bivalves

The Pacific Oyster is a species that has been used in transplant studies in marine and estuarine studies (Waldock et al. 1996). Its shells are usually more difficult to measure because of their irregular shape and protrusions.

Mytilus have been used extensively in the ongoing International Mussel Watch Project to monitor trends of chemical contamination and assess the effects of human activities on coastal and estuarine areas in North America and around the world (O'Connor 1992; Jernelov 1996). Blue Mussels (*Mytilus edulis*) and Bay Mussels (*Mytilus trossulus*) are found on the Atlantic Canadian coast, whereas mostly *M. trossulus* is found on the Pacific Canadian coast. These two species may be easily confused where they co-occur on the Atlantic coast (Freeman et al. 1994; Mucklow 1996), and since their biology and reproductive cycles differ, species identification is essential. *Mytilus* spp. are often referred to as the *M. edulis* complex, in recognition of biochemical differences (Varvio et al. 1988), and may comprise the species *M. edulis*, *M. galloprovincialis* and *M. trossulus*. Several fundamental differences have been observed between *M. edulis* and *M. trossulus*, including gamete incompatibility (Rawson et al. 2003), temporal separation and duration of spawning in Atlantic mussels (Toro et al. 2002), and total egg production and size (Toro et al. 2002). Bay Mussels have smaller eggs, with longer spawning times and less gamete production than Atlantic mussels. The growth rate for *M. edulis* is faster than *M. trossulus* on the east coast of Canada (Penney et al. 2002). Work carried out on caged mussels on the Pacific and Atlantic coasts has confirmed differences in growth and reproductive cycles between the species, showing *M. trossulus* to be more sensitive, smaller, and producing fewer and smaller eggs (Metro Vancouver, unpublished data). Table 9-5 lists several differences noted between the two species over five years of analysis.

Table 9-5: Differences noted between two species of mussels over 5-year study in the Burrard Inlet, Vancouver, British Columbia

<i>Mytilus edulis</i>	<i>Mytilus trossulus</i>
Faster growth	Slower growth
Higher survival	Lower survival
More eggs / larger eggs	Smaller eggs / fewer eggs
Reproduction over 2-3 weeks	Reproduction over 9-14 weeks
Egg production more clearly separated from energy storage	Egg production less separated from energy storage
Less susceptible to leukemia	More susceptible to leukemia
Overall better suited for monitoring	Overall not best suited for monitoring
Not always present (West Coast)	Sometimes co-occur (East Coast)

9.3.4.3 Freshwater Bivalves

Freshwater unionid, or freshwater mussels, have been used in a number of caged studies to examine water column and sediment exposures. Unionid bivalves (such as *Elliptio complanata*, *Anodonta kennerlyi* and *Pyganodon grandis* [formerly *Anodonta grandis*]) or sphaeriid clams might be considered suitable for assessing differences in the survival, growth, condition and reproductive rates of bivalves in freshwater receiving environments. The following authors have discussions on the life cycles of these species: Mackie (1978b), Sandusky et al. (1979), Stephenson and Mackie (1981), and Stewart and Malley (1997). Freshwater mussels are bivalves belonging to the super-family Unionoidea and comprise one of the most endangered groups of organisms in North America (Wolfe et al. 2009). The unionids are notable in that their glochidia require incubation in a vertebrate host for survival to adulthood. Glochidia are the parasitic larval stage of unionid mussels that attach to the fish host after release from the adult mussel. Glochidia remain on the host fish until metamorphosis is completed, the duration of which is dependent on water temperature. The glochidia of different genera are released at different times of the year (Bauer 1994). The various taxa of the genus *Pisidium* may be too small for practical handling and, in addition, are taxonomically difficult for the non-specialist. Additional research may be required to demonstrate the utility of sphaeriid bivalves for use in EEM. Freshwater bivalves do not have fused gonads as marine mussels do, but have distinct gonads. Freshwater bivalves also display distinct seasonal cycles in tissue biochemical content, related mostly to the reproductive

cycle. As with the marine and estuarine bivalves, proteins, glycogen and lipids content are maximal during gonad development and gametogenesis, and minimal during glochidial release. As with their marine counterparts, glycogen presents the most variation (Jadhav and Lomte 1982). Table 8-6 lists the differences between Unionoidea and Sphaeriidae.

Table 9-6: Differences noted between Unionoidea and Sphaeriidae

Unionoidea	Sphaeriidae
Fast growth until maturity; then slow	Slower growth; bivalve can be very small
Life span: < 6 to > 100 years	Life span: 1-4 years
Fecundity: 200 000 to 17 000 000 eggs per female; small eggs	3-24 eggs per female; large eggs
Reproduction: one per year	Reproduction: 3 per year, sometimes continuous
Egg production more clearly separate from energy storage	Egg production less separate from energy storage
Gonochoristic	Hermaphroditic
Age at maturity: 6-12 years	Age at maturity: 0.2-1 year
Less suited for studies: gonochoristic, long-lived, interoparous, often rare, difficult to collect, complicated life cycles (parasitic stage)	Better suited for studies: greater abundance, ease of collection, ease of maintenance, relatively simple life cycle, short life span

9.3.4.4 Source

An important consideration will be whether or not to use farmed mussels or native local mussels. Depending on the parameter being measured, there are advantages and disadvantages to both; the choice should be made in consideration of the circumstances outlined below.

Bivalves should be obtained from a commercial grower when the parameter being measured is growth or chemical accumulation. Native mussels should be used for reproductive and energy parameters, preferably using a natural gradient or control impact design if the exposure period is short, or commercial mussels when the exposure period is longer. Growth should be measured using juveniles obtained from an aquaculture facility. Bivalves obtained from a commercial

facility have an environmental history that is well known, and assurances of being uncontaminated are greater than for animals collected in the wild. In any study, some form of species identification should be confirmed, as results may vary significantly between species, as outlined in Table 9-6. All individuals used in a caged bivalve study should be from the same population for the same parameter. If wild populations are the only possible source, they must be collected from an uncontaminated area. Epiphytic growth on bivalve shells should be removed gently by hand or with a soft brush or scraper. Collection permits for field-collected or transplanted bivalves are required by Fisheries and Oceans Canada and may be required by some local or provincial agencies. In addition, if cages present a potential obstruction to navigation, a permit may be required from the Canadian Coast Guard and a Notice to Mariners may be required. The permitting process should be considered early in the study planning process. It may take several weeks or months to process the necessary permits, as there may be a need for an environmental assessment as required by the Canadian Environmental Assessment Agency.

9.3.4.5 Species Identification

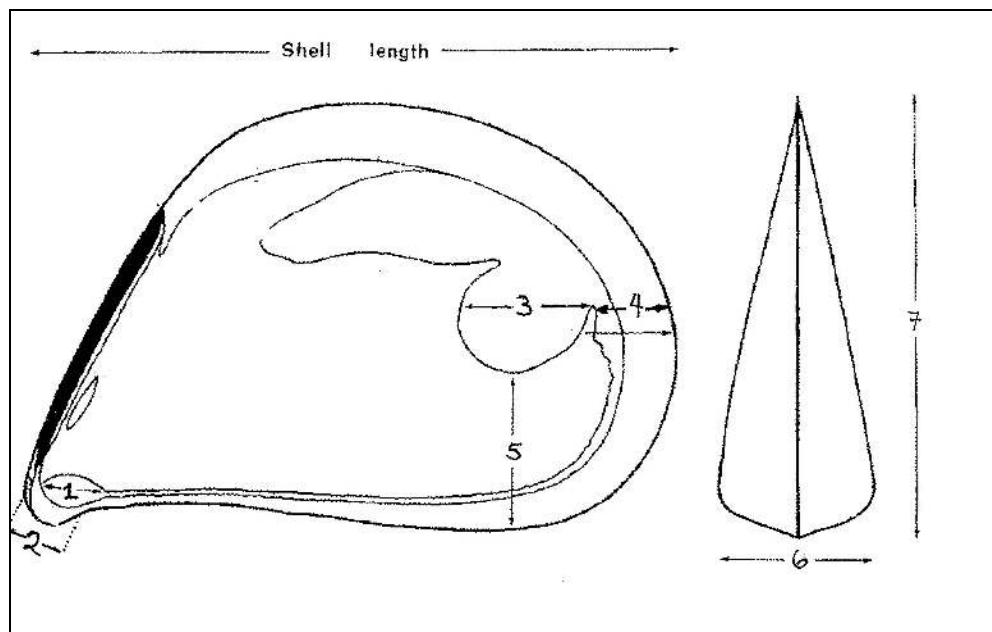
Mussels can be identified by two methods: allozyme analysis or by morphometric measurement combined with statistical analysis (McDonald et al. 1991; Mallet and Carver 1995). For the morphometric measurements, empty mussel shells are scraped to remove any remaining tissues and dried for 4-5 hours (60°C). A minimum of five shell characteristics (listed below) should be measured using a stereo microscope (6.4x magnification):

- the length of the anterior adductor muscle scar;
- the length of the hinge plate;
- the distance between the anterior edge of the posterior adductor muscle scar and shell margin;
- the distance between the ventral edge of the posterior adductor muscle scar and ventral shell margin; and
- the distance between the pallial line and ventral shell margin midway along the shell (Figure 9-12).

Three additional shell characteristics should be measured with callipers:

- shell length;
- shell width; and
- shell height.

Each characteristic should be standardized using $\log_{10}$ and divided by the $\log_{10}$ shell length. These morphometric variables (log-transformed and length-standardized as appropriate) should then be multiplied by their raw canonical coefficients and summed to generate a canonical variate for each individual (Mallet and Carver 1995). *Mytilus edulis* typically has a longer anterior adductor mussel scar, a longer hinge plate, and a greater shell height than *Mytilus trossulus*, resulting in positive values of the standardized canonical coefficients.



© S. St-Jean, adapted from McDonald et al. 1991

Figure 9-12: *Mytilus* spp. shell scars markings

Note: 1) length of anterior adductor muscle scar; 2) length of hinge plate; 3) length of posterior adductor muscle scar; 4) distance between posterior edge of posterior adductor muscle scar and posterior shell margin; 5) distance between ventral edge of posterior adductor muscle scar and ventral shell margin; 6) shell width; 7) shell height.

9.3.4.6 Size and Age

All bivalves used in a caged study should belong to the same age class and be as uniform as possible in size. Typically, juvenile mussels intended for growth measurement should not have more than 5 mm difference in length at the onset of the exposure. This minimizes the number of individuals required to achieve adequate power. Juveniles are the best candidates for this parameter, as most of their energy is directed toward growth. Bivalves, and mussels in particular, have an inverse relationship between energy directed toward growth and energy directed toward reproduction. Juvenile mussels will expend most of their energy on growth with little input toward reproduction, and the ratio between growth and reproduction will slowly shift as most energy is directed to reproduction and little to growth in adults. Typically, mussels can start producing some gametes as young as 1 year old, and by 3 or 4 years old most of their energy will be directed toward reproduction and they will grow at a much slower rate. For energy and reproductive output measurement, mussels older than 3 years of age are recommended (generally at least 4 cm in length), while juveniles (between 2 and 2.5 cm) are recommended for growth.

Mussels can be aged following a combination of the techniques described in Ramon and Richardson (1992) and Sejr et al. (2002). The technique is based on annual growth bands and has been validated using a mark-and-recapture approach (Sejr et al. 2002). An experienced biologist should perform this task.

9.3.4.7 Number of Organisms

A power analysis should be used to determine the minimum number of bivalves needed to detect a specified “effect size.” It is recommended that studies be designed to detect a 20% difference in growth. Data sets to aid in predicting the number of required organisms may be available through Environment Canada. However, when the range in mussel length is an average of 5 mm, 100 mussels will be sufficient to achieve power and to ensure an adequate number of mussels survive the exposure.

The number of animals required per cage for the growth measurements will depend on the study design (e.g., number of cages per station; see more discussion on this in the following sections), species used, age of animals, variability in response to the station, and growth conditions at the station. The number of bivalves required to fill the cages will depend on study design in terms of:

- number of areas (e.g., exposure area plus number of reference areas);
- number of stations per area;
- number of cages per mooring (if the study is designed to address varying depths in the water column); and
- recommended bivalves per cage.

In addition, if tissue samples are required for tissue analysis, consideration must be given to the number of animals required to obtain a sufficient sample size for chemical analysis.

9.3.4.8 Handling and Holding Conditions

Salazar and Salazar (2000) provide detailed guidance on handling and holding conditions for bivalves, and this is summarized below. Test organisms should be handled as little as possible and should be deployed as soon as possible after collection. When handling is necessary, it should be done carefully, gently and quickly so that the bivalves are not needlessly stressed. Bivalves should be kept in well-aerated, clean-flowing water as long as possible between collection, sorting and deployment. If transporting bivalves for extended periods, keep them moist and cool by placing them in a cooler with frozen gel packs or wet ice (wet ice at the bottom of the cooler). Use seaweed or cloth towels to keep bivalves separated from the gel pack or wet ice. Newspaper should be avoided, as it contains ink that can be toxic to mussels.

9.3.5 Study Design

Determining the appropriate study design is critical if the results are going to be meaningful. The study design for caged bivalves should include a number of components:

- sampling design;
- area and station selection;
- replication of cage stations and cages per station;
- timing and duration;
- modifying or confounding factors;

- supplementary measures;
- cage design; and
- mooring systems.

9.3.5.1 Sampling Design

There are six main sampling designs that are recommended for caged mussel marine and freshwater assessments:

- control-impact (C-I) design;
- multiple control-impact (MC-I) design;
- simple gradient (SG) design;
- radial gradient (RG) design;
- multiple gradient (MG) design; and
- control-simple gradient (C-SG) design.

The C-I and MC-I designs are used to determine the magnitude of difference between homogeneous exposed and unexposed areas, while the SG, RG and MG designs examine changes in an effect along an effluent gradient. The C-I and MC-I designs are used when there are few qualitative levels of exposure. It is suggested that multiple reference areas be used rather than increasing sample sizes in one reference station. The SG, RG and MG designs may be used when there are many quantitative levels of exposure (Paine 2000); they may also be useful in discriminating among effects from sources other than metal mine effluent. Guidance on selecting the appropriate design is provided in Chapter 2.

For example, in estuaries with complex tidal regimes and mixing regimes, an MC-I approach may be the most appropriate. An SG approach may be applicable to a river receiving environment where flow is unidirectional.

C-I, C-SG and MC-I designs require some level of replication within the control and impact areas. C-SG is a combination between C-I and SG, which is sometimes useful when more than one reference site is desired. In this case, a control, or reference station, is added to a simple gradient, typically when the very low effluent exposure (far-far field) area does not have conditions similar to the exposed area in terms of depth or other important biotic factors. Replication at each station may not be needed for gradient designs, although an appropriate number of stations are needed in order to discriminate between spatial patterns related to effluent discharge and other spatial patterns in the environment. Replication is discussed below in section 9.3.5.2. Ultimately, it is the responsibility of study designers to develop site-specific study designs that are scientifically defensible, robust and suitably sensitive.

9.3.5.1.1 Area and Station Selection

Chapter 2 provides guidance for the selection of multiple reference locations for a variety of receiving environments and is applicable to caged mussel studies. Reference areas should be as similar as possible to study areas in terms of the water's:

- depth;
- hydrodynamic conditions;
- temperature;
- salinity;
- dissolved oxygen concentration; and
- food availability and quality.

Multiple reference stations may help to identify natural differences and variability among uncontaminated areas. Often, a mixture of simple gradient and multiple control areas, such as C-SG, allows for a more robust study.

The exposure area is defined by plume characteristics. Plume delineation, as described in Chapter 2, should provide sufficient information to model average effluent concentrations with distance from source and to identify the degree of vertical mixing in the water column. This information will assist in the selection of stations and depth of deployment in the water column. More guidance is provided in the sections below.

9.3.5.2 Replication

9.3.5.2.1 Number of Cage Stations

For the exposure area, stations and depths of cage deployment may be chosen to represent a gradient of exposure to the effluent plume. The deployment stations of bivalve cages should be given careful consideration for a number of potentially interfering factors, such as the following:

- In estuarine and marine situations, effluent may be positively buoyant, resulting in a thin layer of “freshwater” effluent floating on denser saline water.
- In tidal situations, the mixing behaviour of effluent may be quite complex, resulting in low confidence about the average exposure concentration to which bivalves may be exposed. Tidal situations may require consideration of exposure area stations that are both “upstream” and “downstream” of the outfall.
- In river discharge situations, effluent plumes may be quite long and narrow, meaning that there is little or no opportunity to meaningfully replicate cage stations within (i.e., across the long axis of) the plume.

For these and other reasons, it might be more meaningful to simply evaluate “distance from outfall” rather than effluent concentrations when designing caged bivalve studies. This approach would remain consistent with one of the objectives of EEM, which is to evaluate the magnitude and geographic extent of effects that may be related to the effluent discharge.

The number of replicate stations must be determined to address the sampling design and EEM objectives. The number of replicate stations and sub-samples within replicate stations are determined by power analysis. The allocation and distribution of replicate stations is dependent upon the sampling design. Guidance on the use of power analysis is provided in Chapter 8. Readers are encouraged to ensure that no pseudo-replication occurs, but rather true replication. Hurlbert (1984) offers more guidance on the subject.

9.3.5.2.2 *Applying Replication to Study Design*

Several parameters are suggested for caged bivalve studies (survival, growth [change in length or wet weight], soft tissue fresh weight, condition, reproduction and energy storage). In order to assess these parameters on a set of bivalves exposed at different locations in the field, many different configurations of cages and bivalves are possible. For EEM studies, it may be advantageous to deploy replicate cages containing multiple bivalves at each station, and to consider only the average performance within each cage. This approach confers the additional advantage of providing redundancy in case one or more cages are lost, and may simplify the construction and deployment of cages. So long as the statistical hypothesis testing is confined to evaluating whether or not there are significant differences between areas (without uniquely attributing effects to effluent exposure), the replication and statistical analysis is valid.

As general guidance, it is suggested that cages containing 20 animals per cage be deployed for the survival measurements and at least 5 cages be deployed at each station to evaluate growth. These cages should be deployed on individual moorings, not 5 cages on one mooring, since the mooring is the most appropriate unit of statistical replication. However, practitioners should be encouraged to explore the power and robustness of potential study designs, using synthetic data (or empirical data where available) as an integral part of the study design process to determine the minimum number of animals needed.

9.3.5.3 *Timing and Duration*

Timing of the studies should be such that:

- it coincides with a high growth period in natural populations so that growth is maximized and differences in growth rate among treatments are more measurable; and
- it does not coincide with a spawning period if the test bivalves are adults.

For growth, survival and chemical accumulation, the duration of exposure should be 60-90 days (see discussion in Salazar and Salazar 2000). This should provide sufficient time for effects on survival and growth to be manifested. However, a minimum of 9 months may be needed to measure energy or reproduction; typically, on the East Coast, cages should be deployed in the summer to ensure that the energy accumulated in the fall (energy reading) is from those sites and that egg production reflects any potential effects of the effluent. Although 3 samplings are required (deployment, energy and reproduction), the cost is not generally prohibitive, as, unlike growth and survival, no measurements are needed prior to deployment.

9.3.5.4 Modifying or Confounding Factors and Supplementary Measures

Results of caged bivalve studies will depend, at least in part, on natural factors such as temperature, food supply, other physicochemical properties of the test environments, species selected, condition of test organisms, exposure method, and handling of test organisms. Exposure and reference areas should be as similar as possible with respect to the factors listed below, to minimize confounding differences. It may be useful to measure some of the factors in order to assist in interpreting results. These factors may include life cycle, behaviour, temperature, lack of acclimation, current speed, salinity, fouling, chemical concentration and food availability. These are discussed further in Salazar and Salazar (2000).

9.3.6 Cage Designs

For growth and survival studies, cages with individual compartments are suggested so that individual bivalves can be tracked through the study. The mesh size should be maximized to allow maximum water flow but small enough to contain the test animals. Individuals are assigned to compartments such that survival, growth and condition can be tracked in each bivalve. Individual test organisms are placed in the mesh bags and separated by using a plastic cable tie or other suitable tie. Sufficient space should be allowed in each compartment to permit test animals to grow during the exposure period.

For reproduction and energy measurements, mussels do not need to be pre-measured; therefore, compartments are not necessary. However, they should still be exposed to relatively uniform conditions, and stringing mussels into socks, clumped in 3-4 individuals, is recommended (Figure 9-13). This significantly reduces the level of effort needed for this design.

A variety of cage designs are described in Salazar and Salazar (2000). A flat (i.e., two-dimensional) cage design, as shown in Figure 9-13, is suggested, as it is a convenient unit to work with. Polyvinyl chloride (PVC) tubing is a convenient material to use for constructing cages. PVC should be water-supply grade obtained from a high-quality source and soaked for at least 24 hours in flowing fresh or seawater before use to remove water-soluble and volatile chemicals. Alternative materials are described in Salazar and Salazar (2000), section 9.

Final cage dimensions depend on the size of the test organism and the number of organisms per cage. Typical bags sizes for species like mussels and clams are 10-15 cm in diameter with 5-mm mesh size. Each mesh bag should be long enough to accommodate the desired number of bivalves per bag, plus sufficient material for attachment to the PVC frame. For freshwater clam deployments, a wide variety of cage designs have been used by investigators. It is the responsibility of the study designer to ensure that cage designs are appropriate to the test species and receiving environment.

9.3.6.1 Mooring Systems

The cages can be suspended in the water column by attaching them to mooring lines that have an anchor or weight on one end (e.g., iron chain links) and a surface or subsurface buoy attached to

the other end (Figures. 9-14, 9-15). Salazar and Salazar (2000) discuss the factors that should be considered for deployment of cages.

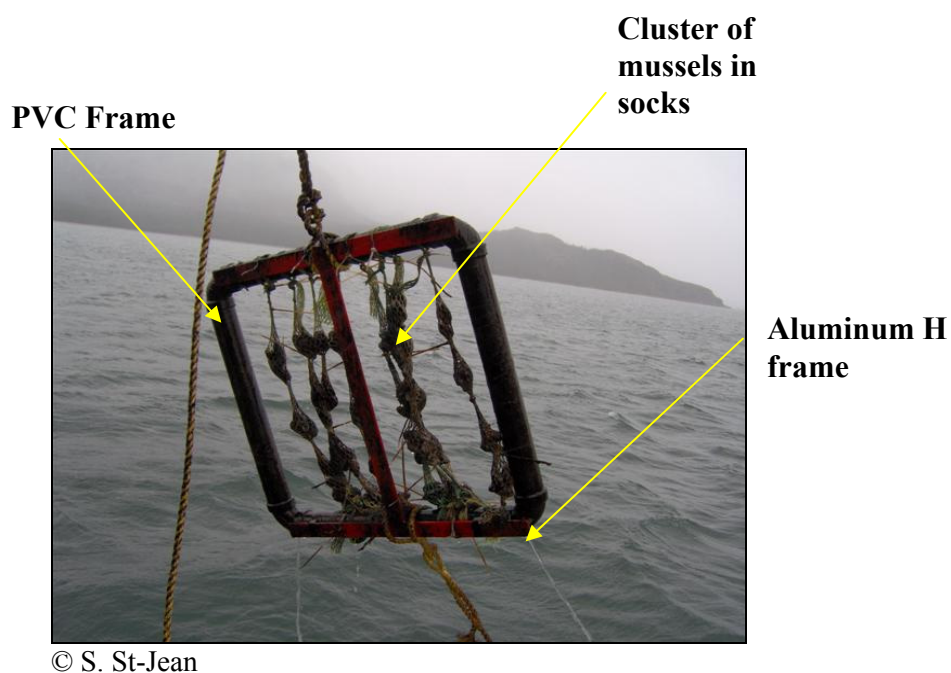
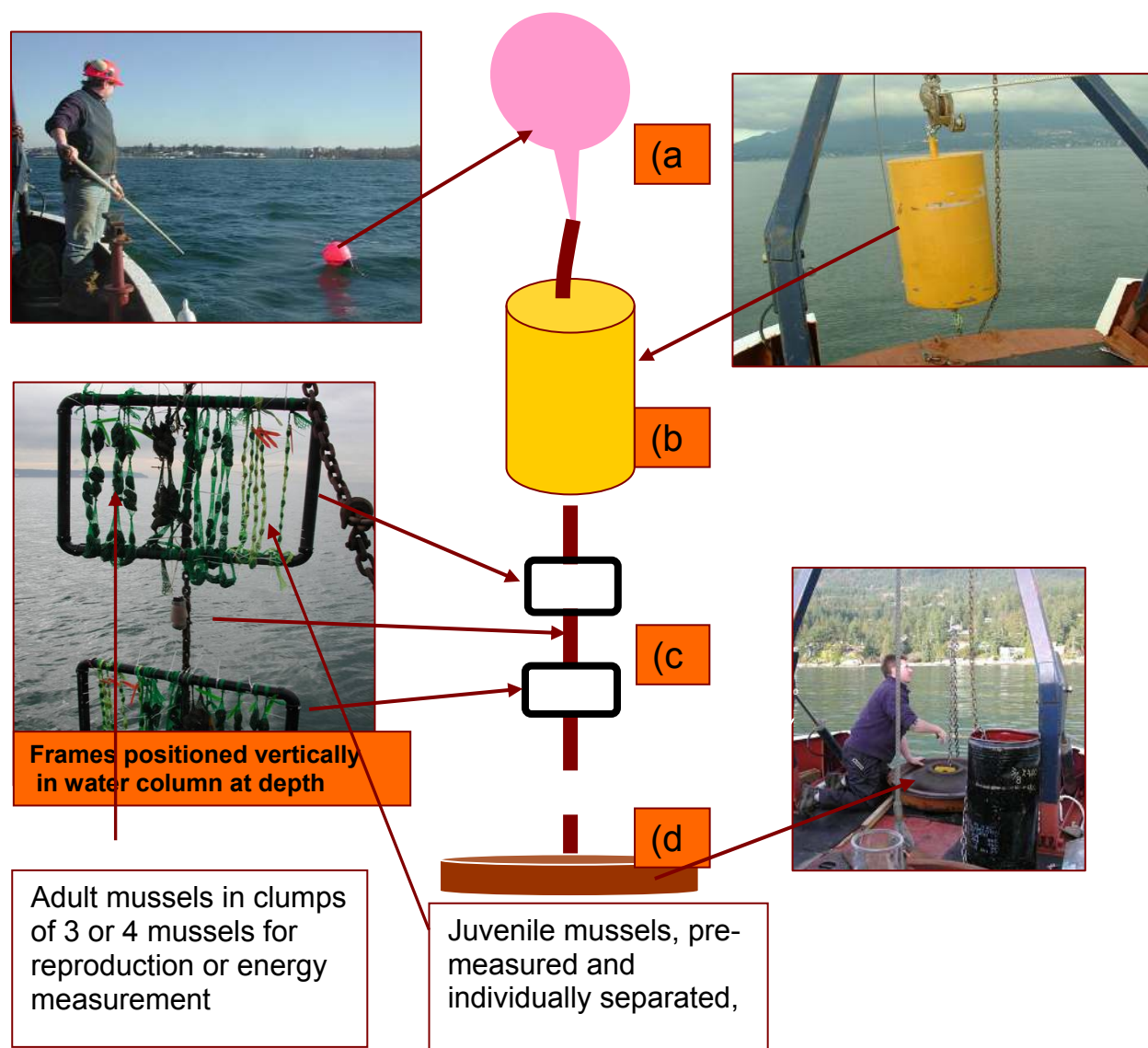


Figure 9-13: Duplicate frame from a caged mussels exposure experiment

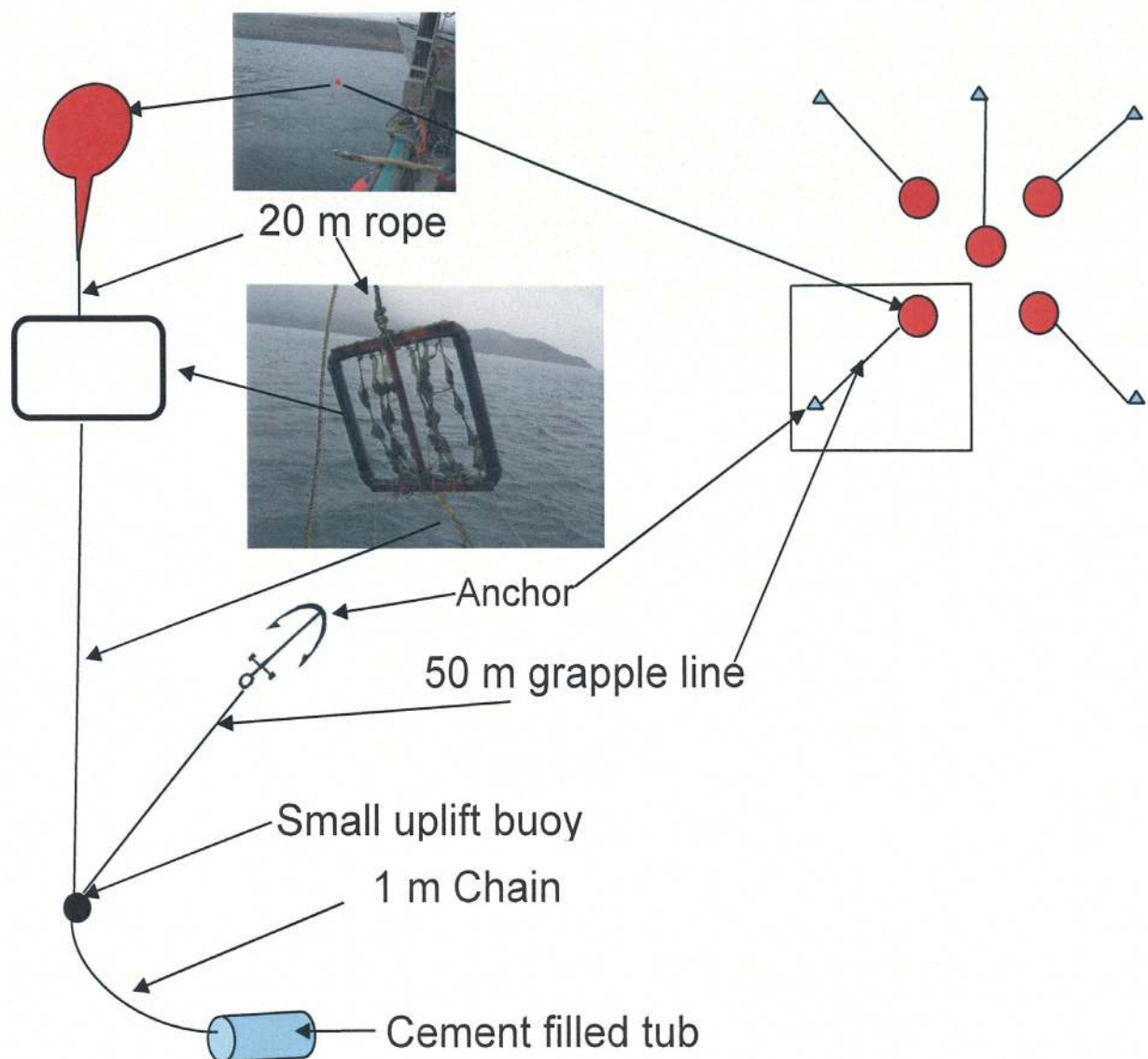


© S. St-Jean

Figure 9-14: Modular mesocosm parts diagram

a) Vertical View

b) Plane View



© S. St-Jean

Figure 9-15: Modular mesocosm flow schematic

9.3.7 Methods for Test Initiation, Cage Deployment and Retrieval and Test Termination

9.3.7.1 Test Initiation

The first step is to sort all bivalves into the desired size range(s). By selecting bivalves within a narrow size/age range, the investigator can be relatively confident that the individuals will have

similar growth potential. Determining the age of wild bivalves may be difficult if not impossible for some species; length is therefore used to obtain individuals with similar growth rates. Commercial growers can often provide bivalves of a known age. The size range selected for test organisms depends on the species, organism supply, and target age. Test organisms should be selected within a narrow size range, and should be kept cool and moist during the sorting stage to prevent stress, damage or death (see section 9.3.4.8). For growth and survival studies, shell lengths, widths, heights and whole-animal wet weight (WAWW) are recorded for each animal and the animals assigned to individual holding containers (e.g., ice cube tray) that are labelled according to their specific location within the cage. This allows measurements of each individual test organism to be repeated at test termination. Animals in a sacrificed sub-sample (statistically appropriate number) of test organisms will be measured for shell length, WAWW, tissue weight and shell weight. For other parameters, mussels of similar size do not need to be measured; they can be placed in the socks, preferably in clumps of 3-4 individuals, as this mimics their natural clumping behaviour and minimizes stress.

For quality control purposes, approximately 20% of the measurements should be repeated and recorded by a different investigator. It is best to use an electronic spreadsheet for recording all measurements in order to reduce transcription errors. It is possible to electronically link measurement tools (e.g., vernier calipers and analytical balance) with a spreadsheet to provide an additional quality control measure.

All test organisms needed to fill the cages, plus a sub-sample for destructive sampling (tissue weight), should be sorted and measured prior to distribution among socks and cages. Test organisms are then distributed among individual mussel socks before they are attached to cages. An example of a labelling and distribution scheme is provided in Salazar and Salazar (2000), section 11.

9.3.7.2 Cage Deployment

Ensure that an appropriate vessel is chartered to assure safe transport and deployment of all cages, moorings and floats. Moorings and floats may be attached to the cages on the vessel while en route to the site. When on-site, cage location should be identified using a global positioning system (GPS) or other reliable method (e.g., nearby onshore reference points). During the exposure period, the cages may be inspected (e.g., by divers) to check for presence, damage and fouling.

9.3.7.3 Cage Retrieval and Test Termination

Cages should be retrieved using various location identifiers (e.g., GPS, depth sounders) and a grappling hook if a retrieval rope was used for each cage.

Details on test termination are provided in Salazar and Salazar (2000), section 11, and outlined here. At the processing site, bivalves from all bags for each cage should be processed together. For growth and survival measurements, it is essential that the order and orientation of each bivalve be maintained during all of the end-of-test measurements. Individuals can be removed

from mesh socks and placed in labelled individual containers (e.g., ice cube trays) to facilitate measurement.

Five to 10 minutes prior to making length and weight measurements, set the tray(s) into a tub containing clean water. Individuals that float indicate that air is trapped between the valves of the shell. When all animals in the tray(s) have opened their valves slightly and are no longer floating (5-10 minutes), length and weight measurements may begin. Start with WAWW and shell length measurements, then carefully shuck each individual and blot the tissue before taking tissue fresh-weight measurements. Note that if tissues are to be analyzed for chemical parameters, care must be taken to not introduce any contaminants from the blotting material. The Mussel Watch protocol (Gulf of Maine Council, 1997) is best suited to dissection for the purpose of chemical analysis.

9.3.8 Effect Indicators

The effect indicators to be measured in caged bivalve studies for EEM studies are survival, growth, condition and reproduction, as described in more detail below. Table 9-7 shows the effect and supporting endpoints used for a caged bivalve study. The statistical procedures are also listed and described in more detail in section 9.3.10. An example of a reporting format for recording survival and growth raw data and endpoints is provided in Table 9-8.

Table 9-7: Caged bivalve study effect indicators and endpoints and related statistical procedures

Effect Indicators	Effect and Supporting Endpoints	Statistical Procedure
Growth	• Change in size (weight and length) over time (end measurements compared to starting measurements) • WAWW • Shell length and width • Soft tissue fresh weight	ANOVA (regression analysis for gradient designs)
Reproduction	• *Mantle somatic index (MSI)^y (similar to the GSI) (gonad weight against body weight)	ANOVA or ANCOVA
Condition	• *Weight (whole-animal dry weight, dry shell or soft tissue weight) related to shell length • Soft tissue weight related to shell weight • Soft tissue weight related to shell volume	ANOVA or ANCOVA
Survival	• *Percentage of individual animals alive per cage at the end of the exposure period • Length frequency analysis	ANOVA

* Caged bivalve effect endpoints used for determining effects. Other supporting endpoints can be used to support analyses.

^y Currently, guidance is only provided on the relative mantle somatic index (similar to the gonad index [GSI]). See section 9.3.8.5.

9.3.8.1 Survival

Survival is not a particularly sensitive indicator of effects in caged bivalves, but it is an important parameter to monitor. Survival can be easily determined and quantified, although it is possible to have some individuals missing at the test end due to shell decomposition. Bivalves are dead if they are gaping open and do not close their shells when touched or tapped. Survival is expressed as percent of individual animals alive per cage at the end of the exposure period.

9.3.8.2 Growth

Growth is a measure of energy use and is a sensitive indicator of effects that is easy to measure. Several types of growth measurements should be made; measuring only one could provide misleading results. Growth measurements, with their expected accuracy, are measured at test initiation and termination as outlined below:

WAWW	± 0.001 g
Shell length	± 0.01 mm
Shell width	± 0.01 mm
Shell height	± 0.01 mm

Growth can be expressed in a number of ways:

- absolute growth = absolute change in value from test initiation to test termination;
- growth rate = absolute change in value per unit of time, typically using one week as the time increment; or
- relative growth = (final weight – initial weight) / initial weight; relative growth may be used when there is a significant difference among cages in initial weights; relative growth is expressed as a proportion and therefore an arcsin square root transformation of relative growth values may be appropriate prior to applying statistics. Green (1979) provides useful advice on this transformation, noting that it is not usually required for proportion data in the range of 0.3-0.7, and that while it may not always help, it probably does no harm, either.

Use the most appropriate expression of growth to suit the study design and site-specific characteristics.

9.3.8.3 Condition

Condition is a measure of how an animal stores its energy, and it can be measured in both adults and juvenile mussels. There is more than one option that may be considered for calculating condition, as described below. Note that some of these methods require measurement of variables that are in addition to those outlined in section 9.3.8.2. The most appropriate method to calculate condition is left to the discretion of the investigators.

Weight (whole-animal dry weight, dry shell or soft tissue weight) related to shell length: This is analogous to the Fulton Condition Index (Ricker 1975; Anderson and Neumann 1996) used in fisheries biology. This relationship may be characterized according to a conventional formula for a straight line (e.g., in Mackie and Flippance 1983), with slope (C) and intercept (b):

$$\log W = b + (C \times \log L)$$

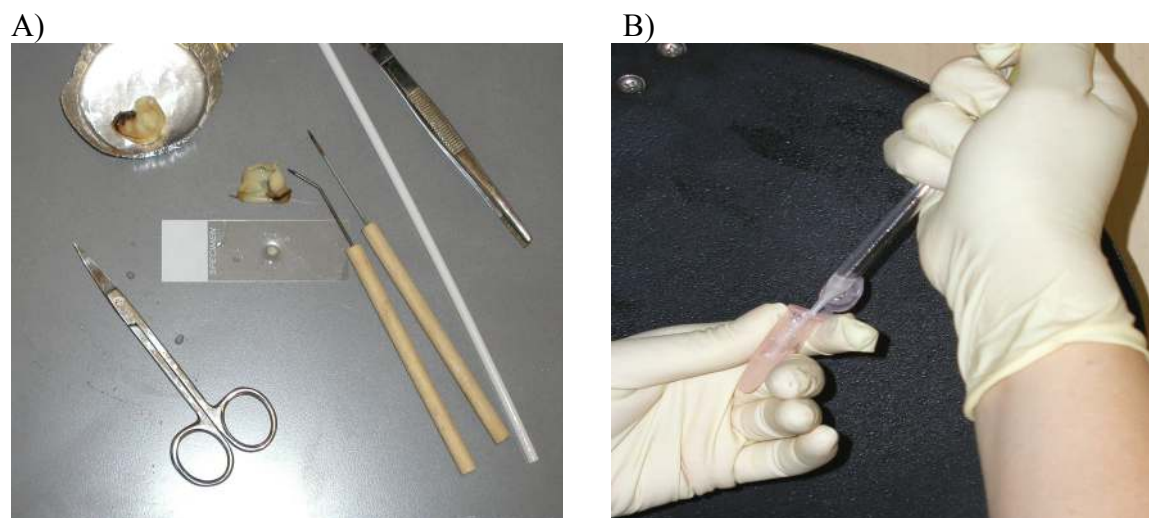
High values of C imply that a bivalve has a relatively high tissue weight at a given length, whereas low values may indicate that an animal is not obtaining sufficient food or is experiencing chronic stress that prevents it from thriving. This method of characterizing condition is suitable for assessing condition in wild bivalves where shell length is expected to be quite variable. However, since animals used for caged bivalve studies are screened for uniform length at test initiation, this method will not be reliable. This method may also not be suitable for bivalves because shell length and tissue weight are influenced by different factors in the environment (Salazar, personal communication).

Soft tissue weight related to shell weight: This method of characterizing condition uses soft tissue weight and shell weight. An ANOVA can be conducted or, more simply, soft tissue weight can be divided by shell weight. Grout and Levings (2000) measured the condition of Blue Mussels as the ratio of tissue weight to shell weight. They found that condition distinguished caged mussels in a high survival zone (condition index 1.10 to 1.42) from caged mussels in a low survival zone (condition index 0.82 to 0.96).

Soft tissue weight related to shell volume: This method was used by Mucklow (1996; based on Seed 1968) to calculate condition by dividing soft tissue dry weight by shell volume, measured as length x width x height. In a study on wild Blue Mussel populations, Mucklow (1996) concluded that seasonal patterns in condition index were variable and influenced by a number of natural factors, including food availability and physiological energy requirements.

9.3.8.4 Energy Measure

As seen in Figure 9-11, energy accumulation also occurs in the mantle, and has an annual cycle. Generally, Blue Mussels will reach their maximal energy content in late fall. When the mantle is at that stage, most of the weight consists of stored glycogen that will be used later for reproduction. Prior to dissection, mussels should be assessed for WAWW and shell measurements (length, weight, height and internal scarring). Mantle lobes should be separated from the body and weighed (mantle wet weight), after which the weight of the remainder of the body should be added to determine the body wet weight. Samples should be dried at 55°C until a constant weight is reached (approximately 2-3 days). Mantle dry weight and body dry weight should both be measured, from which the LSI-like measure can be derived. The bivalve LSI consists of the ratio of the dry weight of the mantle to the dry weight of the animal.



© S. St-Jean

Figure 9-16: A) Mantle plug and tools necessary for its removal; B) Mantle plug after homogenization and ready for assessment

9.3.8.5 Reproductive Effort

Recent research has found that the MSI for mussels, similar to the GSI in fish, can be used to determine reproductive investment. Most of the gametes produced by the mussel are stored in the mantle lobes. The following is a summary of steps to determine the MSI of mussels:

1. Measure the mussel length, width, height and the whole animal weight.
2. Dissect each mussel.
3. Determine the sex.
4. Remove and weigh each mantle lobe and determine the body mass of each individual.
5. If female: one mantle lobe should be used for dry weight and calculation of GSI ratio (ratio of body weight [minus gonads] to gonad). If male, both lobes are used. Female calculation should be extrapolated for both lobes.
6. The second lobe should be used to assess reproductive effort, through egg measurement and count.
7. Calculate the MSI (ratio of body weight [minus gonads] to gonad).

The MSI should be determined when 90% of the mantle lobe consists of gonads. There are numerous factors to consider that will affect the time of spawning: water depth, seasonal temperature, response patterns, and prevalence of different species. Boudreau et al. (in preparation), and St-Jean et al. (2008) conducted studies using MSI on the Atlantic and Pacific coasts of Canada, using different species. They found that Bay Mussels from the Pacific coast did not have a characteristic peak in gamete production preceding spawning, unlike the Blue Mussels from the Atlantic coast. Boudreau et al. (in preparation) also assessed reproductive effort by calculating and weighing the number of eggs in each mantle.

For more detailed information on the MSI, please refer to St-Jean (2003), or contact a regional EEM program coordinator for further information and complete techniques.

Briefly, the supplemental technique consists of the following (Fig. 9-16):

1. Weigh gonad lobe on analytical balance.
2. Extract a small plug of tissue using the straws provided.
3. Weigh the plug.
4. Mince the plug.
5. Using the microscope ocular grid at a magnification of 400x, measure 20 eggs (measurements are of the whole egg and nucleus).
6. Count representative sub-samples in a counting chamber.
7. Repeat Step 6, 2 more times, for a total of 3 counts.

Table 9-8: An example of a field data sheet for recording survival and growth raw data***Caged Bivalve Field Data Sheet**

Site Station Biologist Comments:				Date received Date processed Chemical analysis required? Date shipped Sample weight (chemistry)			Storage To:	
Animal No.	Length (L) (mm)	Height (H) (mm)	Width (W) (mm)	WAWW (g)	Mantle wet weight (WW) (g)	Whole - Animal Dry Weight (g)	Mantle dry weight (DW) (g)	Dry Shell Weight (g)
1								
2								
3								
4								
5								
6								
7								
8								
9								
10								
11								
12								
13								
14								
15								
16								
17								
18								
19								
20								
QA/QC								

Number of mussels alive at end of test _____ Percent survival _____

Site Conditions:

Date	Time	Temperature (°C)	Dissolved Oxygen (mg/L)	Salinity (parts per thousand)	Current Velocity (cm/sec)	Current Direction (degrees)

Field Staff: _____**Notes:**

* Data sheet provided by SSJ Environmental Limited

9.3.9 Quality Assurance and Quality Control

All work should be performed by suitably qualified and trained staff (biologists and technicians). Where contractors are used they should be selected for their specialist expertise. All fieldwork should be carried out following standard operating procedures to ensure overall consistency and that appropriate procedures are followed. All field and laboratory measurements should be made using properly calibrated instruments. All field data should be recorded using standard forms to ensure that all of the required data are collected in a reproducible and standardized format.

Replicate measures should be taken on 20% of all measures in order to verify the accuracy and reproducibility of both field measurements and laboratory analyses. For mussels in socks, this represents at least two mussels per sock.

In data analysis, the first step should be the screening of the data for outliers. A rapid way to screen for outliers is to create scatter plots of pairs of variables with 95% confidence ellipses superimposed. Potential outlier data points can then be identified as those that lie outside the confidence ellipses. Outliers can be the result of a number of causes, including data entry or transcription errors. Where outliers are detected, the data records should be reviewed in order to isolate and if possible correct the source of a potential error. Where no such identification is possible, the analysis should be performed both with and without the outlier, in order to evaluate the influence that the outlier exerts on the results of the data analysis.

Statistical data will be examined to evaluate the degree to which they conform to the underlying assumptions of the analysis (such as normality and homogeneity of variance, or equality of slopes in ANCOVA). Where appropriate, transformations may be applied in order to lessen the magnitude of violations of the underlying assumptions.

9.3.10 Data Analyses

Data analyses and interpretation of the results should be appropriate to the study design. Caged mussel studies for EEM are designed to determine if there are significant effects on biota in the

vicinity of effluent outfalls. This can be accomplished by using a C-I (= reference – exposure) or a gradient (= regression) type of design. Statistical procedures appropriate to each effect indicator (i.e., survival, growth, reproduction, energy and condition) are summarized in Table 9-7.

An ANCOVA should be performed to test the GSI (dry gonad weight), condition (dry body weight) with covariates to remove influences including dry body weight for GSI, and shell length for condition. Where an interaction between treatment and covariate precludes the use of an ANCOVA, stratified subsets of the covariate should be compared in a single-factor ANOVA. When the control groups are not significantly different, they should be grouped for the analyses. However, if a significant difference is found between groups, all controls should be included in the analyses. A Tukey multiple-comparison test can be employed when significant differences are found among groups. Non-normality (probability plot) or heteroscedasticity (Fmax test) that cannot be resolved by appropriate data transformations should be followed by non-parametric analysis using the Kruskal-Wallis test, followed by a Noether multiple-comparison test (Scherrer 1984; Zar 1999). Probit analysis can be employed for survival. The level of significance should be set at $p < 0.05$, and back-transformed means should be accompanied by their 95% confidence interval.

The first step is to generate summary statistics for each parameter (i.e., WAWW, shell length, soft tissue fresh weight) and each cage and station.

The second step is to determine whether there are significant differences among replicate cages for each of the parameters measured before deployment (if they were not measured post-distribution and before deployment). This involves assessing the data for normality and homogeneity of variances.

The final step is to use the appropriate statistical test for the study design. In general, ANOVA and multiple-comparison tests are used for hypothesis testing and comparison among stations. For C-I and MC-I designs using ANOVA and ANCOVA procedures, detailed guidance is provided in Chapter 8 of this guidance document. If statistical differences are found, a multiple-range test, or its non-parametric counterpart, can be used to determine which stations are different from the others. Linear and multiple-regression analyses (using variables regressed on distance) may generally be used to establish relationships among variables along exposure gradients. For ANCOVA analyses of condition, the covariates will be dictated by the condition formula that is used.

9.3.11 Mercury in Fish Tissue

If a mine is required to measure mercury in fish tissue, caged bivalves may be considered. However, there are some precautions that should be taken with this approach, including the following:

- Duration of exposure should be chosen such that mercury is likely to be accumulated to detectable levels in bivalve tissue.

- Consider whether bivalves would be harvested (commercially or recreationally) for human consumption in the area.
- Ensure that the number of bivalves included in the design is sufficient to obtain enough tissue for analysis.

There may be mines for which caged bivalves would not be considered suitable for measuring uptake of mercury. Any proposal to use caged bivalves for mercury measurements would require approval by the Regional Authorization Officer.

9.3.12 Reporting

QA/QC considerations for caged bivalve studies should follow those outlined for the fish survey in Chapter 3 of this guidance document. QA/QC measures apply to the following components of caged bivalve studies:

- study design;
- field sampling;
- sample processing / laboratory analysis;
- data analysis; and
- reporting.

9.4 References

Anderson RO, Neumann RM. 1996. Length, weight, and associated structural indices. In Murphy BR, Willis DW, editors. Fisheries techniques. 2nd edition. Bethesda (MD): American Fisheries Society. p. 447-482.

Andrews S, Parker R. 1999. Draft Report (II). An environmental quality evaluation of Pictou Harbour, Nova Scotia, using caged bivalves; *Mytilus edulis*. Prepared by Environmental Protection Branch of Environment Canada, Dartmouth (NS) for the Pictou Harbour Environmental Protection Project, New Glasgow (NS).

Ankley GT, Jensen KM, Kahl MD, Korte JJ, Makynen EA. 2001. Description and evaluation of a short-term reproduction test with the fathead minnow (*Pimephales promelas*). Environ Toxicol Chem 20:1276-1290.

Ankley GT, Jensen KM, Makynen EA, Kahl MD, Korte JJ, Hornung MW, Henry TR, Denny JS, Leino RL, Wilson VS, et al. 2003. Effects of the androgenic growth promoter 17-beta-trenbolone on fecundity and reproductive endocrinology of the fathead minnow. Environ Toxicol Chem 22:1350-1360.

Applied Biomonitoring. 2000. Caged Mussel Pilot Study Port Alice Mill, Vancouver Island. EEM Program. Environment Canada, Environmental Protection Branch, Pacific and Yukon Region. Regional Manuscript Report MS 00-01.

Atz JW. 1986. *Fundulus heteroclitus* in the laboratory: a history. Amer Zool 26:111-120.

Barber WE, Kevern NR. 1974. Seasonal variation of sieving efficiency in a lotic habitat. Freshwat Biol 4:293-300.

Bauer G. 1994. The adaptive value of offspring size among freshwater mussels (Bivalvia; Unionoidea). J Anim Ecol 63:933-944.

Beckvar N, Salazar S, Salazar M, Finkelstein K. 2000. An in situ assessment of mercury contamination in the Sudbury River, Massachusetts, using transplanted freshwater mussels (*Elliptio complanata*). Can J Fish Aquat Sci 57:1103-1112.

Bothwell ML. 1985. Phosphorus limitation of lotic periphyton growth rates: an intersite comparison using continuous-flow troughs (Thompson River system, British Columbia). Limnol Oceanogr 30:527-542.

Bothwell ML, Daley RJ. 1981. Determination of phosphorus sufficiency and growth rates of benthic algae in the Thompson River, BC, using experimental flowing troughs. IWD Regional Report. Vancouver (BC): Inland Waters Directorate, Environment Canada.

Bothwell ML, Derksen G, Nordin RN, Culp JM. 1992. Nutrient and grazer control of algal biomass in the Thompson River, British Columbia: a case history of water quality management,

In Robarts RD, Bothwell ML, editors. Aquatic ecosystems in semi-arid regions: implications for resource management. (National Hydrology Research Institute Symposium Series 7). Saskatoon (SK): Environment Canada. p. 253-266.

Bothwell ML, Culp JM. 1993. Sensitivity of the Thompson River to phosphorus: studies on trophic dynamics. NHRI Contribution No. 93006. Saskatoon (SK): National Hydrology Research Institute, Environment Canada.

Burt A, Kramer DL, Nakasuru K, Spry C. 1988. The tempo of reproduction in *Hyphessobrycon pulchripinnis* (Characidae), with a discussion on the biology of “multiple spawning”. Environ Biol Fish 22:15-27.

Carlisle DM. 2000. Bioenergetic food webs as a means of linking toxicological effects across scales of ecological organization. J Aquat Eco Stress Rec 7:155-165.

Cash KJ, Culp JM, Dubé MG, Lowell RB, Glozier NE, Brua RB. 2003. Integrating mesocosm experiments with field and laboratory studies to generate weight-of-evidence risk assessments for ecosystem health. Aquatic Ecosystem Health and Management. 6: 177-183.

Chen CY, Stemberger RS, Klaue B, Blum JD, Pickhardt PC, Folt CL. 2000. Accumulation of heavy metals in food web components across a gradient of lakes. Limnol Oceanogr 45:1525-1536.

Clarke A. 1973. The freshwater molluscs of the Canadian Interior Basin. Malacologia 13:1-509.

Clarke A. 1981. The freshwater molluscs of Canada. Ottawa (ON): National Museums of Canada.

Coad BW. 1995. Encyclopedia of Canadian fishes. Waterdown (ON): Canada Museum of Nature and Canadian Sportfishing Productions Inc.

Couillard Y, Campbell PGC, Tessier A, Pellerin-Massicotte J, Auclair JC. 1995a. Field transplantation of a freshwater bivalve, *Pyganodon grandis*, across a metal contamination gradient. I. Temporal changes in metallothionein and metal (Cd, Cu, and Zn) concentrations in soft tissues Can J Fish Aquat Sci 52:690-702.

Couillard Y, Campbell PGC, Tessier A, Pellerin-Massicotte J, Auclair JC. 1995b. Field transplantation of a freshwater bivalve, *Pyganodon grandis*, across a metal contamination gradient. II. Metallothionein response to Cd and Zn exposure, evidence for cytotoxicity, and links to effects at higher levels of biological organization. Can J Fish Aquat Sci 52:703-715.

Courtenay SC, Parker WR, Rawn GP. 1998. Proceedings of a workshop to assess alternatives to the fish survey component of the environmental effects monitoring program for Canadian pulp and paper mills. (Canadian Technical Report of Fisheries and Aquatic Sciences 2233).

Crane M, Burton GA, Culp JM, Greenberg MS, Munkittrick KR, Ribeiro R, Salazar M, St-Jean S. 2007. In situ approaches for stressor and effect diagnosis. *Integr Environ Assess Manag* 3:234-245.

Culp JM, Cash KJ. 1995. Potential impacts of effluent on the Thompson and upper Fraser Rivers. DOE FRAP 1995-10. Vancouver (BC): Fraser River Action Plan, Environmental Conservation Branch, Environment Canada.

Culp JM, Podemski CL. 1996. Design and application of a novel stream microcosm system for assessing effluent impacts to large rivers. In Servos MR, Munkittrick KR, Carey JH, Van Der Kraak GJ, editors. *Environmental fate and effects of pulp and paper mill effluents*. Delray Beach (FL): St. Lucie Press. p. 549-555.

Culp JM, Podemski CL, Cash KJ, Lowell RB. 1996. Utility of field-based artificial streams for assessing effluent effects on riverine ecosystems. *J Aquat Ecosystem Health* 5:117-124.

Culp JM, Lowell RB. 1998. Pulp mill effluent impacts on benthic communities and selected fish species in the Fraser River Basin. In Gray C, Tuominen T, editors. *Health of the Fraser River aquatic ecosystem*, Vol. II. DOE FRAP 1998-11. Ottawa (ON): Environment Canada. p. 13-34.

Culp JM, Lowell RB, Cash KJ. 2000a. Integrating mesocosm experiments with field and laboratory studies to generate weight-of-evidence risk assessments for large rivers. *Environ Toxicol Chem* 19:1167-1173.

Culp JM, Podemski CL, Cash KJ, Lowell RB. 2000b. A research strategy for using stream microcosms in ecotoxicology: integrating experiments at different levels of biological organization with field data. *J Aquat Eco Stress Rec* 7:167-176.

Culp JM, Podemski CL, Cash KJ. 2001. Interactive effects of nutrients and contaminants from pulp mill effluents on riverine benthos. *J Aquat Eco Stress Rec* 8(1):67-75.

Culp JM, Glozier NE, Cash KJ, Dubé MG, Waiser M, MacLatchy DL, Brua B, Firth B, Wilson G. 2004. Cumulative effects investigation of pulp mill and sewage effluent impacts on benthic food webs: a mesocosm example. In Borton DL, Hall TJ, Fisher RP and Thomas JF, editors. *Pulp and paper mill effluent environmental fate and effects, proceedings of the Fifth International Conference on Fate and Effects of Pulp and Paper Mill Effluents*, June 1-4, 2003, Seattle, Washington, USA. Lancaster (PA): DEStech Publications. p. 464-472.

Day KE, Metcalfe JL, Batchelor SP. 1990. Changes in intracellular free amino acids in tissues of the caged mussel (*Elliptio complanata*), exposed to contaminated environments. *Arch Environ Cont Toxicol* 19:816-827.

Drake JA, Huxel GR, Hewitt CL. 1996. Microcosms as models for generating and testing community theory. *Ecology* 77:670-677.

Dubé MG. 1995. Effects of bleached kraft pulp mill effluent on periphyton and chironomid abundance in the Thompson River, British Columbia [master's thesis]. Saskatoon (SK): University of Saskatchewan.

Dubé MG. 2000. Sources of contaminants in the bleached kraft pulping process and their effect on the mummichog (*Fundulus heteroclitus*) [dissertation]. Saint John (NB): University of New Brunswick.

Dubé MG, Culp JM. 1996. Growth responses of periphyton and chironomids exposed to biologically treated bleached kraft pulp mill effluent. *Environ Toxicol Chem* 15:2019-2027.

Dubé MG, Culp JM, Scrimgeour GJ. 1997. Nutrient limitation and herbivory: processes influenced by bleached kraft pulp mill effluent. *Can J Fish Aquat Sci* 54:2584-2595.

Dubé MG, MacLatchy DL. 2000a. Endocrine responses of an estuarine fish *Fundulus heteroclitus* to final effluent from a bleached kraft pulp mill before and after installation of reverse osmosis treatment of a waste stream. *Environ Toxicol Chem* 19:2788-2796.

Dubé MG, MacLatchy DL. 2000b. Reverse osmosis treatment of condensates from a bleached kraft pulp mill: effects on acute and chronic toxicity of process streams and final effluent. In Ruoppa M, Paasivirta J, Lehtinen K-J, Ruonala S, editors. *Proceedings of the Fourth International Conference on Environmental Impacts of the Pulp and Paper Industry*. Helsinki (FI). p. 270-276.

Dubé MG, MacLatchy DL. 2001. Identification and treatment of a waste stream at a bleached kraft pulp mill that depresses a sex steroid in the mummichog (*Fundulus heteroclitus*). *Environ Toxicol Chem* 20:985-995.

Dubé MG, Culp JM, Cash KJ, Glozier NE, MacLatchy DL, Podemski CL, Lowell RB. 2001. Artificial streams for environmental effects monitoring (EEM): development and application in Canada over the past decade. *Wat Qual Res J Can* 37(1):155-180.

Dubé MG, Culp JC, MacLatchy DL, Gillis G, Parker R, Courtenay S, Gilman G. 2002. Utility of mobile, field-based artificial streams for assessing effects of pulp mill effluents on fish in the Canadian Environmental Effects Monitoring (EEM) program. *J Aquat Eco Stress Rec* 9:85-102.

Dubé MG, MacLatchy DL, Firth BK, Culp JM, Glozier NE, Cash KJ. 2004. Using mesocosms to explore confounding factors influencing longnose dace (*Rhinichthys cataractae*) responses to kraft mill effluent on the Wapiti River, AB, Canada. In Borton DL, Hall TJ, Fisher RP, Thomas JF, editors. *Pulp and paper mill effluent environmental fate and effects*, Proceedings of the Fifth International Conference on Fate and Effects of Pulp and Paper Mill Effluents, June 1-4, 2003, Seattle, Washington, USA. Lancaster (PA): DEStech Publications. p. 481-490.

Dubé MG, MacLatchy DL, Kieffer JD, Glozier NE, Culp JM, Cash KJ. 2005. Effects of metal mining effluent on Atlantic salmon (*Salmo salar*) and slimy sculpin (*Cottus cognatus*): using

mesocosms to assess existing effects and predict future consequences. *Sci Total Environ* 343:135-154.

Dubé MG, MacLatchy DL, Hruska KA, Glozier NE. 2006. Assessing the responses of creek chub (*Semotilus atromaculatus*) and pearl dace (*Semotilus margarita*) to metal mine effluents using in situ artificial streams in Sudbury, Ontario, Canada. *Environ Toxicol Chem*. 25(1):18-28.

Environment Canada. 1997. Fish Survey Working Group Final Report. Recommendations from Cycle 1 Review. Ottawa (ON): National EEM Office, Science Policy and Environmental Quality Branch, Environment Canada. EEM/1997/6.

Federal-Provincial Thompson River Task Force. 1976. Sources and effects of algal growth, colour, foaming and fish tainting in the Thompson River system. Victoria (BC): Environmental Protection Division, British Columbia Ministry of Environment, Lands and Parks.

Freeman KR, PerryKL, DiBaccoTG, Scarratt DJ. 1994. Observations on two Mytilid species from a Nova Scotian mussel farm. (Canadian Technical Report of Fisheries and Aquatic Sciences 1969). Halifax (NS): Fisheries and Oceans Canada.

Fritz ES, Meredith WH, Lotrich VA. 1975. Fall and winter movements and activity level of the mummichog *Fundulus heteroclitus*, in a tidal creek. *Chesapeake Sci*. 16:211-215.

Gibbons WN, Munkittrick KR, Taylor WD. 1998a Suitability of small fish species for monitoring the effects of pulp mill effluent on fish. 1. Response of spoonhead sculpin exposed to bleached kraft mill effluent. *Environ Toxicol Chem* 7(11):2227-2237.

Gibbons WN, Munkittrick KR, Taylor WD. 1998b. Suitability of small fish species for monitoring the effects of pulp mill effluent on fish. 2. Comparison of trout perch responses to white sucker exposed to bleached kraft mill effluent. *Environ Toxicol Chem* 17(11):2238-2245.

Goldberg ED, Bowen VT, Farrington JW, Harvey G, Martin JH, Parker PL, Risebrough RW, Robertson W, Schneider E, Gamble E. 1978. The Mussel Watch. *Environ Conserv* 5(2):101-125.

Gray C, Tuominen T. 1998. Health of the Fraser River aquatic ecosystem, Vol. I. DOE FRAP 1998-11. Vancouver (BC): Environment Canada.

Green RH. 1979. Sampling design and statistical methods for environmental biologists. Toronto (ON): John Wiley & Sons.

Grout JA, Levings CD. 2000. Effects of acid mine drainage from an abandoned copper mine, Britannia Mines, Howe Sound, British Columbia, Canada, on transplanted blue mussels (*Mytilus edulis*). *Mar Environ Res* 51(3):265-288.

Gulf of Maine Council on the Marine Environment. 1997. Gulfwatch Project: standard procedures: field and laboratory. Gulfwatch implementation period 1993-2001. August 1997.

Harries JE, Runnals T, Hill E, Harris CA, Maddix S, Sumpter JP, Tyler CR. 2000. Development of a reproductive performance test for endocrine disrupting chemicals using pair breeding fathead minnows (*Pimephales promelas*). Environ Sci Technol 34:3003-3011.

Heins DC, Rabito FG Jr. 1986. Spawning performance in North American minnows: direct evidence of the occurrence of multiple clutches in the genus *Notropis*. J Fish Biol 28:343-357.

Hinch SG, Kelly LJ, Green RH. 1989. Morphological variation of *Elliptio complanata* (Bivalvia: Unionidae) in differing substrata of soft-water lakes exposed to acidic deposition. Can J Zool 67:1895-1899.

Hornbach DJ, Wissing TH, Burky AJ. 1982. Life-history characteristics of a stream population of the freshwater clam *Sphaerium striatinum* Lamarck (Bivalvia: Pisidiidae). Can J Zool. 60:249-260.

Hruska K, Dubé MG. 2004. Using artificial streams to assess the effects of metal-mining effluent on the life-cycle of the freshwater midge (*Chironomus tentans*) in situ. Environ Toxicol Chem 23(11):2709-2718.

Hruska K, Dubé M. 2005. Comparison of a partial life cycle bioassay in artificial streams to a standard beaker bioassay to assess effects of metal mine effluent on *Chironomus tentans*. Environ Toxicol Chem 24(9):2325-2335.

Hurlbert SH. 1984. Pseudoreplication and the design of ecological field experiments. Ecol Monog 54:187-211.

Jadhav ML, Lomte VS. 1982. Seasonal variation in biochemical composition of the freshwater bivalve, *Lamellidens corrianus*. Rivista di Idrobiologia. 21:1-7.

Jenkin, RE, Burkhead NM. 1993. Freshwater fishes of Virginia. Bethesda (MD): American Fisheries Society.

Jensen KM, Korte JJ, Kahl MD, Pasha MS, Ankley GT. 2001. Aspects of basic reproductive biology and endocrinology in the fathead minnow (*Pimephales promelas*). Comp Biochem Physiol Part C. 128:127-141.

Jensen KM, Kahl MD, Makynen EA, Korte JJ, Leino RL, Butterworth BC, Ankley GT. 2004. Characterization of responses to the antiandrogen flutamide in a short-term reproduction assay with the fathead minnow. Aquat Toxicol 70:99-110.

Jernelov A. 1996. The International Mussel Watch: A global assessment of environmental levels of chemical contaminants. Sci Tot Environ 188 (Suppl. 1): 37-44.

Kneib RT. 1986. The role of *Fundulus heteroclitus* in salt marsh trophic dynamics. Amer Zool 26:259-269.

Kneib RT, Stiven AE. 1978. Growth, reproduction, and feeding of *Fundulus heteroclitus* (L.) on a North Carolina salt marsh. J Exp Mar Biol Ecol 31:121-140.

Leblanc J, Couillard CM. 1995. Description de la période de reproduction d'un poisson sentinelle: le choquemort (*Fundulus heteroclitus*) de l'estuaire de la Miramichi. (Rapport technique canadien des sciences halieutiques et aquatiques 2057). Mont-Joli (QC): Fisheries and Oceans Canada.

Lowell RB, Culp JM, Wrona FJ. 1995. Stimulation of increased short-term growth and development of mayflies by pulp mill effluent. Environ Toxicol Chem 14:1529-1541.

Lowell RB, Culp JM, Wrona FJ, Bothwell ML. 1996. Effects of pulp mill effluent on benthic freshwater invertebrates: food availability and stimulation of increased growth and development. In Servos MR, Munkittrick KR, Carey JH, Van Der Kraak GJ, editors. Environmental fate and effects of pulp and paper mill effluents. Delay Beach (FL): St. Lucie Press. p. 525-532.

Lowell RB, Culp JM, Dubé MG. 2000. A weight-of-evidence approach for northern river risk assessment: integrating the effects of multiple stressors. Environ Toxicol Chem 19:1182-1190.

Mackie GL. 1978a. Shell structure in freshwater Sphaeriaceae. (Bivalvia: Heterondonta). Can J Zool 56:1-6.

Mackie GL. 1978b. Are sphaeriid clams ovoviviparous or viviparous? Nautilus 92:145-147.

Mackie GL, Quadri SU, Clarke AH. 1974. Development of brood sacs in *Musculium securis* Bivalvia: Sphaeriidae. Nautilus 88:109-111.

Mackie GL, Flippance LA. 1983. Growth dynamics in *Sphaerium rhomboideum* (Bivalvia: Pisidiidae). Can J Zool 61:868-873.

Mallet AL, Carver CE. 1995. Comparative growth and survival patterns of *Mytilus trossulus* and *Mytilus edulis* in Atlantic Canada. Can J Fish Aquat Sci 52:1873-1880.

Malley DF, Stewart AR, Hall BD. 1996. Uptake of methyl mercury by the floater mussel, *Pyganodon grandis* (Bivalvia, Unionidae), caged in a flooded wetland. Environ Toxicol Chem 15:928-936.

Martel P, Kovacs T, Voss R. 2003. Survey of pulp and paper mill effluents for their potential to affect fish reproduction. Proceedings of the Fifth International Conference on the Fate and Effects of Pulp and Paper Mill Effluents, Seattle, Washington, USA, June 1-4, 2003. p. 78-91.

Mason RP, Laporte JM, Andres S. 2000. Factors controlling the bioaccumulation of mercury, methylmercury, arsenic, selenium and cadmium by freshwater invertebrates and fish. Arch Environ Contam Toxicol 38:283-297.

McDonald JH, Seed R, Koehn RK. 1991. Allozymes and morphometric characteristics of three species of *Mytilus* in the northern and southern hemispheres. Mar Biol 11:323-333.

McGreer ER, Belzer W. 1998. Contaminant sources. In Gray C, Tuominen T, editors. Health of the Fraser River aquatic ecosystem, Vol. I. DOE FRAP 1998-11. Ottawa (ON): Environment Canada. p. 7-22.

McMahon RF. 1991. Mollusca: bivalvia. In Thorp JH, Covich AP, editors. 1991. Ecology and classification of North American freshwater invertebrates. San Diego (CA): Academic Press. p. 315-399.

Metcalf-Smith JL, Green RH, Grapentine LC. 1996. Influence of biological factors on concentrations of metals in the tissues of freshwater mussels (*Elliptio complanata* and *Lampsilis radiata radiata*) from the St. Lawrence River. Can J Fish Aquat Sci 53:205-219.

Mucklow LC. 1996. Effects of season and species on physiological condition and contaminant burdens in mussels (*Mytilus edulis* L. and *Mytilus trossulus* G.): implications for mussel watch programs [master's thesis]. Halifax (NS): School for Resource and Environmental Studies, Dalhousie University.

Nelson JS, Paetz MJ. 1992. The fishes of Alberta. Calgary (AB): University of Calgary Press.

Newell RIE. 1989. Species profile of the blue mussel, *Mytilus edulis*. US Fish Wildl Ser Biol Rep 82(11):40.

Ni IH, Wang WX, Tam YK. 2000. Transfer of Cd, Cr and Zn from zooplankton prey to mudskipper *Periophthalmus cantonensis* and glassy *Ambassis urotaenia* fishes. Mar Ecol-Prog Ser 194:203-210.

O'Connor. 1992. Recent trends in coastal environmental quality: Results from the first five years of the NOAA Mussel Watch Project. Washington (DC): U.S. Department of Commerce, National Oceanic and Atmospheric Administration (NOAA), and National Ocean Service.

[OECD] Organisation for Economic Co-operation and Development. 2001. Detailed review paper: appraisal of test methods for sex hormone disrupting chemicals. (OECD Series on testing and assessment 21). ENV/JM/MONO(2002)8.

[OECD] Organisation for Economic Co-operation and Development. 2006. Report of the validation of the 21-day fish screening assay for the detection of endocrine substances (Phase 1B). (OECD Series on testing and assessment 61). ENV/JM/MONO(2006)29.

Paine MD. 1990. Life history tactics of darters (Percidae: Etheostomatiini) and their relationship with body size, reproductive behaviour, latitude and rarity. J Fish Biol 37:473-488.

Paine MD. 2000. EEM Design. Short-Course presented at the 27th Annual Aquatic Toxicity Workshop, October 1-4th, St. John's, NF.

Panter GH, Hutchinson TH, Lange R, Lye CM, Sumpter JP, Zerulla M, Tyler CR. 2002. Utility of a juvenile fathead minnow screening assay for detecting (anti-)estrogenic substances. *Environ Toxicol Chem* 21(2):319-326.

Parker R, Smith N. 1997. A synopsis of the results of environmental effects monitoring studies at 19 pulp and paper mills in Atlantic Canada. In *Proceedings of the Third International Conference on Environmental Effects of Pulp and Paper Mill Effluents*. Rotorua, NZ. p. 432-441.

Parrott JL. 2005. Overview of methodology and endpoints in fathead minnow lifecycle tests assessing pulp and paper mill effluents. *Water Qual Res Can* 40:334-346.

Parrott JL, Wood CS. 2001. Fathead minnow lifecycle tests for detection of endocrine disrupting substances in effluents. *Water Qual Res J Can* 37:651-667.

Penney RW, Hart MJ, Templeman N. 2002. Comparative growth of cultured blue mussels, *Mytilus edulis*, *M. trossulus* and their hybrids, in naturally occurring mixed-species stocks. *Aquat Res* 33:693-702.

Podemski CL. 1999. Ecological effects of a bleached kraft pulp mill effluent on benthic biota of the Athabasca River [dissertation]. Saskatoon (SK): University of Saskatchewan.

Podemski CL, Culp JM. 1996. Nutrient and contaminant effects of bleached kraft mill effluent on benthic algae and insects of the Athabasca River. In Servos MR, Munkittrick KR, Carey JH, Van Der Kraak GJ, editors. *Environmental fate and effects of pulp and paper mill effluents*. Delay Beach (FL): St. Lucie Press. p. 571-580.

Pollock MS, Fisher SE, Squires AJ, Pollock RJ, Chivers DP, Dubé MG. 2008. Relative body size influences breeding propensity in fathead minnows: implication for ecotoxicology testing procedures. *Water Qual Res J Can* 43(4):257-264.

Pollock MS, Dubé MG, LaPorte J, Clarke L, Schryer R. 2009. Investigating the link between pulp mill effluent and endocrine disruption: attempts to explain the presence of male and female hermaphrodites in the Wabigoon River, Ontario. *Environ Toxicol Chem* 29(4):952-965.

Portt CB, Coker GA, Ming DL, Randall RG. 2006. A review of fish sampling methods commonly used in Canadian freshwater habitats. (Canadian Technical Report of Fisheries and Aquatic Sciences 2604). Cat. No.Fs97-6/2604E. Ottawa (ON): Fisheries and Ocean Canada.

Ramón M, Richardson CA. 1992. Age determination and shell growth of *Chamelea gallina* (Bivalvia: Veneridae) in the western Mediterranean. *Mar Ecol Prog Ser* 89:15-23.

Rawson PD, Slaughter C, Yund PO. 2003. Patterns of gamete incompatibility between the blue mussels *Mytilus edulis* and *M. trossulus*. *Mar Biol* 143:317-325.

Rees HL. 1984. A note on mesh selection and sampling efficiency in benthic studies. *Mar Poll*

Bull 15:225-229.

Resh VH, McElravy EP. 1993. Contemporary quantitative approaches to biomonitoring using benthic macroinvertebrates. In Rosenberg DM, Resh VH, editors. Freshwater biomonitoring and benthic invertebrates. New York (NY): Chapman & Hall. p. 159-194.

Ricker WE. 1975. Computation and interpretation of biological statistics of fish populations. Bulletin 191. Ottawa (ON): Department of the Environment and Fisheries and Marine Service.

Rickwood CJ, Dubé MD, Hewitt LM, Kovacs T, Parrott JL, MacLachy DL. 2006a. Use of paired fathead minnow (*Pimephales promelas*) reproductive test: Part I: Assessing biological effects of final bleached kraft pulp mill effluent using a mobile bioassay trailer system. Environ Toxicol Chem 25:191-201.

Rickwood CJ, Dubé MD, Hewitt LM, Kovacs T, Parrott JL, MacLachy DL. 2006b. Use of paired fathead minnow (*Pimephales promelas*) reproductive test: Part II Source identification of biological effects at a bleached kraft pulp mill. Environ Toxicol Chem 25:202-211.

Rickwood CJ, Dube MG, Weber L, Driedger K, Janz DM. 2006c. Assessing effects of metal mining effluent on fathead minnow (*Pimephales promelas*) reproduction in a trophic-transfer system. Environ Sci Technol 40:6489-6497.

Rickwood CJ, Dubé MG, Weber LP, Lux S, Janz DM. 2008. Assessing effects of a mining and municipal sewage effluent mixture on fathead minnow (*Pimephales promelas*) reproduction using a novel, field-based trophic-transfer artificial stream. Aquat Toxicol 86(2):272-286.

Roberts WE. 1988. The sculpins of Alberta. Alberta Naturalist 18:121-127.

Salazar MH, Salazar SM. 1995. In situ bioassays using transplanted mussels: I. Estimating chemical exposure and bioeffects with bioaccumulation and growth. In Hughes JS, Biddinger GR, Mones E, editors. Environmental toxicology and risk assessment – Third Volume. Philadelphia (PA): American Society for Testing and Materials. p. 216-241.

Salazar M, Salazar S. 2000. Standard guide for conducting in-situ field bioassays with caged marine, estuarine and freshwater bivalves. (Draft dated November 17, 2000). Philadelphia (PA): American Society for Testing and Materials.

Sandusky MJ, Sparks RE, Paparo AA. 1979. Investigations of declines in fingernail clam (*Musculium transversum*) populations in the Illinois River and Pool 19 of the Mississippi River. Bull Amer Malacol Union 1979:11-16.

Scherrer B. 1984. Biostatistique. Québec (QC): Gaëtan Morin.

Scott WB. 1967. Freshwater fishes of eastern Canada. Toronto (ON): University of Toronto Press.

Scott WB, Crossman EJ. 1973. Freshwater fishes of Canada. (Fisheries Research Board of Canada Bulletin 184). Ottawa (ON): Fisheries Research Board of Canada.

Scott WB, Scott MG. 1988. Atlantic fishes of Canada. Can Bull Fish Aquat Sci 219:314-316.
Scrimgeour GJ, Culp JM, Glozier NE. 1993. An improved technique for sampling lotic invertebrates. Hydrobiologia 254:65-71.

Seed R. 1968. Factors influencing shell shape in *Mytilus edulis* L. J Mar Biol Ass UK 48:561-584.

Sejr M, Jensen KT, Rysgaard S. 2002. Annual growth bands in the bivalve *Hiatella arctica* validated by a mark-recapture study in NE Greenland. Polar Biol 25:794-796.

Shaw JL, Maund SJ, Marshall SJ, Hill IR. 1995a. Fathead minnow (*Pimephales promelas*) reproduction in outdoor microcosms: an assessment of the ecological effects of fish density. Environ Toxicol Chem 14(10):1763-1772.

Shaw JL, Maund SJ, Hill IR. 1995b. Fathead minnow reproduction in outdoor microcosms: a comparison to bluegill sunfish reproduction in large mesocosms. Environ Toxicol Chem 14(10):1753-1762.

Skidmore DA, Chew KK. 1985. Mussel aquaculture in Puget Sound. Seattle (WA): Washington Sea Grant Report. WSG 85-4.

Slack KV, Averett RC, Geeson PE, Lipscomb RG. 1973. Methods for collection and analysis of aquatic biological and microbiological samples. In Techniques of water-resources investigations of the United States Geological Survey, Chapter 4A, Book 5. Washington (DC): U.S. Department of the Interior, Geological Survey. p. 1-165.

Sohoni P, Typer CR, Hurd K, Caunter J, Hetheridge M, Williams T, Evans M, Toy R, Gargas M,

Sumpter JP. 2001. Reproductive effects of long-term exposure to bisphenol a in the fathead minnow (*Pimephales promelas*). Environ Sci Technol 35:2917-2925.

Stephenson M, Mackie GL. 1981. Effects of 2, 4-D treatment on growth and survival of fingernail clams (Bivalvia: Pisidiidae) in artificial pond ecosystems. Bull Amer Malacol Union 1981:19-22.

Stewart, PL. 1994. Environmental requirements of the Blue Mussel (*Mytilus edulis*) in eastern Canada and its response to human impacts. Canadian Technical Report of Fisheries and Aquatic Science, 2004.

Stewart R, Malley DF. 1997. Technical evaluation of molluscs as a biomonitoring tool for the Canadian mining industry. Aquatic Effects Technology Evaluation (AETE) Program Project 2.3.1. Sponsored by Canada Center for Mineral and Energy Technology (CANMET) and the Mining Association of Canada (MAC).

St-Jean SD. 2003. Supplemental techniques for cage bivalve. National Water Research Institute Contribution number: 03-422.

St-Jean SD, Courtenay SC, Parker RW. 2003. Immunomodulation in blue mussels (*Mytilus edulis*) exposed to a pulp and paper mill effluent in Eastern Canada. Wat Qual Res J Can 38(4):647-666.

St-Jean SD, Stephens RE, Courtenay SC, Reinisch CL. 2005. Detecting p53 family proteins in hemocytic leukemia cells of *Mytilus edulis* from Pictou Harbour, Nova Scotia. Can J Fish Aquat Sci 62:2055-2066.

St-Jean S, Courtenay S, Coomber C, Bishay F. 2008. Responses to environmental stressors of wild vs. caged mussels. EEM Science Symposium, Gatineau, Quebec. 29-30 April 2008.

Suess MJ. 1982. Examination of water for pollution control. A reference handbook. Vol. 3. Biological, bacteriological and virological examination. Oxford (UK):Permagon Press.

Toro JE, Thompson RJ, Innes DJ. 2002. Reproductive isolation and reproductive output in two sympatric mussels species (*Mytilus edulis*, *M. trossulus*) and their hybrids from Newfoundland. Mar Biol 141:897-909.

[US EPA] United States Environmental Protection Agency. 1982. User's guide for conducting life-cycle chronic toxicity tests with fathead minnows (*Pimephales promelas*). Duluth (MN): Environmental Research Laboratory. U.S. EPA/600/8-81/011.

[US EPA] United States Environmental Protection Agency. 1996. Prehatching development of the fathead minnow *Pimephales promelas*. Washington (DC). EPA/600/R-96/079.

[US EPA]. 1999. Screening level reproduction assay with the fathead minnow (*Pimephales promelas*) Duluth (MN): EPA-NEERL-DUL 3544.

[US EPA] United States Environmental Protection Agency. 2002. Comparative evaluation of fathead minnow assays for detecting endocrine disrupting chemicals. Washington (DC): Endocrine Disruptor Screening Program. EPA/68/W-01/023.

[US EPA] United States Environmental Protection Agency. 2007. Short term test method for assessing the reproductive toxicity of endocrine-disrupting chemicals using the fathead minnow (*Pimephales promelas*). Report No. 600R01067.

Varvio S-L, Koehn RK, Vkinola R. 1988. Evolutionary genetics of the *Mytilus edulis* complex in the North Atlantic region. Mar Biol 98:51-60.

Waldock MJ, Thain JE, Waite ME. 1996. An assessment of the value of shell thickening in *Crassostrea gigas* as an indicator of exposure to tributyltin. In Champ MA, Seligman PF, editors. Tributyltin: Environmental Fate and Effects. London (UK): Chapman and Hall. p. 219-237.

Weber CI. 1973. Biological field and laboratory methods for measuring the quality of surface waters and effluents. Cincinnati (OH): U.S. Environmental Protection Agency. EPA-670/4-73-001.

Wiederholm T. 1980. Use of benthos in lake monitoring. J Wat Poll Cont Fed 52:537-547.

Williams JD, Warren ML Jr, Cummings KS, Harris JL, Neves RJ. 1993. Conservation status of freshwater mussels of the United States and Canada. Fisheries 18(9):6-22.

Wolfe BA, Burkhard MJ, Leavell S, Weiss RB, Kuehn K, Valentine H, Watters GT. 2009. Health and stress assessment in freshwater mussels following translocation to a captive setting. In Proceedings of the International Symposium of the Freshwater Mollusk Conservation Society, 19-24 April 2009, Baltimore, Maryland. p. 37.

Wrona FJ, Culp JM, Davies RW. 1982. Macroinvertebrate subsampling: a simplified apparatus and approach. Can J Fish Aquat Sci 39:1051-1054.

Xu Y, Wang WX. 2002. Exposure and potential food chain transfer factor of Cd, Se and Zn in marine fish *Lutjanus argentimaculatus*. Mar Ecol-Prog Ser 238:173-186.

Zar JH. 1999. Biostatistical analysis. Fourth edition. Upper Saddle River (NJ): Prentice Hall.

Table of Contents

10.	Information Management and Interpretative Reports.....	10-1
10.1	Electronic Reporting	10-1
10.1.1	Effluent and Water Quality Monitoring.....	10-1
10.1.2	Biological Monitoring.....	10-1
10.2	Interpretative Reports.....	10-2
10.2.1	Interpretative Report Content	10-2
10.2.1.1	Interpretative Report for Standard Biological Monitoring Studies to Assess Effects	10-2
10.2.1.2	Interpretative Reports for Magnitude and Geographic Extent and Investigation of Cause Studies.....	10-4
10.2.2	Interpretative Report Structure	10-4

10. Information Management and Interpretative Reports

Environmental effects monitoring (EEM) reports shall be submitted electronically in the format provided by Environment Canada as per section 23 of the *Metal Mining Effluent Regulations* (MMER). Reports shall be submitted in writing to the Authorization Officer if an electronic format is not provided or is impracticable. The following sections of this chapter provide additional information on the reporting format of the different EEM reports.

The MMER state the required dates for data and report submissions, and Chapter 1 of this document provides an additional description of reporting requirements. Contact information for Authorization Officers and EEM Coordinators is available at www.ec.gc.ca/esee-eem/default.asp?lang=En&n=92476010-1.

10.1 Electronic Reporting

10.1.1 Effluent and Water Quality Monitoring

Environment Canada provides the Regulatory Information Submission System (RISS) for electronic reporting and submission of effluent and water quality monitoring studies (effluent characterization, sublethal toxicity testing and water quality monitoring). RISS is available on the Internet at the following webpage: <https://www.riss-sitdr.ec.gc.ca/riss/Global/Index.aspx>. For reporting requirements not supported by RISS, a hard copy should be submitted to Environment Canada. These requirements include the methodologies used to conduct effluent characterization and water quality monitoring as well as the quality assurance and quality control measures implemented and data related to the implementation of those measures.

The Effluent and Water Quality Monitoring Report must be submitted to the Authorization Officer no later than March 31 each year (MMER, Schedule 5, section [s.] 8). Information on the content of the Effluent and Water Quality Monitoring Report is discussed in Chapter 5.

10.1.2 Biological Monitoring

The data from biological monitoring studies must currently be submitted to Environment Canada using a standardized template in Excel format. One copy of the completed standardized template should be submitted to the National EEM Office at eam-esee@ec.gc.ca, and an additional copy should be submitted to the appropriate EEM Regional Coordinator. Biological monitoring data from standard surveys, and the components of magnitude and extent studies or investigation of cause (IOC) studies that

can be accommodated by the standardized template, must be submitted electronically. For example, if an adult fish survey is conducted as part of an IOC study, the data from the fish survey must be submitted electronically using the standardized template. The software previously available for inputting electronic EEM biological monitoring data (i.e., the EEM – Metal Mining Data Entry Software v. 2.1 for EEM biological monitoring) is being upgraded. Until the updated reporting software becomes available, the standardized Excel template, with data entry instructions, is available at <http://www.ec.gc.ca/ese-eem/default.asp?lang=En&n=2EEF6671-1>.

10.2 Interpretative Reports

Interpretative reports are submitted to the Authorization Officer. As some of the information provided in the interpretative reports is not supported in an electronic format provided by Environment Canada (e.g., maps) these reports should be submitted in hard copy.

The first interpretative report shall be submitted not later than 30 months after the date the mine becomes subject to the MMER, or not later than 42 months after the date the mine becomes subject to the MMER if historical information was submitted (MMER, Schedule 5, s. 18). Subsequent interpretative reports must be submitted 24, 36 or 72 months after the day on which the most recent interpretative report was required to be submitted, depending on the results of the previous studies (MMER, Schedule 5, s. 22). The final interpretative report shall be submitted to the Authorization Officer not later than 36 months after the day on which the notice to close the mine was provided (MMER, Schedule 5, s. 26). Further information on the reporting frequency is provided in Chapter 1.

10.2.1 Interpretative Report Content

The required content for interpretative reports depends on the type of biological monitoring study being conducted (MMER, Schedule 5, s. 17, 21 and 25). The reporting requirements for the varying types of biological monitoring studies are outlined below.

A mine could be conducting different types of studies on fish and benthic invertebrates at the same time (see Chapter 1 for further details). The interpretative report would then provide the information related to these different studies.

10.2.1.1 Interpretative Report for Standard Biological Monitoring Studies to Assess Effects

For standard biological monitoring studies that are conducted to assess effects, the interpretative report should contain the following information (MMER, Schedule 5, s. 17):

1. a description of any deviation from the study design that occurred while the biological monitoring studies were being conducted, and of any impact that the deviation had on the studies;
2. the latitude and longitude of the sampling areas in degrees, minutes and seconds, and a description of the sampling areas sufficient to identify their location;
3. the dates and times when samples were collected;
4. the sample sizes;
5. the results of the data assessment and supporting raw data of the benthic invertebrate community study, including the mean, median, standard deviation (SD), standard error, and minimum and maximum values for:
 - total benthic invertebrate density
 - taxa richness
 - evenness index (Simpson's Evenness Index)
 - similarity index (Bray-Curtis Index)
 - if the benthic invertebrate community survey is conducted in an area where it is possible to sample sediment, the total organic carbon content of sediment and the particle size distribution of sediment;
6. the results of the data assessment and supporting raw data of the fish population study, including the mean, median, SD, standard error, and minimum and maximum values, for indicators of growth, reproduction, condition and survival that include, where practicable, the length, total body weight and age of the fish, the weight of its liver or hepatopancreas and, if the fish are sexually mature, the egg weight, fecundity and gonad weight of the fish;
7. the results of the data assessment and supporting raw data of the fish tissue analysis, including the mean, median, SD, standard error, and minimum and maximum values of the concentration of total mercury wet weight in the fish tissue;
8. the identification of the sex of the fish and the presence of any lesions, tumours, parasites or other abnormalities;
9. the results of the statistical analysis performed to determine if there is a statistical difference between the sampling areas, the probability of correctly detecting an effect of a predefined size and the degree of confidence that can be placed in the calculations;
10. the identification of any effect on:
 - the fish population
 - fish tissue
 - the benthic invertebrate community;
11. a summary of the results of effluent characterization, sublethal toxicity testing and water quality monitoring;
12. the conclusions of the biological monitoring studies, taking into account:
 - the results of the statistical analysis conducted on the fish and benthic invertebrate survey data
 - the probability of correctly detecting an effect of a pre-defined size (power analysis) and the degree of confidence that can be placed in the calculations

- the results of any previous biological monitoring studies
 - the presence of anthropogenic, natural or other factors that are not related to the effluent under study and that may reasonably be expected to contribute to any observed effect
 - a description of quality assurance or quality control (QA/QC) measures that were implemented and the data related to the implementation of those measures;
13. a description of the impact of the results on the study design for subsequent biological monitoring studies; and
14. the date of the next biological monitoring study.

10.2.1.2 Interpretative Reports for Magnitude and Geographic Extent and Investigation of Cause Studies

If the two most recent interpretative reports indicate a similar type of effect on the fish population, fish tissue or the benthic invertebrate community, the interpretative report should contain, in addition to the information mentioned in section 10.2.1.1, the magnitude and geographical extent of the effect on fish population, fish tissue or the benthic invertebrate community (MMER, Schedule 5, subsection [ss.] 21(1)). A similar type of effect is defined as a statistically significant difference in the same endpoint in the same direction in two consecutive studies.

An interpretative report submitted when the mine conducts an IOC study may not necessarily contain all of the same information as the interpretative reports for standard biological monitoring (described in section 10.2.1.1) or the magnitude and geographic extent studies. The IOC interpretative report contains the cause of the effect on fish populations, fish tissue or the benthic invertebrate community and, if the cause was not determined, an explanation as to why it was not determined and a description of any steps that need to be taken in the next study to determine that cause (MMER, Schedule 5, ss. 21(2)).

10.2.2 Interpretative Report Structure

This section outlines the information that is recommended for inclusion in an interpretative report. The information in this section is generic to all biological monitoring study types and can be provided for each study. The information can be reported under the following categories:

Site Description and Mine Update: A synopsis and update of the information provided in the study design, especially with regards to mine history and operations, and ecological aspects of the study area. It is also important to indicate any changes or deviations from the initial study design. Information recommended for this section falls under the following three categories (other information may be provided on a site-specific basis):

Synopsis of mine history and operations

- significant changes that have been made to the mine site (e.g., altered process or operations, new effluent treatment or operations)
- a summary of any violations of MMER requirements during the EEM study
- relevant historical information concerning MMER and EEM

Synopsis of ecological aspects of study area

- any ecological variations that have occurred in the study area since submission of the study design, such as:
 - any new factors, natural or anthropogenic, that may affect the study area, including any new confounding factors
 - any unusual significant events that may have occurred (e.g., floods, spills)
- any new information that was not available at the time the study design was submitted

Synopsis of study design

- any changes from initial study design and the rationale for such changes

Location: Include any information pertaining to the mine site, and sampling areas and stations such as:

- comprehensive sampling areas and station location maps
- the latitude and longitude, and a description of the sampling areas and stations sufficient to identify their location
- rationale for choosing sampling areas and stations
- photographs of the sampling areas and stations

Methods: Include information on procedures and techniques used to perform the study, problems that may have occurred and their solutions, and justification of any variations in methods from those stated in the original study design. If a tracer study is used to delineate the effluent plume in the receiving environment, identify the type of chemical or biochemical tracer, the method used, and the justification behind the selection.

Data: Include all raw data in tables in appendices, and ensure that raw data are also reported electronically as previously described in section 10.1.2. The benthic invertebrate data should be organized by taxonomic level (e.g., order, family, genus and species). Also include all data and results of QA/QC assessments (e.g., standard reference materials, duplicates, field blanks, calibration charts, fish-aging validation results, and, in the case of benthic laboratory sub-sampling procedures, the accuracy, precision and recovery rate).

Statistics: Indicate the methods and tests used for statistical analysis, including a justification for these choices, assessment of data variability, data transformations, outliers or extreme values (also provide scatter plots for fish and benthic invertebrate surveys to help identify outliers and other unusual data), data screening results (for

benthic invertebrate data) and interactions (for analysis of co-variance [ANCOVA]). Further details on conducting statistical analyses are provided in Chapter 8.

Results: Include summary statistics of the raw data (including mean, median, SD, standard error, and minimum and maximum values) and statistical analysis results (including *p*-value and effect size) in table or figure formats for fish and benthic invertebrate surveys. Discuss the effect of outliers or extreme values on the results, if any, and include a comparison of habitat between sampling areas. The results of the power analysis conducted to evaluate the ability to detect a specified change for a given level of confidence should also be included.

A summary of the results of effluent characterization, sublethal toxicity testing and water quality monitoring reported annually in RISS must also be included in the interpretative report to help interpret the biological results.

Discussion: Include a discussion of the observed similarities or differences between sampling areas; discuss the implications of the factors affecting the interpretation of the results, such as any confounding factors, the validity of reference areas selected and the methodologies used, power analysis results, trend detection (e.g., interaction), and the problems encountered during the study, if any, and how they may affect the results; provide a summary of adherence to data quality objectives, standard operating procedures, and identification of any QA/QC problems; include a comparison of present results to previous studies in the same area or other mine studies in the literature; and discuss implications of the results on subsequent phases, taking into account the problems encountered, if any, and suggesting potential site-specific solutions/approaches for future phases.

Conclusions: Provide an overall assessment of the biological monitoring studies of all the combined results and discussions. Include suggestions on how future EEM studies could be improved. The information provided here will help in the development of studies in subsequent phases.

Table of Contents

11.	Public Involvement in Metal Mining EEM	11-1
11.1	Overview.....	11-1
11.2	Objective and Potential Scope of Public Involvement in EEM.....	11-2
11.3	Mechanisms for Public Involvement	11-3
11.3.1	Ongoing Mechanisms	11-4
11.3.2	Other Mechanisms	11-5
11.4	Public Involvement Plan	11-6
11.4.1	Provision of EEM Information to the Public	11-7
11.4.2	Proposed Mechanisms for Public Involvement	11-7
11.4.3	Proposed Mechanisms to Respond to Public Input	11-11

11. Public Involvement in Metal Mining EEM

11.1 Overview

The objective of this guidance is to help facilitate public involvement in the environmental effects monitoring (EEM) program, particularly on a site-specific basis. Mines are strongly encouraged to provide opportunities for public involvement in all aspects of the EEM program. Public input can play an important role throughout the EEM program, including the early planning steps prior to the initiation of EEM, preparation of the site characterization and the first study design, data interpretation for each EEM study at a site, and decisions regarding next steps in the EEM program at a site.

Following recommendations of the Whitehorse Mining Initiative (WMI) and the Assessment of the Aquatic Effects of Mining in Canada (AQUAMIN), public involvement is an important component of the metal mining EEM program. The WMI recommended the following as an underlying principle for public involvement: “More effective approaches to environmental management can be developed, and the public trust in mining enhanced, when the public and other stakeholders are fully informed and participate in decision making related to the public interest in all stages of mining.” The WMI and AQUAMIN recommended the use of public liaison committees (PLCs) (see section 11.3.1) as a mechanism for public involvement.

The focus of this chapter is on the “public” as a stakeholder group, as members of the public are often not significantly involved in programs such as EEM but may be able to make important contributions to the EEM program. However, it is important to recognize that the public is just one of several important stakeholder groups with an interest in the metal mining EEM program. For the purposes of this document, a stakeholder is defined as any person or group that has an interest in, is affected by, or has an effect on a watershed where a mine is operating, or has a role in decisions made pertaining to that watershed. The public includes all stakeholders with the exception of mining companies or associations, and federal, provincial, territorial, Aboriginal or municipal government departments or agencies.

It is important to bear in mind that “the public” is not a homogeneous group, and that within “the public” there may be several different stakeholders, each with different interests and concerns. For example, these interest groups may include Aboriginal groups, environmental non-governmental organizations, community groups, commercial and/or sport fishers, and concerned individuals.

11.2 Objective and Potential Scope of Public Involvement in EEM

The objective of public involvement in EEM is to ensure that decisions made regarding metal mining EEM are a result of informed, inclusive and fair consultation with the public. Effective public involvement in EEM may result in:

- improved EEM study design;
- improved decision making in EEM;
- increased relevancy of EEM;
- increased degree of trust between all stakeholders, and established/improved working relationships;
- improved public education, resulting in increased awareness and understanding of EEM issues; and
- improved communication between stakeholders.

To achieve this objective and to derive maximum benefit from public involvement, mines are strongly encouraged to facilitate public involvement in a wide range of EEM activities, such as:

Pre-EEM: Before metal mines begin EEM, the public could have input regarding the site-specific objectives of EEM, as well as the site-specific questions to be addressed. This input would help in shaping the overall direction of the EEM program.

Site characterization: Public involvement during the preparation of the site characterization can be invaluable. Given their knowledge of local conditions, the public may make significant contributions to the description of the study area and possible confounding influences, particularly with respect to the fisheries. In addition, the public may be able to help in the identification of valued ecosystem components (VECs), which are elements of the environment valued for biological, scientific, socio-economic, aesthetic or cultural reasons. VECs may be used to help refine the site-specific objectives and site-specific questions. VECs could include, for example, a fish species of cultural or economic significance, a reach of a stream valued for recreational purposes, or a scenic view. If available, such information should be used in designing the EEM study.

Study design: Following the public involvement in site characterization, public involvement in study design may contribute to the establishment of site-specific environmental quality objectives, development of reporting procedures, and identification of suitable exposure and reference areas. In addition, the public may review study designs prior to the commencement of monitoring.

Monitoring activity: The public may be actively involved in some aspects of the monitoring work, particularly sample collection. For example, the U.S. Environmental Protection Agency has developed a protocol for the collection of water samples by

community groups. In British Columbia, the Pacific Streamkeepers Federation has developed a handbook that includes such topics as water sample collection, stream habitat surveys and stream invertebrate surveys. The existence of such programs points to the fact that, with proper training, the public can be involved in monitoring work. In the field, the public could augment the work of professionals in environmental monitoring. Public involvement can help increase the cost-effectiveness of monitoring, while at the same time increasing the awareness and knowledge of the public. Involving the public in monitoring provides an opportunity for the mine to train and educate the public in EEM, and to increase public awareness and understanding of EEM components.

Data assessment and interpretation: The public may review EEM interpretative reports, and have input on decisions regarding the next steps in the monitoring program. This input would be provided recognizing that there are certain aspects of the program that cannot be significantly altered.

11.3 Mechanisms for Public Involvement

Mines are encouraged to establish mechanisms for public involvement as early as possible in the EEM process, recognizing that mechanisms may evolve as relations with the public change. There is a range of mechanisms by which the public could be involved in metal mining EEM. The WMI and AQUAMIN strongly recommended the formation of PLCs as a mechanism for public involvement. However, the appropriate mechanism for a particular site depends, in part, on the intended scope and degree of public involvement at that site.

A mine may employ more than one mechanism to facilitate public involvement. This may be particularly helpful in cases where a mine wishes to use a complementary mechanism to reach a broader segment of the public. It may also be useful in cases where a mine wishes to use more than one mechanism to reach different segments of the public that may have strongly divergent interests, or in cases where more than one language is spoken within a community.

It is very important to note that, at sites where effective public involvement mechanisms are already in place, mines are encouraged to use those mechanisms to address EEM issues, rather than establish new mechanisms.

It is essential that a mine commit an appropriate level of financial and staffing resources to support the mechanisms to be implemented. If resources are not adequate, it is unlikely that the mechanisms will be effective. To facilitate public involvement, the mine may need to make some resources available to public participants in order to cover expenses associated with participation. This may be particularly important in remote areas where there may be costs associated with travel.

Mechanisms by which the public could be engaged include those described below. Note that the mechanisms identified here are not exhaustive. There are other potential mechanisms that a mine may choose to use.

11.3.1 Ongoing Mechanisms

Public Liaison Committee

The WMI and AQUAMIN recommended that the public be involved as actively as possible, and that sharing authority on certain issues is desirable. Further, AQUAMIN recommended that Environment Canada, in consultation with other stakeholders, develop guidelines for the establishment of PLCs, including the reporting of information to the public.

It is strongly recommended that PLCs be formed to facilitate public involvement, whenever there is sufficient public interest. PLCs can help ensure that decisions regarding EEM are made in an open, transparent and inclusive manner. If a PLC is to be established, the following factors should be considered:

1) Stakeholders should be involved as soon as possible and should participate in process design, including:

- setting the terms of reference for the PLC;
- determining the focus of discussion, within the scope determined by the mine operator; and
- identifying participant funding needs, and what expenses may be covered.

2) All stakeholders should have a clear understanding of the process, including:

- the objective of EEM;
- how the public is to be involved;
- the objective of the public involvement process;
- the scope of public involvement, and the scope of decisions to which the public has input;
- who the final decision makers are, and the fact that decisions will be communicated to the PLC and general public, complete with reasons for the decision; and
- the consequences of not reaching consensus.

3) Membership in the PLC should be determined on a site-specific basis, and should include representatives of the company and the public. Membership may also include:

- representatives of the federal government;

- representatives of relevant provincial/territorial/Aboriginal government departments or agencies; and
- company employees.

Meetings of the PLC should be open to the general public, including anyone wishing either to observe or participate in a specific meeting, thus ensuring that the process remains open and transparent.

4) Accurate, credible and timely information should be equally available to all participants.

5) Consensus among the participants on the PLC should be the ultimate goal. In the case of consensus not being reached, the PLC would forward its findings, including full discussions of dissenting viewpoints, to the decision-making body. The decision makers would thus have a clear understanding of the situation and the different options presented, as a basis upon which to make a final decision.

Public Liaison Contact

Where there is insufficient public interest to establish a PLC, a mine may identify a member of the public to volunteer to act as a Public Liaison Contact. The Public Liaison Contact should be a person not connected with the mine, either as an employee or contractor, or as a relative of an employee or contractor. It may be difficult to identify an appropriate person to act as Public Liaison Contact, but care should be taken to try to ensure that the person identified is acceptable to the public.

The Public Liaison Contact would receive copies of EEM-related correspondence between the mine and the Authorization Officer, and other relevant documentation such as any information prepared by the mine that is specifically intended for the public. The Public Liaison Contact would be the primary point of contact with the public for the mine, would play a role in ensuring broader distribution of relevant documentation to the public, and could assist the mine in planning and implementing short-term and/or complementary mechanisms for public involvement.

11.3.2 Other Mechanisms

Open houses: drop-in events designed to allow the public to obtain information and respond at their convenience. Open houses may consist of a visual display, together with handouts and knowledgeable staff to answer questions and solicit opinions.

Public meetings: opportunities to inform the public and for the public to make formal and informal presentations, and to exchange comments. To be effective, public meetings need to follow an agenda. A representative of the mine or a neutral party should chair public meetings.

Workshops: carefully planned forums designed to air certain issues and share different points of view. Workshops are usually limited to a small number of invited participants. A facilitator, whose role is to encourage dialogue, structure input toward the workshop goals and summarize results, may chair the workshops.

Community visits: visits by mine staff to community groups in order to interact directly with local citizens. These provide an opportunity to interact with the public in their “domain,” meaning such visits may be more conducive to constructive informal dialogue. This may be a valuable option to consider in remote areas.

Site visits: interested participants visit a mine site to obtain first-hand information and orientation. Such a visit provides an opportunity for direct contact and exchange of information with the public, and provides the public with an enhanced understanding of the project.

Electronic communications: Internet sites or other means for a mine operator to make information available and receive feedback. These means provide a cost-effective way to make large amounts of information available, and simplify logistics given that there is no need for building address lists, copying, and mailing out documents. However, appropriate in-house technical expertise is required. Electronic communications should not be used exclusively, because individuals without access to the necessary computer hardware and software will not be able to participate.

11.4 Public Involvement Plan

To help in the development and implementation of public involvement processes, mines are encouraged to prepare public involvement plans. The first plan should be prepared as early as possible in the EEM program, and plans should be updated periodically as the EEM program and public involvement activities at a site evolve. Mines are encouraged to provide copies of public involvement plans to the Authorization Officer. Plans provided to the Authorization Officer will be kept on file so that there is a public record of plans submitted.

The objective of the public involvement plan is to outline mechanisms by which the mine proposes to provide information on EEM to the public, seek input from the public, and respond to input from the public.

Note that as part of the public involvement plan, the mine employee who is responsible for public involvement activities should be identified.

11.4.1 Provision of EEM Information to the Public

The timely provision of information to the public is essential to public involvement processes. As a minimum, mines should make available to the public the executive summaries and text of each public involvement plan, EEM study design and EEM interpretative report. In addition, mines are encouraged to provide to the public any other documentation that may be helpful.

The public involvement plan should include the following elements:

a) A description of how the mine proposes to inform the public that information regarding the EEM program exists, and how the mine proposes to distribute this information to the public. Options for informing the public and distributing information include:

- advertisements in local media, including newspapers and radio;
- notifications on the mine website;
- notifications to community organizations, local government, resources users and labour unions;
- information kiosks at community centres or meeting places (e.g., shopping malls, town halls); and
- making information available in local libraries and any other appropriate venues.

b) A description of measures that may be taken to provide information in a form that is understandable to the public. In order to help the public understand information such as study designs, interpretative reports and other documents, consideration should be given to:

- language(s) of the community;
- education level of the public targeted in the communications;
- sensitivity to appropriate cultural communication styles; and
- acknowledging localized sensitivities in communicating with various segments of the public.

11.4.2 Proposed Mechanisms for Public Involvement

The public involvement plan should include the following elements:

a) A description of current conditions at the site. Such a description, with respect to the factors listed immediately below, is important in providing a context for the public involvement plan. Gathering this information may be very helpful to a mine planning public involvement activities, because it helps establish realistic expectations for those activities. The factors are:

- degree of previous or current public involvement activities;

- location and size of potentially affected communities;
- age, size and history of the mining operation;
- nature of historical or current environmental concerns; and
- human use of aquatic resources in the watershed.

b) Proposed objective and goals of public involvement. The overall objective of public involvement for metal mining EEM is to ensure that decisions made regarding metal mining EEM are made as a result of informed, inclusive and fair consultation with the public. Site-specific objectives and goals may be established. Clear objectives and goals will provide a basis for more effective relationships established with the public. Identifying desired results of public involvement will assist not only in the design and implementation of the public involvement plan, but in the evaluation of the effectiveness of public involvement.

The public involvement plan should also outline the mine's expectations for the proposed public involvement activities, and outline in a clear and transparent manner the proposed steps that may be taken in the event that expectations are not being met.

c) Proposed scope of public involvement. Define which EEM activities the public may be involved in and which decisions may be influenced by public input. The timing of the preparation of the public involvement plan should be consistent with the proposed scope, particularly if the mine plans to involve the public in pre-EEM activities or site characterization.

d) Defined principles for public involvement. The following are principles upon which public involvement in EEM should be based (principles appropriate to a particular site will depend on both the scope and degree of public involvement):

Open and transparent: Once a public involvement process has been initiated, all decisions related to the scope of the process should be completed in an open and transparent manner, so that all stakeholders involved, including the public, are aware of the decision-making process and the nature of the decisions. Unless the process is open, fair and equitable, agreement may not be reached and, if reached, may not last.

Purpose-driven: Participants need to clearly understand the objectives and scope of the public involvement process, and see a clear need for their involvement. To be meaningful, public involvement processes should take place while options for decisions are still open. All stakeholders involved, including the public, need a common understanding of the objective, and an acceptance that a public involvement process is an appropriate mechanism to achieve this objective.

Inclusive, not exclusive: All stakeholders, including the public, should be given the opportunity to influence and participate in the process. Mines need to identify all stakeholder groups, including public groups, that have a significant interest in the

outcome, including those who will be affected by the outcome, those who will be responsible for implementing it, and those who could undermine the outcome if not involved.

Voluntary participation: Stakeholders participate voluntarily. All stakeholders, including the public, should support the public involvement process, and will need to invest the time necessary to make it work.

Flexibility: Flexibility should be designed into the process. Operating within the framework of the metal mining EEM requirements and guidance, stakeholders should be able to work together to design site-specific public involvement processes. The initial process design may evolve as the stakeholders become more familiar with the issues, the process and each other. It is necessary to incorporate the feedback of participants in an ongoing evaluation of the process.

Equal opportunity: All stakeholders, including the public, should have equal access to relevant information, and the opportunity to participate effectively throughout the process. Whenever possible, stakeholders should have the opportunity to choose their own representatives to the process.

Respect for diverse interests: Acceptance of the diverse values, interests and knowledge of the stakeholders involved, including the public, is essential. The mine needs to allow time for other stakeholders to explore and develop common interests despite their different values. Increased understanding fosters trust and openness, which assists the stakeholders to move beyond bargaining over positions to exploring their underlying interests and needs.

Accountability: The participants, including the public, are accountable both to their constituencies and to the public involvement process that they have agreed to establish.

Cost efficiency: Public involvement processes should be carried out in a cost-effective manner. At the same time, a realistic time frame should be considered to allow participants to effectively liaise, consult and exchange information with their constituencies

Implementation: Commitment to implementation of recommendations/decisions of public involvement activities, and feedback to the public regarding implementation, are essential elements of any agreement on public involvement. In cases where recommendations/decisions are not implemented, those involved should be informed, and a rationale should be provided.

e) Proposed timelines. Realistic timelines are essential to public involvement in EEM. The public involvement plan should identify deadlines and milestones for key decisions

that may be influenced by public involvement. Timelines should provide optimum opportunities for public involvement, while allowing the mine to meet required deadlines and milestones. Mines should be aware of the scheduling and resources constraints of the public. In scheduling public involvement activities, it may be helpful to identify the schedules and availability of the key public stakeholders first, and then the availability of other public stakeholders, where appropriate. When legislated timelines are tight, and there may not be time for adequate public involvement, this situation should be communicated to the public.

f) Proposed degree of public involvement. Following on the recommendations of AQUAMIN and the WMI, it is recommended that the public be involved to the fullest extent possible at all mine sites. As the degree of public involvement increases, so do the expectations and need for commitment regarding:

- the level of skill and expertise necessary for all participants;
- resource requirements (time and money);
- expectations that consideration be given to input throughout the EEM process; and
- expectations that input would influence the final decisions made by the mine or governing body.

Options for the potential degree of public involvement include the following:

Shared authority: The public is involved in the decision-making process, to a degree agreed upon by all stakeholders. Within the terms of such an agreement, the public is an equal partner in the decision-making process. The public is formally engaged through the establishment of PLCs. However, under such an arrangement, the government and regulated industry, not the PLCs, will ensure that regulatory requirements are met.

Joint planning: The public is engaged through ongoing consultation in all phases of EEM, from objective setting to review of results. This consultation occurs in part through PLCs, with a mandate and terms of reference agreed upon by all stakeholders. In addition, broader public consultation may occur through the use of other mechanisms. All stakeholders participating are accountable to the consultation process, and there is an obligation that mines and regulatory agencies give due consideration to the results of the process.

Ongoing public consultation: Public forums and/or outreach activities to local community organizations are held on an ongoing basis. The objective of these efforts is to provide regular updates (including response to input from previous forums) and to obtain regular input from the public. The scheduling and frequency of these activities are at the discretion of the mine, and there is no formal ongoing relationship with the public, although informal relationships may develop.

Public consultation: Through public forums and/or outreach activities targeting local community organizations, public input is sought on various EEM issues, including study design. The public also has an opportunity to review monitoring results. No ongoing working relationships with the public are established.

Information feedback: Information on the status of the EEM program, including monitoring results, is presented to the public at open forums or through another means. Forums provide opportunities for the public to comment.

Information: Information on the status of the EEM program, including monitoring results, is available to the public, but there is no formal mechanism for public input or involvement in EEM.

g) Resources for proposed public involvement activities. It is essential that a mine ensure that adequate resources are in place to support the proposed public involvement activities. If adequate resources are not committed, even the best plans may fail. It is a mine's prerogative to keep internal resourcing issues confidential, but a mine should provide assurances to stakeholders, including the public, that it has committed adequate resources to the proposed public involvement process.

The public involvement plan should clarify the types of financial resources that may be available to support the participation of public stakeholders, as well as the types of expenses that would be covered and to what degree and by what type of mechanism they would be covered.

11.4.3 Proposed Mechanisms to Respond to Public Input

Essential to the public involvement plan is the proposing of mechanisms to respond to public input. Such proposals are needed to provide assurance to the public that their input will be considered. The broader the proposed scope of public involvement activities and the greater the degree of proposed involvement, the higher the public's expectations will be that the recommendations/decisions of public involvement activities are going to be considered. Therefore, proposed mechanisms to respond to public input should be developed to a level of detail appropriate to meet the expectations of the public.

Mechanisms to respond to public input should include mechanisms for decision making based on public involvement, and mechanisms for informing the public of decisions made (including a rationale for not accepting recommendations from the public).

Table of Contents

12.	Investigation of Cause.....	12-1
12.1	Introduction.....	12-1
12.2	Goal and Expectations of Investigation of Cause	12-1
12.3	Developing Hypotheses and Study in IOC	12-2
12.4	Using Response Patterns and Population Dynamics	12-4
12.4.1	Overview of Metal Mining EEM Results	12-4
12.4.2	Fish Response Patterns	12-4
12.4.3	Possible Effect Profiles for Fish	12-5
12.4.3.1	Nutrient Limitation/Toxicity	12-5
12.4.3.2	Metabolic Disruption	12-5
12.4.3.3	Eutrophication	12-6
12.4.4	Benthic Community Response Patterns.....	12-6
12.4.5	Possible Effect Profiles for Benthos	12-6
12.4.5.1	Eutrophication/Smothering	12-6
12.4.5.2	Toxicity/Smothering	12-7
12.4.5.3	Meromictic/Hypoxic Conditions	12-7
12.5	Tiered (Elimination) Approach.....	12-8
12.5.1	Toxicity Reduction Evaluation and Toxicity Identification Evaluation	12-8
12.5.1.1	Toxicity Reduction Evaluation.....	12-8
12.5.1.2	Toxicity Identification Evaluation.....	12-9
12.5.2	Investigation of Cause Framework Example of Toxicity Reduction Evaluation and Toxicity Identification Evaluation	12-9
12.5.2.1	Defining Response Patterns	12-10
12.5.2.2	Source Identification & Selective Operation of In-plant Processes	12-11
12.5.2.3	Chemical Isolation and Characterization	12-11
12.6	Integrated (Weight-of-Evidence) Approach	12-12
12.6.1	Using Weight of Evidence	12-12
12.6.2	The Sediment Quality Triad, as an Example of Weight of Evidence	12-14
12.7	Metal Toxicology and Bioaccumulation.....	12-16
12.7.1	Ecotoxicity of Metals.....	12-17
12.7.2	Selenium Toxicity.....	12-18
12.7.3	Biomarkers (Metallothionein).....	12-18
12.7.4	Using Bioaccumulation Methods.....	12-19
12.7.5	Methods for Correlating Bioaccumulation and Toxicity	12-20
12.7.5.1	Tissue Residue Approach	12-20
12.7.5.2	Biotic Ligand Model	12-21
12.7.5.3	Subcellular Fractionation	12-21
12.8	Field Study Approaches for Fish and Benthos	12-21
12.8.1	Field Study Design.....	12-22
12.8.2	Caged Organisms	12-22
12.8.3	Mesocosms (Fish and Invertebrate)	12-23
12.8.4	Benthic Transplant Devices	12-23
12.8.5	Macrophytes, Periphyton, Phytoplankton, Chlorophyll	12-23

12.8.5.1	Macrophytes.....	12-24
12.8.5.2	Periphyton	12-24
12.8.5.3	Phytoplankton.....	12-25
12.8.5.4	Indicators for Phytoplankton and Periphyton	12-25
12.9	Laboratory Methods and Toxicity Tests	12-26
12.9.1	Histology and Physiological Parameters	12-26
12.9.2	Toxicity Tests.....	12-27
12.9.3	Fish Toxicity Testing: Fathead Minnow Lifecycle.....	12-27
12.9.4	Whole Sediment Toxicity Testing	12-28
12.9.4.1	Survival and Growth of <i>Hyaella azteca</i>	12-29
12.9.4.2	Survival and Growth of the Freshwater Midges <i>Chironomus dilutus</i> or <i>Chironomus riparius</i>	12-29
12.9.4.3	Survival and Reproduction of the Oligochaete <i>Tubifex tubifex</i> ...	12-30
12.9.4.4	Survival and Growth of the Mayfly <i>Hexagenia limbata</i>	12-30
12.9.5	Metals in Overlying Water in Whole-Sediment Toxicity Tests	12-30
12.9.6	Sediment Pore Water Toxicity Testing.....	12-31
12.9.6.1	Sea Urchin Fertilization Bioassay.....	12-31
12.9.6.2	Microtox Acute Test	12-31
12.9.6.3	Acute Lethality with <i>Daphnia magna</i>	12-32
12.9.6.4	Reproduction and Survival Using <i>Ceriodaphnia dubia</i>	12-32
12.9.6.5	Growth Inhibition of the Alga <i>Pseudokirchneriella subcapitata</i>	12-32
12.10	Tools for Effluent and Water Quality Analysis	12-32
12.10.1	Measurement of Dissolved Metals during Investigation of Cause ..	12-33
12.10.2	Metal Speciation for Metals of Concern.....	12-34
12.10.3	Measurement of Reagents & Reagent By-products used in Processing... ..	12-35
12.10.4	Monitoring of Flow and Loadings in the Exposure Area	12-36
12.11	Tools to be Considered for Sediment Analysis.....	12-37
12.11.1	Sediment Monitoring as a Tool for Investigation of Cause.....	12-37
12.11.2	Sediment Mass Transport	12-37
12.11.3	Sediment Depositional Rate & Sediment Dating for Historical Trends.....	12-38
12.11.4	Sediment Coring	12-39
12.11.5	Sediment Chemistry.....	12-39
12.11.5.1	Partial Metal Concentrations in Sediment (Partial Extraction), Sequentially Extracted Metals, SEM/AVS Ratios.....	12-39
12.11.5.2	C/N Ratio for Marine Sediment	12-40
12.12	Sediment Pore Water Analysis	12-41
12.13	References.....	12-43

List of Tables

Table 12-1: Examples of approaches that may be used to determine possible causes during IOC	12-3
Table 12-2: Formalized set of causal criteria forming part of a weight-of-evidence approach for assessment of mining effluent effects.....	12-13

List of Figures

Figure 12-1: Tiered framework for investigation of cause in environmental effects monitoring.....	12-10

12. Investigation of Cause

12.1 Introduction

This chapter is a compilation of all the Investigation of Cause (IOC) sections that were included in various chapters of the 2002 version of the Metal Mining (MM) Environmental Effects Monitoring (EEM) Guidance Document. Several methods and concepts have been updated with new information.

An EEM MM IOC workshop was hosted jointly by the Metal Mining Association of Canada and Environment Canada (EC) in December 2009. The objective of the workshop was to explore various aspects and challenges related to the IOC phase of the Metal Mining EEM program. Background information on IOC, environmental studies, potential causes of effects, various case studies, and tools and approaches for conducting IOC studies were presented. The workshop provided a forum for discussion and development of IOC for metal mines, and allowed research needs to be identified. The proceedings of the workshop were used to further develop this IOC chapter for the Metal Mining EEM Program, and further details can be found within the proceedings, which were published in 2012 and are available on the EEM website at www.ec.gc.ca/esee-eem.

12.2 Goal and Expectations of Investigation of Cause

The goal of an IOC study is to determine the cause of confirmed effects (i.e., what is responsible for the effects) (see Chapter 1, section 1.3.2.1 for information regarding confirmed effects). The guidance presented in this chapter is intended to allow flexibility in designing IOC studies in order to accommodate site-specific needs, and should be complemented by all available information, including publicly available scientific literature. As with all EEM studies, IOC studies are required to use validated methods and good scientific practice.

Acceptable validated methods include those whose soundness has been established on an authoritative basis, such as standardized methods/procedures prepared by recognized international agencies (for example, EC, the Organisation for Economic Co-operation and Development (OECD), the United States Environmental Protection Agency (US EPA), the American Society for Testing and Materials (ASTM), the International Organization for Standardization (ISO), the European Union (EU) and the World Health Organization (WHO)), as are methods published in peer-reviewed journals within the literature and accepted as scientifically defensible protocols. In addition, non-authoritative methods that have not been peer reviewed may be used provided that evidence is presented to demonstrate their soundness.

Components of good scientific practice include scientific integrity, experimental design, laboratory safety, error analysis, quality assurance and control, critical data interpretation,

and accurate record keeping (i.e., method documentation and data collection, documentation, curation and access) (OECD 2007, IARC 2008, Deutsch Forschungsgemeinschaft 1998). EEM studies are expected to be of sound science, i.e. “organized investigations and observations conducted by qualified personnel using documented methods and leading to verifiable results and conclusions” (SETAC 1999). Sound science implies that the data and conclusions are supported by the high standards of the scientific method, which include the development of a readily testable hypothesis or hypotheses, the use of systematic and well-documented experimental or analytical methods (adequate sample sizes, appropriate controls), the use of appropriate data analysis tools (statistics or models), and the articulation of conclusions that address the hypothesis(es) and that are supported by the results (SETAC 1999). The Society of Environmental Toxicology and Chemistry (SETAC) (1999) makes note of important cautions with respect to data interpretation, including statements of certainty vs. uncertainty, causation vs. correlation, absence of evidence vs. evidence of absence, and potential for misinterpretation (overlooking variables, inadequate sample sizes, lack of appropriate controls, bias, anecdotes). In addition, Bosker and Munkittrick (2009) raise caution regarding the following generic quality assurance / quality control (QA/QC) and statistical issues within EEM monitoring studies: insufficient sample size, failure to remove outliers, uncertain exposure, poor reference site selection, incomplete/poor reporting, data entry error, and wrong data reported.

12.3 Developing Hypotheses and Study in IOC

In developing the IOC study design (see Chapter 2, 2.2, and Chapter 1, 1.4.2.2), mines are encouraged to outline potential hypotheses that could explain the causes of the confirmed effects, taking into account any identifiable response patterns. Multiple effects may have the same cause, and may be captured through establishing an overarching hypothesis. In other cases, multiple hypotheses may be constructed to address multiple confirmed effects. Tools appropriate to address the established hypotheses could then be proposed. The scope of work could range. It may be sufficient to examine and present solid evidence using existing data, alone or in combination with new data (weight-of-evidence approach) and/or a literature review. Alternatively, it may be necessary to conduct full field and/or laboratory studies.

The table below shows examples of different IOC approaches to examine various hypotheses of possible causes. The following sections of this document suggest many ways of investigating the cause of effects.

Table 12-1: Examples of approaches that may be used to determine possible causes during IOC (not meant to be prescriptive or exhaustive; please consult the scientific literature)

Hypotheses (Possible Cause)	Approach	Example
Habitat, nutrients	Collection of biological data with supporting measures to characterize habitat, water quality and sediment chemistry that would allow for expanded statistical analyses (e.g., correlative, multivariate approaches). Effluent chemistry samples (extensive if necessary) collected simultaneously may allow for links to be made between conditions in the receiving environment to possible causative components in the mine effluent. Characterization of unusual events or conditions may provide further knowledge.	1) Standard benthic survey with corresponding water, sediment and effluent chemistry to explore relationships with habitat, nutrient and contaminant measures. 2) Collection of data on temperature, nutrients and primary productivity (chlorophyll <i>a</i>, periphyton) in the receiving environment with corresponding measures of nutrients in effluent to study a nutrient-related cause of effect(s), such as eutrophication, and may include modelling and/or mass balance component.
Natural variability	See Chapters 2, 3, 4, 6, 7, 8 and 9.	
Contaminants	Analysis of metal body burdens, alone or in combination with other measurements e.g., fish gonad or liver histology, to look for signs of abnormal development (gonads) or signs of pathology (e.g., lesions on livers); measurement of liver enzymes.	1) Analysis of trace elements in amphipods collected in the reference and exposure area, and comparison of levels to published “critical body concentrations.” 2) Biomarker approach for fish e.g., measure metallothionein concentration and/or other low molecular weight proteins that are known to be induced by exposure to metals. 3) Analysis of trace elements in tissues (e.g., liver, viscera, whole body, and muscle, as appropriate for identified effect and/or contaminant of concern) of fish collected in the reference and exposure area; determines whether contaminants are bioavailable and being accumulated. 4) Use of a tiered approach (toxicity reduction evaluation / toxicity identification evaluation) to determine the source of the contaminants in the waste streams.
Food limitation	Literature review on the diet of fish to understand preferences and examination of existing benthic data to assess food availability, alone or in combination with other information, such as stable isotopes.	Measures of stomach contents, relevant prey types (e.g., benthic, planktonic prey), or water quality parameters such as turbidity, temperature.

12.4 Using Response Patterns and Population Dynamics

12.4.1 Overview of Metal Mining EEM Results

The national analyses of metal mining EEM data (Lowell 2007; Environment Canada 2012) suggest that the fish response patterns (fish were older, thinner (reduced condition), with smaller livers and gonads) may be due to toxicity or habitat alteration. The data also suggest that benthos response patterns (significantly reduced taxon richness, changes in community structure in exposed areas as measured by the increased Bray-Curtis Index and Simpson's evenness endpoints, and increased density) may be due to toxicity/habitat alteration, or eutrophication for some mines. To date, reported mine effluent effects on fish usability (as measured by mercury in fish tissue) appear to be minimal.

12.4.2 Fish Response Patterns

It is important to understand that the response of the fish population sampled is a snapshot in time, and that the response should not be assumed to represent a step in a progression of responses that may lead from one steady-state condition to a new one (Environment Canada 2010). The response is also a reflection of how the existing fish are performing, and not an indication of the mechanism of impact. For example, with an increase in the quantity or quality of food or habitat, there should be an increase in growth, size, reproductive investment (gonad size), and condition. A faster growth rate typically results in fish reproducing at a younger age, and therefore lowers age-to-maturity (the age at which the fish start reproducing). In combination, these changes in the population normally decrease the average age of the population. When the population adapts to its new carrying capacity, parameters should revert to reference levels, but at a higher density of fish. As well, an acutely lethal accidental discharge may be reflected later in time as an increase in food resources, because there is a lower density of fish and the same amount of food is available. Thus, the effects may not result in longer-term changes in the fish community, because the population is maintaining equilibrium, and corrective actions may not be ecologically effective or cost effective.

Not all species will respond directly to stress, but some may respond indirectly due to changes in predation pressure or food availability. The observed response pattern can, however, be used to interpret results and design studies for the next phase. It is important to look at supporting data to help with interpretation and study design. Gibbons and Munkittrick (1994) and Munkittrick et al. (2000) grouped fish characteristics according to age structure (mean age or age distribution), energy expenditure (growth rate, reproductive rate) and energy storage (condition, liver weight). They assigned an increase, decrease or no change to each characteristic, to come up with a generalized response pattern that could be used to provide direction for research into causal factors. Moreover, the nature of the response pattern over successive monitoring phases enables characterization of the status of the system in question, eventually enabling management

decisions to be made regarding the effectiveness of current regulations (Environment Canada 2010). The IOC/IOS chapter of the Pulp and Paper EEM Technical Guidance Document (Environment Canada 2010) and Munkittrick et al. (2000) give detailed information regarding effect profiles and interpretable response patterns. Possible interpretable response patterns that mines could consider, for example, are nutrient limitation, toxicity, eutrophication, and metabolic disruption. Other response patterns exist and may also be applicable.

12.4.3 Possible effect profiles for fish

Three effect profiles are described below; others exist and may be applicable.

12.4.3.1 Nutrient Limitation/Toxicity

Nutrient limitations affecting fish health (which may be characterized by decreased condition factor, liver size and gonad size) can result initially in a decrease in fish growth and reproduction. Over time this can lead to an increase in age of the population, because fewer young are being produced (Gibbons and Munkittrick 1994). A prolonged problem with food availability and performance will eventually lead to a reduction in population size below the carrying capacity of the system, and the performance parameters in fish (growth, condition) may begin to recover as the population continues to grow older and the population grows smaller.

Chemical toxicity may cause an increase in liver weight, with a decrease in condition and gonad weight. It has been suggested that the increased liver size is associated with increased activity of detoxification processes. It is important to note, however, that chemical toxicity can result in increased liver detoxification enzymes without the presentation of enlarged livers, and vice versa. Enlarged livers may be an indicator of altered energy storage due to toxicity but not directly related to an increase in the detoxification enzymes (Munkittrick et al. 1994, 2000).

12.4.3.2 Metabolic Disruption

As noted in Hewitt and Servos (2001), the Government of Canada (in the *Canadian Environmental Protection Act, 1999*) has defined endocrine-disrupting substances as substances that have the ability to affect the synthesis, secretion, transport, binding, action or elimination of hormones, with effects on the maintenance of homeostasis, reproduction, development or behaviour of an organism. Effects on reproduction may be characterized by reduced gonad size and increased indicators of energy use (growth, liver size, condition), indicating that there is energy available but fish are not directing this energy to reproduction (Munkittrick et al. 1991). Heavy metals in particular are known to be metabolic or endocrine disrupters, including cadmium, copper, lead, iron, mercury, and selenium (Hewitt and Servos 2001; Fajreaus-Van Ree 2004). Such heavy metals have been associated with impaired stress responses and adrenal function in fish in the laboratory environment and in the wild (Brodeur et al. 1998; Brodeur et al. 1997).

Hontela et al. (1992) showed that fish exhibit an exposure-related decrease in condition factor and growth efficiency, and a reduced capacity to elevate blood cortisol.

12.4.3.3 Eutrophication

The eutrophication pattern (characterized by increases in gonad weight, liver weight and condition) can be a result of either a decrease in population size or an increase in available habitat and food resources. Decreased population size may be associated with increased predation because of an increased abundance of predators, or an increase in mortality due to the aging population. However, with an increased reproductive rate, the long-term results could be an increase in younger fish, which could eventually lead to limitation of food resources (Gibbons and Munkittrick 1994), if the younger fish are not kept in check by predation.

12.4.4 Benthic Community Response Patterns

The EEM benthic community survey effect indicators (and effect endpoints) are density (number of animals per unit area), taxa richness (number of taxa), similarity index (Bray-Curtis Index) (measure of community structure differences between 2 assemblages) and evenness index (Simpson's Evenness Index) (how evenly individuals are distributed among the taxa). By comparing these effect endpoints between an exposure area and a reference area, or along a gradient, it is possible to detect structural differences in the benthic community. This information can be used to determine the amount of energy available for the fish, and thus is a measure of fish habitat health. The IOC/IOS chapter of the Pulp and Paper EEM Technical Guidance Document (Environment Canada 2010) and Munkittrick et al. (2000) give detailed information regarding certain effect profiles and some interpretable response patterns. Possible interpretable response patterns that mines could consider, for example, are nutrient limitation, toxicity, and eutrophication. Other response patterns exist that may also be applicable.

12.4.5 Possible Effect Profiles for Benthos

Two effect profiles are described below; others exist that may also be applicable.

12.4.5.1 Eutrophication/Smothering

Eutrophication, or nutrient enrichment, is a process of over-fertilization of a water body by nutrients, resulting in the production of more organic matter than the self-purification reactions of that water body can overcome (Chambers et al. 2001). The degree of eutrophication affecting benthic invertebrates can range from mild to moderate or pronounced. In the case of mild to moderate eutrophication, the typical response is an increase in the abundance¹ and number of benthic invertebrate taxa (taxon richness)

¹ Abundance is defined as the number of benthic invertebrate individuals. The term density is used when abundance is expressed per unit area sampled.

relative to reference conditions. Pronounced eutrophication (decreased taxon richness, increased density) will begin to shift the composition of the benthic community. Hyper- or severe eutrophication may be observed when the abundance and taxon richness of benthic organisms decline; at this stage, negative impacts on fish stocks and plant life are usually observed as oxygen is depleted by decomposing organic matter (Environment Canada 2007; Lowell et al. 2000).

Some studies in the literature suggest that phosphorus complexes with metals in the water column as suspended solids; these complexes then deposit in the sediment (Ledo et al. 2004; Pereira et al. 2008). Under certain conditions (reducing environment, seasonality (fall/winter) and daylight), the phosphorus is released, causing eutrophication (Ledo et al. 2004; Pereira et al. 2008). Ammonia/nitrates are commonly elevated due to blasting materials, cyanide breakdown, or sewage inputs (CCME 2009); these can act as toxicants but could also contribute to enrichment.

Pronounced eutrophication is commonly associated with an increase in the number and type of pollution-tolerant (benthic) taxa (e.g., oligochaetes, chironomids or nematodes) and a decrease in the number and type of sensitive species (e.g., mayflies, stoneflies or caddisflies). Severe eutrophication frequently masks toxicity effects that may otherwise have been measured (Lowell et al. 2000).

12.4.5.2 Toxicity/Smothering

Trace metals are potentially toxic to organisms and can impact individual and population performance. Decreases in both taxon richness and abundance are typically a sign of overall inhibitory effects, such as toxicity or smothering (Lowell et al. 2000). Such population-level changes are reflected in the benthic community structure, with shifts towards an increasingly simple, and often predictable, species composition as sensitive taxa become rare or disappear and more metal-tolerant taxa dominate (Clements 2004; Pollard and Yuan 2006; Courtney and Clements 2002; Canfield et al. 1994).

Hypoxic or anoxic water conditions can develop when oxygen is consumed by decomposing organic matter. If currents are weak and the organic matter is not being flushed from the area, these conditions may generate potentially toxic reduced compounds such as methane, ammonia and hydrogen sulphide (Pearson and Rosenberg 1978). Toxic and anoxic conditions can lower invertebrate feeding rates, which potentially can lower invertebrate growth and biomass (which in turn will affect the fish community) (Lowell et al. 2000).

12.4.5.3 Meromictic/Hypoxic Conditions

Mine effluent can produce stratification in some deep water lakes, resulting in meromictic conditions in these lakes. Normal seasonal mixing does not occur, causing vertical and temporal differentials of dissolved metals, and meromictic cycling of some metals. Dissolved oxygen at depths below the chemocline of the lake is not replenished, resulting

in oxygen depletion (Sánchez et al. 2008; Szarek-Gwiazda and Żurek 2006; Campbell and Torgersen 1980). This ultimately leads to toxic conditions and the loss of habitat.

12.5 Tiered (Elimination) Approach

12.5.1 Toxicity Reduction Evaluation and Toxicity Identification Evaluation

12.5.1.1 Toxicity Reduction Evaluation

A toxicity reduction evaluation is a site-specific step-wise diagnostic approach to resolving toxicity issues (Novak and Holtze 2012). Protocols to investigate the probable causes of toxicity have been developed by the US EPA (1989) and are known as Toxicity Reduction Evaluation (TRE) and Toxicity Identification Evaluation (TIE) studies.

The general objectives of TRE/TIE processes (US EPA 1989) are to determine those actions necessary to reduce the effluent's toxicity to acceptable levels. A six-tier approach was developed, based on US EPA (1989), (1991a), (1991b), (1993a) and (1993b), and Novak and Holtze (2012), directed toward the reduction of toxicity of the whole effluent rather than specific components within the effluent. The six tiers are:

- 1) Information and data acquisition
- 2) Evaluation of remedial actions in operation/process to reduce effluent toxicity
- 3) TIE of the effluent:
 - Phase I: Characterization of toxicity through various treatments
 - Phase II: Identification of suspected toxicants
 - Phase III: Confirmation of toxicants
- 4) Source(s) investigation/identification of the toxicity in the facility
- 5) Toxicity Treatability Evaluation for reduction of toxicity in the final effluent
- 6) Confirmation and removal of toxicity

12.5.1.2 Toxicity Identification Evaluation)

The TIE approach uses the responses of organisms within an appropriate bioassay format to detect the presence of active agents. This approach characterizes the active substances of interest in a complex matrix comprising three phases: characterization, identification, and confirmation of toxicants.

1. Phase I aquatic TIE methods were originally developed for use with acute lethality tests using fathead minnows or *Ceriodaphnia dubia* (US EPA 1991a), but have been adapted for sublethal testing (US EPA 1991b).
2. Phase II aquatic toxicity identification is described in US EPA document (1993a).
3. Phase III aquatic toxicity confirmation is described in US EPA document (1993b).

TIE protocols also exist for sediment and pore water toxicity (US EPA 1991c and 1997) and marine toxicity (phase I only) (US EPA 1996).

The standard US EPA Phase I effluent characterization treatments involve subjecting the effluent or water to various treatments that remove or separate specific chemical classes (for example, filtration, aeration, extraction, chelation). This process of elimination is used to determine how its chemistry and toxicity changes after such treatment. A significant portion of toxicity observed in industrial effluents is often attributed to pH effects. Therefore, pH adjustment is used throughout Phase I to provide more information on the nature of the toxicants. In all cases, the toxicity of treated samples is compared to non-treated samples to determine which approach, if any, reduced toxicity.

One of the most important benefits of the TRE/TIE process is that it incorporates the responses of organisms into the assessment of complex mixtures to determine the identity of the substance(s) responsible for toxicity. Attempts to use chemical screening alone to identify substances responsible for toxicity are typically unsuccessful.

Determining the cause of transient or non-persistent toxicity can be difficult and may require the testing and analysis of a large number of samples.

Case studies using TRE/TIE were presented at the 2009 IOC Workshop by Novak and Holtze (2012).

12.5.2 Investigation of Cause Framework Example of Toxicity Reduction Evaluation and Toxicity Identification Evaluation

The Investigation of Cause Framework is an application of the TRE/TIE approach in the IOC context. It was developed by Hewitt et al. (2003), and is described in further detail in the Pulp and Paper EEM Technical Guidance Document IOC Chapter (Environment Canada 2010). The framework consists of step-by-step questions that follow a tiered approach (Figure 12-1). The questions are defined by a continuum of investigative phases, each providing more information regarding the cause of the effect with

concomitant investments of time and resources. A review of relevant information concerning mine history, process type, process or operational changes, extent and magnitude of effects, and response patterns observed in EEM phases is critical before decisions can be made regarding the initial steps and direction of the IOC. The guidance for addressing these questions has evolved through review of the published literature and the ongoing results of the Pulp and Paper EEM program (Hewitt et al. 2003).

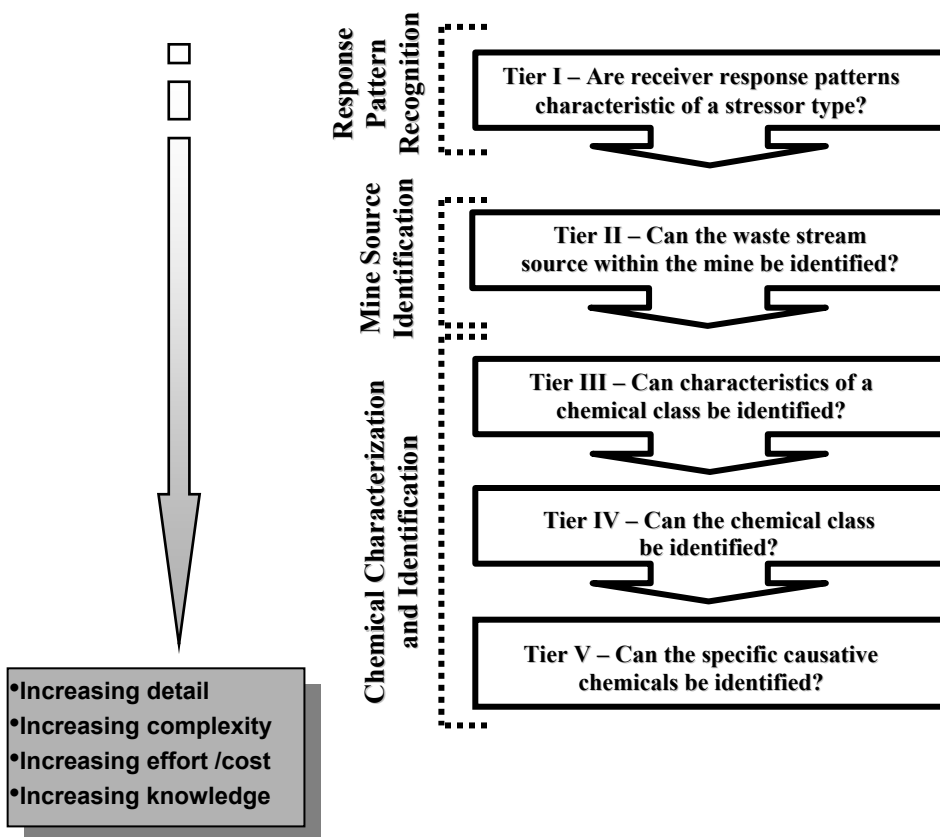


Figure 12-1: Tiered framework for investigation of cause in environmental effects monitoring (Hewitt et al. 2003)

12.5.2.1 Defining Response Patterns

The first step is to define the response patterns. Characterizing the type of stressor(s) based on response patterns will greatly benefit any IOC at the outset, by narrowing the focus and making the best utilization of potentially limited resources. Once the response pattern has been defined, the framework (Figure 12-1) is tiered to identify the stressor for that pattern type.

12.5.2.2 Source Identification & Selective Operation of In-plant Processes

The purpose of source identification is to attempt to specify or isolate specific waste streams within the manufacturing or treatment process that are responsible for the observed effects measured in the receiving environment. Studies should begin with a systematic investigation of individual process streams when looking for the source of effects within the mine.

Determining the source(s) of the effect has several potentially important outcomes, including 1) focusing further investigations to a particular area of the mine for a more detailed inventory of process stream sources, quantities, and waste stream qualities and toxicities; 2) identifying an area of the mine where operations can be reviewed to ensure that “normal operations” are occurring and to eliminate anomalies; 3) evaluating the potential for source treatment and the consequences in terms of final effluent quality; and 4) focusing subsequent detailed investigations of the waste stream source(s), including identification of chemicals.

Generally, acute toxicity tests are performed on different waste streams within the plant. Using the acute toxicity data, it is possible to isolate the waste streams that could be used for longer-term sublethal testing. The focus should be on those effluent constituents that would be carried through primary treatment and affect final effluent quality. Further information can be found in Environment Canada (2010), Rickwood et al. (2006) and Martel et al. (1997). Longer-term testing can occur using a variety of approaches, including mesocosms, laboratory bioassays, partial to full life-cycle studies, etc.

12.5.2.3 Chemical Isolation and Characterization

This step represents a TIE. If identification of waste streams has not provided sufficient information to characterize the effect, the procedures below can be used for isolating and identifying the responsible substances.

Chemical characterization and identification can be very complex. Compound characteristics, compound classes or the compounds themselves can be identified; however, each step becomes more costly and complex. The approaches described in this section are designed to identify specific characteristics of the chemical(s) that are responsible for effects under investigation. Information from field studies, source identification, laboratory studies, and the exposure profile may provide indications of chemical class.

Phase I of the TIE involves characterizing the chemical class: 1) determining the characteristics of the active agents and 2) establishing whether the effect is caused by the same substances. This step utilizes specific methods for isolating active chemicals and proposing structures for their identification. The physicochemical properties of the active substances can be described using effluent manipulations coupled to a bioassay that either duplicates the field effects or is mechanistically linked to them (see section 1.2.6.2.6 of

the Pulp and Paper EEM Guidance Document, IOC Chapter) (Environment Canada 2010). Each test is designed to alter the substances themselves or change their bioavailability so that information on the nature of the substances can be obtained.

Phase II of the TIE involves identifying the chemical class. Active components are further isolated or separated from inactive substances for their identification and confirmation. These methods are specific to the classes of chemicals and utilize bioassay responses (see section 1.2.6.2.6 of the Pulp and Paper EEM Guidance Document, IOC Chapter) (Environment Canada 2010) to evaluate the success or failure of extraction, separation and concentration of bioactive substances. Chemical isolation steps proceed in an iterative fashion, directed by bioassay responses until further isolations are not possible or candidate chemicals are identified. Once there is strong evidence that one or more candidate chemicals are associated with the response, the third phase of the TIE (below) can be initiated.

Phase III of the TIE involves identifying the specific causative chemical. This step involves techniques for confirming that the proposed substances are in fact responsible for the observed toxicity. This is usually accomplished through a weight-of-evidence assemblage of information that collectively establishes the identity of the active compounds, and establishing cause of the effect consistently over time so that amelioration efforts can adequately address the effect. Confirmatory approaches include the following:

- 1) Correlation approach: A strong, consistent relationship between the concentrations of the suspected agents and the bioassay response can be established.
- 2) Symptom approach: Different active substances often produce different symptoms in response. By comparing exposures of the effluent sample to those of pure, suspected active substances, one can obtain further evidence on whether the suspected agents are responsible. Examples of symptoms include species sensitivities, shapes of dose-response curves, and time for the effect to occur.
- 3) Spiking approach: Suspected agents are added to the effluent to determine if a proportional response in the bioassay is obtained.

12.6 Integrated (Weight-of-Evidence) Approach

12.6.1 Using Weight of Evidence

Distinguishing among the cumulative impacts of multiple stressors (which sometimes have confounding effects) requires establishing a definitive causality link to the mine effluent that is being evaluated. Environmental monitoring of aquatic ecosystems is particularly prone to these impediments because these ecosystems often receive multiple interacting effluent discharges. Assessments of monitoring results often rely, in large part, on field monitoring data that can only show correlation (rather than clear cause and effect) between the mine effluent and the observed effect. Establishing a strong causal link, however, can benefit from a weight-of-evidence approach, which is the process of

combining information from a variety of sources, i.e., multiple lines of evidence. The strength of the causal link can be evaluated by using a formalized set of criteria originally developed in the field of epidemiology. These criteria are outlined in Table 12.2, which is based on Fox (1991), Suter (1993), Gilbertson (1997), Beyers (1998), Culp (1999), Culp et al. (2000), and Lowell et al. (2000). In many cases, not all of the criteria may be satisfied or the findings for some criteria may conflict with those for other criteria. In these cases, it can be useful to assign weight to each criterion in terms of its relative importance for a given assessment. In particular, assessments that include evidence from field and laboratory experiments will be stronger than those based on field monitoring alone. It is often not necessary to satisfy all of the criteria to provide a strong causal argument (e.g., see Beyers 1998). The advantages to applying formalized criteria to assessments of the data generated by an EEM program include: 1) helping to tie together diverse assemblages of data on the effects of multiple stressors; 2) more clearly assigning causality; and 3) identifying important informational gaps (i.e., criteria that have not yet been addressed). This kind of approach can help to make assessments of mine effluent effects more rigorous and robust by combining alternative approaches and investigations of cause methods.

Table 12-2: Formalized set of causal criteria forming part of a weight-of-evidence approach for assessment of mining effluent effects

Causal criteria for weight-of-evidence approach
<u>Spatial correlation</u> of stressor and effect along gradient from more to less exposed areas
<u>Temporal correlation</u> of stressor and effect relative to time course of exposure
<u>Plausible mechanism</u> linking stressor and effect
<u>Experimental verification</u> of stressor effects under controlled conditions and concordance of experimental results with field data
<u>Strength</u> : steep exposure/response curve
<u>Specificity</u> : effect diagnostic of exposure to a particular stressor
<u>Evidence of exposure</u> (contaminants or other indicators) in body of affected organisms
<u>Consistency</u> of stressor/effect association among different studies within the region being studied
<u>Coherence</u> with existing knowledge from other regions where the same or analogous stressors and effects have been studied

12.6.2 The Sediment Quality Triad, as an Example of Weight of Evidence

The use of the sediment quality triad (SQT) approach is recommended during investigation of cause of effects on benthic invertebrate communities (Parker and Dumaresq 2002). The AETE (Aquatic Effects Technology Evaluation Program) endorsed the use of the SQT for use on a site-specific basis, but not for “routine” monitoring (ESG 1999).

The SQT originally integrated three components (lines of evidence): sediment chemistry to determine chemical contamination, sediment toxicity bioassays, and benthic community structure to determine the status of resident biota (Chapman 1992). Using the different lines of evidence, statistical correlations can be made and integrated in a weight-of-evidence analysis (qualitative and quantitative), making the data more powerful than if each component were interpreted individually (see Alden 1992; Chapman 1992; Warwick and Clarke 1991).

The traditional SQT, which is based on correlation, can provide definitive conclusions regarding the pollution status of contaminated sediments (e.g., exposure and effects). It thus provides the minimum level of information necessary, i.e., a screening-level risk assessment upon which hypotheses regarding cause can be proposed. In the risk assessment paradigm provided by Chapman and Hollert (2006), causation is addressed after SQT, in which initial studies are relatively simple and serve to either provide definitive conclusions or to indicate gaps where more information is needed. Definitive conclusions regarding causation usually cannot be determined without further studies. The investigation of cause is examined at a higher tier, i.e., a more detailed assessment. Chapman and Hollert suggest four categories of additional lines of evidence (LOE) that can be added to the SQT:

- 1) Direct replacements for or additions to the existing alteration LOE: bottom fish histopathology, crab exposure and health.
- 2) Variations on existing LOE: *in situ* or sediment contact toxicity tests (toxicity LOE); colonization experiments (alteration LOE).
- 3) Additional LOE: crab exposure and health; surface water; bacterial community structure; biomagnification / secondary poisoning; biomarkers.
- 4) LOE to determine causation: habitat morphology (TIEs); effect-directed analyses (EDAs); critical body residues (CBRs); *in situ* sediment contact assays and laboratory toxicity tests; biomarkers; and mechanism-specific endpoints.

Overview of the Sediment Quality Triad Study Approach

- 1) sample collection
- 2) sediment chemistry
- 3) laboratory toxicity testing of sediment

- 4) identification and enumeration of benthic invertebrates in sediment samples
- 5) collection of fish for tissue analysis
- 6) additional analysis
- 7) data analysis and report preparation

1) Sample Collection

Samples are collected from reference and exposure areas or over an exposure gradient, to help delineate the geographic extent of the effect. The total number of sampling locations should be at least ten, which should provide enough data for a robust statistical analysis. Samples at each location should be collected for chemistry, benthos and sediment toxicity.

2) Sediment Chemistry

Chemical analysis of the sediments should include physical characteristics (particle size and total organic content), nutrients (e.g., nitrate, nitrite, phosphorous, ammonia, Total Kjeldahl nitrogen (TKN), general chemistry (e.g., potassium, chloride, sulphate), and metals (total and partially extractable) (ESG 1999). Chemical parameters relevant to the site should be analyzed based on previous information from the site, particularly from water and effluent chemistry. Analyses of redox potential should also be considered.

3) Laboratory Toxicity Testing of Sediment

A battery of three tests is ideal for the triad approach, involving three different test species (MOEE 1997; ESG 1999) (formerly MOEE Ontario Ministry of the Environment and Energy, now MOE Ontario Ministry of the Environment). This provides a safeguard that a biological effect is not missed if one test species is not sensitive to the chemicals present in the sediments. It also provides test organisms from different trophic levels in the ecosystem so that the major relevant trophic groups (and their respective levels of complexity with differing capabilities for metabolizing and depurating chemicals) are represented. The recommended tests use benthic invertebrates and fish, and incorporate different biological parameters (i.e., survival and growth). Any tests with organisms that can incorporate reproduction should also be considered (e.g., *Ceriodaphnia dubia*).

The three recommended tests are:

- *Hyalella azteca* 14-day growth and survival (Environment Canada 1997a)
- *C. dilutus* (formerly *Chironomus tentans* (Shobanov et al. 1999)) 14-day growth and survival (Environment Canada 1997b)
- Fathead minnow 28-day toxicity (Bedard 1992; Van Geest et al. 2011a, 2011b) and bioaccumulation (Van Geest et al. 2011a; 2011b)

4) Identification and Enumeration of Benthic Invertebrates in Sediment Samples

The identification and enumeration of benthic invertebrates should be conducted as detailed in Chapter 4 of this guidance document.

5) Collection of Fish for Tissue Analysis

This step is optional and is useful, in particular in areas where there is a concern about the usability of fish for human consumption. The collection of fish tissue for metals analyses would provide information on whether or not chemicals in resident biota may affect the usability of fish for human consumption. Tissues can be frozen for future analysis (MOEE 1997).

6) Additional Analysis

In order to obtain enough lines of evidence to determine causation, Chapman and Hollert recommend (2006) additional studies from their four categories (discussed in 12.6.2 above). Choices of species to use in toxicity tests and lines of evidence to investigate should be made based on contaminants, receptors and stressors of potential concern.

In the model proposed by Chapman and Hollert (2006), causation can be determined using the following lines of evidence: *in situ* sediment chemistry, benthos colonization, TIE, effect-directed analysis or bioassay-directed fractionation (to identify organometallic contaminants), bioaccumulation / critical body residues, habitat morphology, and sediment dynamics. Choices regarding these lines of evidence will depend on findings of the standard screening level SQT (chemistry, toxicity, and benthic invertebrate analysis).

7) Data Analysis and Report Preparation

The last step in the SQT approach is to statistically analyze the data obtained from the benthic survey, chemical analysis and toxicity tests and integrate this in a weight-of-evidence approach.

12.7 Metal Toxicology and Bioaccumulation

At the 2009 Metal Mining IOC Workshop, two particular publications on metal toxicity (US EPA 2007 and Adams et al. 2011, and the references contained within them) were discussed and should be referred to for further information. The Framework for Metals Risk Assessment (US EPA 2007) provides guidance regarding properties of metals, such as environmental chemistry, bioavailability, and bioaccumulation. As part of this framework, a series of papers related to metals were developed, one of which was McGeer et al. (2004), which discusses bioavailability and bioaccumulation of metals. The definition of bioaccumulation, as defined in McGeer et al. (2004) and US EPA (2007), is the net accumulation of a metal in a tissue or whole organism that results from environmental exposure media such as air, water, soil and sediment, as well as diet.

Bioavailability is defined as the extent to which bioaccessible metals absorb onto or into and across biological membranes of organisms (McGeer et al. 2004; US EPA 2007). Bioaccumulation measures are usually an effective indicator of metal bioavailability. Excess accumulation of some metal species is potentially toxic to organisms. However, it is important to note that different types of organisms will accumulate metals to varying degrees (Adams et al. 2011).

12.7.1 Ecotoxicity of Metals

In the 2002 version of the Metal Mining EEM Guidance Document, it was assumed that bioaccumulation could be used to directly infer the cause of toxicity, based upon invertebrate research (Borgmann and Norwood 1997, 1997a, 1997b) and organic chemical models (Campbell 2012). There have been many advances on the topic of metal bioaccumulation in relation to toxicity, and the current state of the science indicates that the relationship between bioaccumulation and toxicity is confounded by physiology (Adams et al. 2011; US EPA 2007). It is now known that not all metals that accumulate in an organism interact at the site(s) (e.g., biochemical receptors, organelles, cells, organs, tissues) of toxic action (McGeer et al. 2010). Correlations between toxicity and metal bioaccumulation do not always indicate a cause and effect relationship, particularly when the relationship is confounded, as is the case for metal bioaccumulation and toxicity. In addition, extrapolating correlations from single metal exposures in laboratory conditions to complex effluents with a mix of potential contaminants would require extensive validation. Direct correlation between accumulation and toxicity has been observed for a few metals (non-essential, non-regulated metals such as cadmium and thallium) and organometallic forms, and in invertebrates only (McGeer et al. 2012; Borgmann and Norwood 1997a, 1997b, 1999; Borgmann 2000). Bioaccumulation of copper and zinc in invertebrate tissues is less useful because these elements are regulated to varying degrees, although their background concentrations can be high (McGeer et al. 2012).

Metals must first interact with the cell membrane in order to elicit a biological response or accumulate within an organism. Unlike organic contaminants, metals in the aquatic environment (with the exception of organometallics and neutral metal complexes) are unable to cross the cell membrane by simple diffusion; rather, they are transported (facilitated) across the cell membrane (Campbell 1995, 2012). Organism physiology sequesters or detoxifies significant amounts of toxic metals. The biologically or metabolically active portion of metal that is available to contribute to toxicity must do so at the site of action and, further, that total metal is at best a surrogate for the fraction of metabolically active metal at the site of action (McGeer et al. 2010).

Because metals are detoxified once they enter the intracellular environment (Vijver et al. 2004; Luoma and Rainbow 2008), simple predictions of metal-induced toxicity on the basis of metal quotas or burdens (tissue residue approach), as is often done with organic contaminants, are rarely applicable (Campbell 2012). Metals bound to inducible metal-binding proteins such as metallothionein, or precipitated into insoluble concretions consisting of metal-rich granules, can be considered to be a biologically detoxified metal,

as compared to metals in metal-sensitive fractions such as organelles and heat-sensitive proteins (Campbell 2012). A corollary to this model of metal accumulation is that metal tolerance or resistance will be related to the ability of an organism to prevent "inappropriate" metals from binding to sensitive sites. The binding of an "inappropriate" metal to a metal-sensitive site, often termed "spillover" or incomplete detoxification, could be the precursor to the onset of metal-induced stress (Campbell and Hare 2009).

12.7.2 Selenium Toxicity

Selenium is a trace metal that has a high potential risk as a toxicant. Selenium, an essential dietary element, can accumulate in the tissues of oviparous fish through dietary exposure (Palace 2012). Although not toxic to adult fish, at slightly elevated concentrations the element is deposited in the egg, and can induce teratogenicity (Holm et al. 2005). Because the biogeochemistry of selenium is complex, tissue-based criteria (as opposed to aqueous concentration) are generally accepted as the most reliable indicator in potential selenium toxicity in fish (US EPA 2004). The most reliable indicator of potential selenium toxicity is measurement in eggs or ovary, although species-specific toxicity curves are required to achieve accurate risk assessment (Holm et al. 2005). Muscle can be used as a surrogate for egg and ovary; however, correlations between these tissues depend on the partitioning of selenium between muscle and egg, which varies with the reproductive cycle and between species (NAMC 2008). Because of the uncertainty of residency of some highly mobile fish and because of the shorter reproductive cycle of some fish species, reproductive toxicity may not be adequately detected between EEM phases (Palace 2012). Therefore, selenium toxicity may be an important factor to investigate as a source of toxicity in IOC. Much information on selenium toxicity was presented in the Toxicity of Metals and General Aquatic Toxicity streams at the Aquatic Toxicity Workshop (ATW) in 2011.

12.7.3 Biomarkers (Metallothionein)

Biomarkers are tools used for assessing metal-specific exposure. Significant increases in biomarker responses (either over time or relative to reference sites) could indicate that metals are bioavailable in the receiving environment. Metallothionein (MT), a low-molecular-weight metal-binding protein (US EPA 2007), is considered as a biomarker of heavy metal pollution in aquatic environments and has received much attention. In the domain of biomarkers for metals, MT conforms to most of the criteria defined for such tools. For example, MT responds specifically in a dose-dependent manner to changes in ambient levels of a trace metal or of a group of trace metals (Cd, Cu, Zn, Ag). In addition, high levels of MT may be associated with deleterious effects on organisms and populations because these high concentrations are indicative of metal detoxification, but if this capacity is exceeded, then there will be non-specific metal binding to cellular targets of toxicity.

There is no field evidence yet that MT concentrations respond to metals like As, Ni, Pb, and Hg. In addition, the MT biomarker will not track exposure to non-metallic chemicals such as ammonia and cyanide. MT constitutes an excellent biomarker of metal-induced effects; however, the extrapolation of lab data is cautioned (US EPA 2007).

Additional information on MT can be found in Couillard 1997, Couillard et al. 1999, Giguère et al. 1999, Gillis et al. 1999, Laflamme et al. 1998, Perceval et al. 1999, Pinel-Alloul et al. 1999, Wang et al. 1999, and US EPA 2007.

12.7.4 Using Bioaccumulation Methods

For assessing metal toxicity, the Framework for Metals Risk Assessment (US EPA 2007) should be referred to, as it addresses questions related to the methods or tools used to reflect metal bioavailability, and scientifically based approaches used to determine metal bioaccumulation (US EPA 2007). In addition, Adams et al. (2011) summarize the work from a SETAC expert subgroup on metal contaminant bioaccumulation. This subgroup evaluated the potential use of metal tissue residues for predicting effects in aquatic organisms (Adams et al. 2011).

The following hierarchical approach for assessing whether metals are responsible for confirmed effects in metal mining receiving environments was presented by Peter Campbell at the 2009 Metal Mining IOC Workshop (Campbell 2012) (the approach considers metal-organism interactions previously discussed, and would be applicable for non-essential metals):

- *Identify* appropriate biomonitor organisms that live in the exposure and reference areas. Criteria for selecting biomonitor organisms are well established (Phillips and Rainbow 1993). If appropriate indigenous organisms are unavailable, caged organisms may be considered.
- For each biomonitor species, *choose* the appropriate target organ, (e.g., gill or liver, or, for smaller species, whole-body concentrations) (subcellular fractionation approach).
- *Compare* the levels of bioaccumulated metals in specimens collected in the reference and exposure areas. The observation of markedly increased metal levels in the exposed specimens is evidence that one or more of these metals may be responsible for the confirmed effects. Note that by relying on metal bioaccumulation (rather than on the determination of metal concentrations and metal speciation in the receiving environment), one can circumvent the need to estimate the bioavailability of the metals found in the water body receiving the mining effluent.
- To refine the interpretation of the bioaccumulated metal levels, *determine* the subcellular partitioning (subcellular fractionation) of the metals in the target organisms (Campbell and Hare 2009). However, this distinction between biologically detoxified metals and metals that have “spilled over” onto metal-sensitive sites would only be justified for biomonitoring species that demonstrate a

clear threshold response to increasing metal exposure (see discussion in Campbell et al. (2008) and Campbell and Hare (2009)).

Campbell (2012) also described a method to separate the relative importance of individual metals within a mixture. This method compares gene transcription between exposed and reference organisms. It shows some promise as a discriminatory biomonitoring tool to detect and differentiate among metal contaminants (Pierron et al. 2009; Walker et al. 2008).

12.7.5 Methods for Correlating Bioaccumulation and Toxicity

Three methods for correlating bioaccumulation and toxicity are briefly discussed below.

12.7.5.1 Tissue Residue Approach

The use of bioaccumulation as an indicator of toxicity is known as the tissue residue approach (TRA) and it relies on establishing a dose-response relationship to link tissue residues with toxic effects. Its premise is that tissue concentrations are a better surrogate for characterizing toxicity than external exposure concentrations (water, sediment/soil or diet). Meador et al. (2011) indicate that five exposure-based metrics can be used in the TRA, all of which have proportional relationships to each other: external exposure concentration, whole-body concentration, organ concentration, target concentration, and receptor concentration. TRAs, relating accumulation to toxicity, have been developed for some metals (e.g., selenium, copper, methylmercury, cadmium, zinc, arsenic, cobalt, chromium, manganese) in some fish and invertebrates (Borgmann 2000; Tessier et al. 1993), including in *Hyalella* (Adams et al. 2011; Borgmann et al. 2001; Norwood et al. 2007; Borgmann et al. 1991; Borgmann and Norwood 1997a, 1997b; Borgmann and Norwood 1999; Holm et al. 2005; Hodson 1990; Ridal et al. 2010; Meador et al. 2011). The use of the TRA for metals, however, can be problematic (Meador et al. 2011). Studies have shown that bioaccumulation relationships with *Hyalella* may be unique (Adams et al. 2011) and the interpretation of data derived from *Hyalella* studies may not be straightforward (Wang et al. 2004). Adams et al. (2011) concluded that the TRAs for metals other than organometals are not currently supported as they have not led to the development of a generalized approach to estimate effects, and under the circumstances a determination of applicability may require field validation. Sappington et al. (2011) discuss the benefits and limitations of incorporating a TRA into an ecological assessment. Therefore, caution is advised when considering the use of TRAs for IOC, and US EPA (2007), Adams et al. (2011), Sappington et al. (2011) and references herein should be considered for further information. Even though TRAs are a complex means for determining toxicity relationships, there is still a potential use for understanding causes of adverse effects in certain scenarios (e.g., indication of metal exposure to determine which metals are being taken up by organisms).

12.7.5.2 Biotic Ligand Model

Metals in the aquatic environment (with the exception of species such as organometallics and neutral metal complexes) are transported across respiratory and digestive membranes by facilitated transport, involving either membrane carriers or channels (ionic binding to cell ligands, assimilable ligand such as thiosulfate). Such transport is normally a function not of the total dissolved metal but rather of the free metal ion concentration. This is the foundation for the Biotic Ligand Model (BLM) (Campbell 2005, 2012). The BLM is a tool that can be used to quantitatively evaluate how several water chemistry parameters affect the speciation and bioavailability of metals in the aquatic environment (Niyogi and Wood 2004; CCME 2007). The BLM is one example of when metal bioaccumulation can be applied successfully as an indicator of impact. The BLM is based on predictions of the binding of metal at the site of toxic action, and the approach integrates metal interactions along the exposure–uptake–accumulation–toxicity pathway (Di Toro et al. 2000, 2001; McGeer et al. 2010; McGeer et al. 2012). Metal uptake within the BLM is based on estimates of the relative bioavailability of different dissolved forms (species) of metal. These interactions between metal species and the biotic ligand are incorporated into the equilibrium modelling framework. The strength of BLMs is that they simultaneously account for the geochemical speciation as well as the relative binding of metal species (or not) to the site of toxicity. More information on the BLM can be found in Paquin et al. (1999, 2002a, 2002b), CCME (2007), US EPA (2003), Santore et al. (2001, 2002), Niyogi and Wood (2004) and McGeer et al. (2010).

12.7.5.3 Subcellular Fractionation

Subcellular fractionation methods characterize bioaccumulation in a tissue or a whole organism. Using the following processes, homogenization, centrifugation and heat treatment, a distinction can be made between metal in detoxified and metabolically active forms within a cell (Campbell et al. 2008; Campbell and Hare 2009). Subcellular partitioning of bioaccumulated metal has the potential to provide valuable information on metal toxicity. Partitions include metal-rich granules, cellular debris, organelles (e.g., mitochondria, microsomes and lysosomes), heat-sensitive proteins, and heat-insensitive proteins.

Studies have grouped subcellular compartments into metal-sensitive pools such as mitochondria and heat-sensitive proteins (where metal toxicity may occur) and metabolically inactive pools such as granules and heat-insensitive proteins (where metal accumulation is benign).

12.8 Field Study Approaches for Fish and Benthos

Several IOC study tools can be used in the field for fish and benthos studies. Cost effectiveness and availability from consulting, academic or government laboratories should be considered.

12.8.1 Field Study Design

Gradient designs can be used to associate stressor level and geographic distance between outfall. Co-occurrence of contamination with metal effects is one line of evidence in establishing metal effects. Gradients of contamination offer the opportunity to determine co-occurrence of metals with predicted patterns of response (Luoma and Rainbow 2008). Field study designs such as gradient designs and control impact designs are presented in other chapters. Gradient designs are discussed in greater detail in chapters 2, 3, 4 and 9. Other field study designs such as control impact designs are presented in chapters 2 and 4.

12.8.2 Caged Organisms

Caged bivalves are discussed in Chapter 9 (Alternative Methods) of this Technical Guidance Document.

IOC studies can make use of non-mobile fish and invertebrates or caging techniques. Cage bioassays are often limited to the use of older (larger) and less sensitive individuals, because the cage mesh size must be large enough to allow adequate water exchange to prevent cage effects such as fouling, sedimentation and low dissolved oxygen concentration.

Fish have been used in *in situ* caging (Muir et al. 1991). Some debate has existed regarding the use of caged fish, due to concerns regarding interactive effects of stress in captive fish (Courtenay 2002). It was suggested that small-bodied fish with potentially smaller home ranges may be less susceptible to stress and more appropriate for caging (Palace et al. 2005; Bandler et al. 2012). This method may be useful in IOC, particularly where effects of discharges need to be resolved at a small spatial scale, if optimized to reduce stress (verification of caging effect on fish weight).

The caging of amphipods for *in situ* assessments of biological responses to environmental conditions is an established methodology (Grapentine 2012; Borgmann et al. 2007; Couillard et al. 2008; Mulliss et al. 1996; Muir et al. 1991; VanWingaarden et al. 1996; Henry et al. 1994). Experiments can be designed to (a) test for direct effects of effluents or contaminated sediment, (b) link particular metals to toxicity and benthic impairment, (c) identify important contaminant exposure and uptake pathways, (d) quantify stressor-response relationships, and (e) examine effects of natural factors and other anthropogenic stressors.

The method is well developed, not technically difficult, addresses multiple contaminant exposure pathways, is amenable to high experimental replication, incorporates multiple biological endpoints, and is low to moderate in cost. Amphipods are held in screened plastic cages in pre-selected locations in the receiving environment, followed by recovery and analysis for survivorship, growth and/or concentrations of metals in tissues. Cages

can be designed for exposure in the water column, sediment and water column, or pore water, and are deployable over most substrates. Each cage can hold up to 15 amphipods without showing significant mortality after deployment. Food can be added at the start or during the experiment if it is not an experimental factor. Amphipods placed in cages can be held in buckets for several hours or days without mortality, allowing for easy transport. The following endpoints can be measured: the number of survivors, mean individual growth over the exposure period, estimation of reproduction, and whole-body concentrations of relevant elements. Analysis of these endpoints can determine: effects of location or stressor level on amphipod health and populations; metal bioaccumulation and bioavailability; and whether the metal concentrations in tissues are associated with adverse effects for the test organism or species that feed on the amphipods. The strength of this method is to distinguish between sediment and water effects. Standardization may be required for tests involving amphipods other than *Hyaella azteca*.

Other aquatic invertebrates that have been used in field cages include midges, leeches, snails (Henry et al. 1994), and the water flea (*Daphnia magna*) (van Wingaarden et al. 1996).

12.8.3 Mesocosms (Fish and Invertebrate)

Mesocosms are helpful for identifying specific effluent contributing factors and are elaborated in depth in Chapter 9.

12.8.4 Benthic Transplant Devices

The Benthic Transplant Device evolved from spatial and gradient designs in mining and pulp and paper EEM studies conducted in rivers, creeks, reservoirs and complex marine environments over many EEM studies (Thomas 2012). This method employs the use and relocation (transplant) of indigenous benthos populations and associated substrate/habitats, and has been effective in elucidating comparative trends in benthos communities and substrate quality with space and time (Thomas 2012). A key feature of this approach is the ability to tease out subtle biophysical, chemical or benthic differences between sites, such as: changes with exposure/distance measured; benthos abundance (density) /diversity; sediment quality (chemistry/physical measures/appearance); biomass changes; bacterial presence and composition; and benthic deformities/symmetry. Furthermore, this method may contribute to separate historical and current effluent effects from other natural and human-created confounding sources and variability, and in determining the effluent exposure scenario.

12.8.5 Macrophytes, periphyton, phytoplankton, chlorophyll

In addition to benthic invertebrates, phytoplankton, macrophytes and periphyton can be evaluated in IOC. AETE (1997) evaluated each of these biotic categories on the basis of sensitivity, ecological relevance, validity and repeatability, site specificity, and applicability.

12.8.5.1 Macrophytes

AETE (1997) reviewed many studies that have demonstrated a reduction in species diversity and abundance in macrophytes exposed to metal-polluted waters. The absence of macrophytes from impact areas where they would normally be expected can be regarded as an indication of pollution (AETE 1997). Impacts on macrophyte density and biomass could substantially affect the other members of the freshwater community, as they may provide habitat for higher trophic levels within the community. Macrophytes are sedentary and provide a measure of the impacts of the bioavailable fraction of metals over a time-integrated picture (Whitton et al. 1981). Macrophyte community composition appears to be more of a long-term indicator of environmental stress rather than a tool for early detection of potential impacts. Kelly (1988), Haslam (1982), Small et al. (1996), and Sortkjaer (1984) have developed methods of monitoring macrophytes for assessment of pollution. There are several frequently encountered macrophyte species recommended as potential biomonitoring candidates in Canadian environments: *Leptodictyum riparium*, *Potamogeton epihydrus* v. *nuttallii*, *Potamogeton* sp., *Fontinalis* spp., *Potamogeton richardsonii*, *Ericaulon esotangulare*, *Elocharis acicularis* (AETE 1997; Whitton et al. 1981; Crawford and Luoma 1993). At the same time, it should be appreciated that macrophyte populations vary seasonally in most rivers and lakes, with marked seasonal variations in biomass, and are really only available as a biomonitoring species during a relatively short growing season (Hellowell 1986). The use of macrophytes for the marine environment is only feasible in intertidal hard-substrate areas. If this is deemed to be a major habitat that is potentially affected by mining effluents, shoreline quadrat studies of percent coverage and biomass per unit area, major species, and possibly associated epifauna such as amphipods may be appropriate as supportive indicators (AETE 1997).

12.8.5.2 Periphyton

Periphyton is generally non-mobile, and integrates effects of environmental variables in a relatively short time frame. The microalgal component of periphyton is important as both a primary producer in rivers and littoral zones in lakes (Cattaneo and Kalff 1980), and as a food source for invertebrates (Whitton 1984). As a food source for higher trophic levels, periphyton potentially plays an important role in contaminant transfer between trophic levels. Also due to their sedentary nature, periphyton are a good indicator of local conditions, and play a key role at the interface between substratum and surrounding water, perhaps by influencing biogeochemical pathways and dynamics. Organisms within the periphyton mat have relatively high turnover rates. Hence, they are among the first organisms to respond to environmental stress, and among the first to recover. Sensitive species are often replaced by more tolerant species (Austin 1983; Austin et al. 1985). As a result, periphyton communities often reflect the current environmental conditions (Lowe and Pan 1996) and could be used as an early warning indicator.

Periphyton has been used as an indicator for water quality assessments in different systems by a large number of investigators (Chessman 1985; Clements and Kiffney

1994). Various sampling methods have been used, including scraping, brushing or aspiration. Structural measurements of periphyton generally involve counts of organisms, providing information on abundance, species richness and community structure, whereas functional measures study changes in primary production, respiration and detrital processing (Clements 1991). Periphyton communities are often species-rich and spatially compact relative to other aquatic groups, and representative samples can be collected from a few square centimetres of substratum (Lowe and Pan 1996). However, some authors have suggested that there is difficulty obtaining representative and uniform samples because of spatial heterogeneity and difficulty sampling organisms (Weitzel et al. 1979; Kutka and Richards 1986; St-Cyr 2000). Regardless of potential sampling problems, many studies have characterized algal communities in streams and have related abundance of specific taxa to the presence of metals (Clements 1991; Whitton 1984). Studies in lakes have similarly identified shifts in species composition with loss of metal-sensitive species (Austin and Deniseger 1985; Austin et al. 1985; Roch et al. 1985). Lowe and Pan (1996) concluded that benthic algal monitoring data would be most valuable when combined with a suite of additional monitoring data, including physical and chemical measurements and the analysis of other biota of aquatic communities such as invertebrates.

12.8.5.3 Phytoplankton

The structure of phytoplankton communities can be altered due to environmental stressors that affect certain sensitive species and not others. This approach has been used widely to examine the effects of metal pollution and has been applied to long-term monitoring progress at mine sites. Standing crop and biomass (usually measured as chlorophyll *a*) of phytoplankton communities are typically assessed in any study in which primary producers are considered, and have been used as indicators in long-term environmental programs at mine sites (AETE 1997). Consistent patterns of tolerance or intolerance to metal pollution are found in some species, and they may serve as an early warning of metal pollution (AETE 1997). On a long-term basis, biomass does not appear to be a useful measure of the effects of metal contamination on primary producers, since long-term shifts in species composition or reduction in grazing pressure can counteract any short-term depression in algal biomass (Yan 1979). Further, biomass disperses, and density and biomass, tend to be similar between impact and reference areas (AETE 1997).

12.8.5.4 Indicators for Phytoplankton and Periphyton

The measurement of chlorophyll *a*, chlorophyll *b*, and phytoplankton periphyton biomass may be helpful in cases where upstream discharges of nutrients represent confounding influences, or where nutrient enrichment from mine effluent is suspected. The concentration of photosynthetic pigments is often used to estimate phytoplankton biomass.

Phytoplankton and periphyton biomass tends to be patchy in its distribution, and varies with time as blooms develop and waters stratify. It is more suitable for lake than for river settings, although evidence for significant phytoplankton biomass in riverine samples is nevertheless an indicator that nutrients reflecting trophic status upstream may be, and probably are, relevant in assessing nutrient (as opposed to contaminant) influences on the downstream structure of benthic communities, zooplankton and presumably fish as well. Criteria have been established to classify eutrophication based on periphyton and phytoplankton (Dodds and Welch 2000; Dodds 2006). Nutrients and many other factors such as light, temperature, velocity, suspended solids, and physical disturbances may affect periphyton community composition and biomass (Culp et al. 1996; Chambers et al. 2006; Azim and Asaeda 2005; Biggs and Kilroy 2000). Relationships between nutrients and periphyton biomass have been investigated (Dodds and Welch 2000; Chambers et al. 2001).

Frequently, two standard measures of biomass are used: chlorophyll *a*, which is used as an indicator of the total amount of autotrophic organisms in the sample, and ash-free dry mass (AFDM), which is a measure of the total amount of organic material in the sample. The AFDM includes living autotrophic and heterotrophic microorganisms, plus dead periphyton, micro-invertebrates, and possibly terrestrial debris. It is recommended to analyze both of these parameters, because they provide complementary information and they can be combined to form a ratio called the “autotrophic index” (Weber 1973). More information on the use of chlorophyll *a* as an indicator of biomass can be found in Weitzel (1979). Methods for determining AFDM are described in Ridley Thomas et al. (1989) and Aloï (1990).

Chlorophyll *a* can be measured using spectrophotometry, fluorometry, or liquid chromatography, and taking into account chlorophyll *b* and *c*, chlorophyll degradation products (e.g., chlorophyllides, pheophorbides and pheophytins), and turbidity. More information can be found on the analytical methods in Standard Methods for the Examination of Water and Wastewater (method 10200 H) (American Public Health Association (APHA), American Water Works Association (AWWA), and the Water Environment Federation (WEF), 2001).

12.9 Laboratory Methods and Toxicity Tests

12.9.1 Histology and Physiological Parameters

The impacts of contamination can be seen not only at the ecosystem, community, population and individual levels, but also at cellular, subcellular and molecular levels (Peplow and Edwards 2005). At the 2009 Metal Mining IOC workshop (Environment Canada 2012), work was presented in regards to studying histological and physiological parameters to determine the cause of increased weight and size of livers and gonads in fish (Sharpe et al. 2012). Gonad histology was used to assess fish sex and state of maturity (due to small fish size), and to observe asynchronous or abnormal development. Liver histology was used to assess gross pathology (deformities, abnormalities and

parasites), liver mitotic index (indicator of cell proliferation due to tumours or arsenic for example), lipid vacuolation and glycogen content. In addition, physiological parameters such as liver glycogen and triglycerides were measured. These indices taken together gave an indication of gonad and liver function. Some indicators and methods include: diagnostic approaches to hepatic toxicity (Wolf and Wolfe 2005), histopathological approaches (Hinton et al. 1990, 2005; Costa et al. 2009; Werner et al. 2003), and physiological and biochemical indicators (Taylor 2004; Weber et al. 2003; Werner et al. 2003).

12.9.2 Toxicity Tests

Laboratory lethal and sublethal toxicity tests using fish, benthic species as test organisms, and whole sediment, sediment pore water, effluent, and water as test media, can provide a direct determination of toxicity. Toxicity tests can be used in conjunction with many of the monitoring tools and other field tests described elsewhere in this chapter to help interpret results in situations of confounding factors, multiple discharges, or habitat modifications from historical effluent discharges. These toxicity testing methods may also be useful in identifying the cause for the absence of taxa at exposure sites.

There is an array of laboratory test methods available; this chapter refers to some relevant methods, but many more are available from Canada and other international bodies, such as Environment Canada's Biological Methods Division, as well as the US EPA, OECD, EU, WHO and so forth. It is recommended that toxicity testing be conducted by accredited laboratories that follow recognized standard protocols.

Lethal and sublethal toxicity tests have a number of limitations that must be considered. Because laboratory toxicity test methods do not determine toxicity *in situ*, they might not account for toxicity due to overlying water, or for the effects of overlying water chemistry on contaminant toxicity in the testing of sediments or sediment pore water. These problems will be most severe when the chemistry of the overlying water used in the toxicity test differs substantially from *in situ* water, or when conditions in the field or during toxicity tests are such that contaminant concentrations in overlying water are far from equilibrium with the sediment (Parker and Dumaresq 2002).

12.9.3 Fish Toxicity Testing: Fathead Minnow Lifecycle

The fathead minnow whole lifecycle laboratory test can be used for assessing long-term exposure effects of effluents on fish (Parrott et al. 2012; Parrott 2005; Parrott and Bennie 2009). This lifecycle test encompasses all critical windows of exposure during the life cycle of the fish: egg, larvae, developing and mature juvenile, reproduction of adult fish, and survival of the F1 generation.

Lifecycle assays are expensive and lengthy; however, they mimic the effects of real environmental exposures, and provide data that are impossible to obtain with shorter fish

exposure bioassays. The method could provide valuable data in cases where capture of wild fish is difficult (Parrott et al. 2012).

12.9.4 Whole Sediment Toxicity Testing

Sediment toxicity tests may be used to evaluate potential contamination in aquatic environments (Parker and Dumaresq 2002). These tests provide a direct method to determine chemical availability of contaminants in sediment, and can be used as an evaluation tool in conjunction with chemical data and other contaminant analysis, benthic community analysis, and other sediment monitoring tools such as the Sediment Quality Triad (see section 12.6.2). Whole sediment toxicity tests provide direct measurement of toxicity by evaluating the effects of exposure-area sediments on test organisms relative to effects due to control or reference-site sediments. Test results are expressed in terms of survival, growth and reproduction in comparison with a reference site. Effects can also be expressed as the difference (in %) with response in control or reference sediments.

Whole sediment toxicity testing can be used to corroborate that changes in benthic invertebrate communities are due to toxicity of sediments and not other physical or biological factors. Adverse effects on benthic invertebrate community structure could be due to sediment toxicity, but could also be due to other factors (predation, habitat differences, etc.). Concurrent impairment of benthic invertebrate community structure and sediment toxicity implicates the sediment itself as the cause of effects on the benthic invertebrate community. These tests also provide important information for interpretation of field effects in situations where benthic invertebrate community data are inconclusive, or when only pollution-tolerant species are present in both exposure and reference areas.

Batteries of tests can be used to demonstrate sediment toxicity in more than a single species. The use of a battery of tests helps to demonstrate the significance and universality of the sediment toxicity response in different levels of organism complexity (i.e., number of test species responding). The relative responses observed in sediment tests with different species can aid in the identification of the cause of sediment toxicity if information is available in the literature on the relative sensitivity of different species to different metals. Multi-species testing can also aid in explaining the relationship between benthic invertebrate community structure and sediment toxicity. Use of multiple species in toxicity tests allows for more direct comparisons of abundance/absence of benthic invertebrate species with toxicity test results for the same or closely related species. Standard test procedures are available (Bedard et al. 1992; ASTM 1997, 2010; Environment Canada 1997a, 1997b).

The limitation of whole-sediment toxicity tests is that they are laboratory tests that do not determine *in situ* toxicity directly. Furthermore, if a standard overlying water is used (as is commonly done), instead of water collected from the study site, the effects of differences in overlying water quality on sediment toxicity, if present, are not taken into account.

12.9.4.1 Survival and Growth of *Hyalella azteca*

Hyalella azteca is a sediment-dwelling amphipod that has been routinely used in both field and laboratory studies to investigate the source and cause of sediment toxicity (Shuhaimi-Othman et al. 2006; Borgmann et al. 2004; Borgmann et al. 2005a, 2005b; Couillard et al. 2008; Ingersoll et al. 2000; Borgmann and Norwood 2002; Nowierski et al. 2005). Environment Canada has an existing standard method (EPS 1/RM/33: Test for Survival and Growth in Sediment Using the Freshwater Amphipod *H. azteca*) (Environment Canada 1997a). The endpoints for the EPS 1/RM/33 method are survival (lethal toxicity) and growth (by dry weight) at the end of a 14-day toxicity test for sediments and effluent/water. The test may be run as a single or multi-concentration assay (Environment Canada 1997a).

Taylor et al. (2012) modified the EPS 1/RM/33 method to employ both a sediment toxicity test and an aqueous-only toxicity test to separate the effects of current effluent discharge from that of historical accumulation of metals and other toxic substances in the sediment. The sediment toxicity test involves preparing various sediment/water overlay combinations: three sediment types (exposure site, field reference, lab reference) are each overlaid with one of three water types (standard lab dilution water, reference water or receiving water) (nine maximum combinations in total). Using the above combinations, it is possible to determine the toxicity of the effluent or the sediment independently. By using a combination of field and lab sediment and water samples, the study design also takes into account the influence that site-specific water chemistry (e.g., dissolved organic carbon (DOC) in receiver) may have on toxicity (Taylor et al. 2012) (Borgmann et al. 2005c; Borgmann 2002; Environment Canada 1997a, Ingersoll et al. 2000). The aqueous-only test is based on the EPS 1/RM/33 method and a draft method for an aqueous-only method developed by Borgmann et al. (2005). The test design is based on a 14-day static-renewal exposure examining both survival and growth of the test organisms. Tests on field reference water, field sample, and lab dilution water sample are conducted in the absence of sediment, but with an artificial substrate for the test organisms.

Combined results from the study presented at the Metal Mining IOC Workshop (Taylor et al. 2012) indicate the *H. azteca* test method (EPS 1/RM/33) is a quick, cost-effective, readily available screening tool when conducted with benthic surveys and chemical analysis (Sediment Quality Triad approach). It can be used to isolate the biological effects of historical contamination in sediment from current effluent quality, taking into account temporal variability, and toxicity of metals in water and sediment. However, temporal variability in effluent constituents should be taken into consideration when deciding if current effluent quality may impact benthic invertebrate communities.

12.9.4.2 Survival and Growth of the Freshwater Midges *Chironomus dilutus* or *Chironomus riparius*

Chironomids have been used extensively in sediment toxicity tests in the United States and Canada. The standardized test method by Environment Canada is described in EPS 1/RM/32

(Environment Canada 1997b): Test for Growth and Survival in Sediment Using Larvae of Freshwater Midges (*Chironomus tentans* or *Chironomus riparius*). A summary and additional references pertaining to chironomid test procedures is available in Environment Canada (1997b). The EPS 1/RM/32 method exposes ten-second (*C. dilutus* formerly *C. tentans*) or first (*C. riparius*) instar organisms to a sediment sample and bioassay water. At the end of the test, the sediment is sieved and the organisms (dead and surviving) are recovered, and the surviving animals are counted, dried and weighted. The two endpoints calculated are the mean percentage of organisms that survived the exposure and the mean dry weight per surviving animal (Environment Canada 1997b).

12.9.4.3 Survival and Reproduction of the Oligochaete *Tubifex tubifex*

This test is described in ASTM (2010).

Tubifex individuals are exposed to equal amounts of sediment and bioassay water. Sexually mature individuals (aged 8 weeks) are introduced and incubated for 28 days. The production of cocoons indicates reproduction of the organisms. At the end of the test, the sample is sieved and the number of surviving adults, the number of full and empty cocoons, the number of young less than 500 µm and the number of young greater than 500 µm are counted as measurements of survival and reproduction.

12.9.4.4 Survival and Growth of the Mayfly *Hexagenia limbata*

This test is described in the Ontario Ministry of the Environment “Laboratory Sediment Biological Testing Protocol” (Bedard et al. 1992) and in ASTM (2010).

The test uses early instar *Hexagenia limbata* mayfly nymphs that are 3–4 months old and that are laboratory-reared from field-collected eggs. Ten to fifteen *H. limbata* individuals are exposed for 21 days to bioassay water and sediment. Animals are not fed during this time. At the end of the test, the number of surviving animals is counted and the animals are dried and weighed to determine dry weight and to give an indication of growth. Survival and growth in the test sediment are compared statistically to survival and growth in the control.

12.9.5 Metals in Overlying Water in Whole-Sediment Toxicity Tests

Measurement of metals in overlying water of whole-sediment toxicity tests can be used for the quantification of bioavailable metal or the identification of cause of sediment toxicity. Measurement of dissolved metals in the overlying water can provide information on the relative bioavailability, and hence potential toxicity, of metals in sediments. However, the toxicity of metals in sediments is not proportional to total metal concentrations in sediment, and bioavailability can vary greatly from one sediment to another. Metal bioavailability, especially to *H. azteca*, often appears to be primarily due to dissolved metals in the overlying water (Deaver and Rodgers 1996; Warren et al. 1998; Borgmann 2000; Borgmann and Norwood 1999a). If data on metal toxicity in water are available, and if toxicity is known to be primarily due to dissolved metals for the benthic

invertebrate species in question, water concentrations can sometimes be used to infer the cause of toxicity. The use of metal concentrations in overlying water is restricted to static toxicity tests in which the overlying water is not renewed.

The main limitation of this type of measurement is that the toxicity of metals in water can be reduced substantially by complexation with DOC leaching from sediments. Water concentrations resulting in a given level of toxicity may, therefore, be higher in sediment toxicity tests than in water-only tests (Borgmann 2000). This effect can be reduced by using a large water-to-sediment ratio in the toxicity test (Borgmann and Norwood 1999b). Water chemistry may also affect metal toxicity, and metal concentrations in overlying water should be compared to metal concentrations causing toxicity in water-only tests only if the major ion composition of the water used in these tests is similar. This technique will not help identify the cause of toxicity if the solid phase of the sediment contributes significantly to toxicity.

12.9.6 Sediment Pore Water Toxicity Testing

The concentrations of contaminants in pore water may be more highly correlated with toxicity to aquatic organisms (Liber et al. 2011; Cairns et al. 1984; Nebeker et al. 1984; Schuytema et al. 1984; Knezovich and Harrison 1988; Giesy and Hoke 1990) than those in bulk sediments (Patrick et al. 1977; Adams et al. 1985; Shaner and Knight 1985; van de Guchte and Mass-Diepeveen 1988; Di Toro 1989; Ankley et al. 1991; Carr and Chapman 1992). Therefore, toxicity testing using sediment pore water can be an important complement to whole-sediment toxicity testing. The relative importance of pore water and whole sediments as sources of toxic contaminants to aquatic organisms appears to depend on the species of test organism and the type of contaminant (Knezovich and Harrison 1987; Giesy and Hoke 1990; Harkey et al. 1994).

There is a relatively large amount of literature describing toxicity testing with pore water (Burton 1992, 1998). However, it is not as extensive as whole sediments, and there are few standardized methods for toxicity testing of freshwater organisms with pore water. Environment Canada (1992) has two standard methods using pore water for Echinoids (Sea Urchins and Sand Dollars) and luminescent bacteria.

Tests that may be used for pore water toxicity testing are briefly described below. Note that these are standard tests for toxicity testing of water and effluent, and are not unique to pore water testing.

12.9.6.1 Sea Urchin Fertilization Bioassay

The sea urchin fertilization test is described in the sublethal toxicity chapter (Chapter 6, Table 6.2).

12.9.6.2 Microtox Acute Test

The Microtox acute test is used to evaluate the effect of effluent exposure on light production by the naturally luminescent marine bacteria *Vibrio fischeri*. The result is expressed as the concentration where light output is reduced by 25% or 50% (IC₂₅, IC₅₀).

The Microtox test is frequently used on site by many industries. The test is described in the manufacturers' handbook *Microtox Manual: A Toxicity Testing Handbook* (Microbics Corp.). The Microtox test is a rapid-screening bioassay kit, which measures toxic effects on the light output of a standardized luminescent bacterial culture. The main limitation of this method is that the diluent (dilution water) is a saline solution and the test organism is a marine bacterium, with little relevance to Canadian mining environments, as most mine effluents are deposited in freshwater environments.

12.9.6.3 Acute Lethality with *Daphnia magna*

The purpose of this test is to determine the concentration of test water that causes 50% mortality to *Daphnia magna* during a 48-hour exposure period. This standard test uses groups of less than 24-hour-old *D. magna* neonates in a range of test water concentrations. The *D. magna* lethality test is described in detail in the EPS 1/RM/11 (Environment Canada 1990, 1996).

12.9.6.4 Reproduction and Survival Using *Ceriodaphnia dubia*

This test is described in the Sublethal Toxicity Chapter (Chapter 6, Table 6.2).

12.9.6.5 Growth Inhibition of the Alga *Pseudokirchneriella subcapitata*

This test is described in the Sublethal Toxicity Chapter (Chapter 6, Table 6.2).

12.9.6.6 Sediment Pore Water Toxicity Testing Using *Hyaella azteca* and *Chironomus dilutus*

The standard tests for *Hyaella azteca* (Environment Canada 1997) and *Chironomus dilutus* (Environment Canada 1997a), described in sections 12.9.4.1 and section 12.9.4.2, respectively, have recently been used for sediment pore water toxicity testing. Liber et al. (2011) evaluated the hypothesis that pore water metal concentrations are better correlated with sediment toxicity to benthic organisms than whole-sediment metal concentrations. Using *H. azteca* and *C. dilutus* they found that, in some cases for the specific metals examined, their approach has potential to predict sediment toxicity using sediment pore water metals data (Liber et al. 2011).

12.10 Tools for Effluent and Water Quality Analysis

At sites where there are effects on fish or the benthic invertebrate community, there are a number of effluent characterization and water quality monitoring techniques available

that may help in understanding the nature and cause of effects. In addition, effluent characterization of additional mine-related contaminants from other sources, particularly non-point sources, may be appropriate at some sites as part of IOC. In some cases, these techniques could also provide valuable information during investigation of cause monitoring, as supporting environmental data, when determining the extent and magnitude of effects observed in fish or the benthic invertebrate community. The techniques described below are recommended, and mines may use individual techniques or combinations of techniques, as appropriate, to address site-specific questions. More information on effluent characterization and water quality monitoring can be found in Chapter 5.

12.10.1 Measurement of Dissolved Metals during Investigation of Cause

The measurement of total and dissolved metals in effluent and exposure-area water quality samples in IOC would assist in determining which metals might be causing or contributing to the observed effects (Parker and Dumaresq 2002). The rationale for measuring dissolved metals during IOC is based on the theory that it is the metals in the dissolved fraction, particularly the free metal ions, that are the most bioavailable to aquatic organisms. There is qualitative evidence to support this theory, especially in defined synthetic media (i.e., laboratory bioassays). However, this relationship appears to break down in natural waters, particularly in the presence of natural dissolved organic matter (ESG 1999).

The normal procedure for measuring dissolved metal concentrations involves the immediate filtration of the raw water sample through a 0.45-micron filter and then preserving the filtered sample with nitric acid to a pH of less than 2.0 to keep the dissolved metals in solution until the analysis. There are differences in the filter size used, but the 0.45-micron filter is the most commonly used. There are concerns that certain amounts of colloid-bound metals, which are not really dissolved, can pass through this size of filter membrane, and some researchers recommend filter-pore sizes as small as 0.1 microns to minimize this concern (EVS 1997).

In a report by Hall (1998), three filter systems were evaluated (syringe, in-line, and vacuum filters). The recommendations from this study are based upon the contamination contributed by these filters and their ease of use. Before a system is chosen, it is important to evaluate its propensity to become clogged with sample water. Anyone planning on completing dissolved metals analyses is strongly encouraged to consult Hall (1998) prior to sample collection.

Peepers, a dialysis device, and diffusive gradient in thin films can be used for measuring bioavailable metal in overlying water (Liber and Doig 2000). Bioavailable metals can be compared to dissolved metal concentrations, which gives a better prediction of toxicity.

12.10.2 Metal Speciation for Metals of Concern

Metals are neither created nor destroyed by biological or chemical processes. However, these processes are capable of transforming metals from one species to another (US EPA 2007). The form in which a metal (or metalloid) occurs (i.e., the speciation of the metal) can have a significant effect on the bioavailability, bioaccessibility and toxicity of that metal to aquatic organisms (Tessier and Turner 1995; Stumm and Morgan 1996; US EPA 2007). As a result, understanding the speciation of contaminants of concern can be important to understanding the nature and causes of effects on aquatic ecosystems (Parker and Dumaresq 2002). Langmuir et al. (2004) discuss the environmental chemistry of metals and provide examples of how speciation impacts behaviour and effects.

Metal speciation can be measured through direct analysis, or estimated with modelling tools. Two of the most important factors affecting metal speciation are pH and oxidation state. Thus, no matter which method is to be used to assess metal speciation, it is important to take accurate field measurements of pH and dissolved oxygen. In addition, if samples are to be analyzed for metal speciation, it is important to ensure that proper sample collection, preservation and storage procedures are followed.

The speciation, or at least valency determination, of some metals in water can be determined analytically, although such a service may be limited. Analytical methods include methods using ion-exchange, electrochemistry, size exclusion chromatography, voltammetry, X-ray absorption fine structure spectroscopy, inductively coupled plasma mass spectrometry, diffusive gradient thin film, ion-pair reversed phase high performance liquid chromatography, gel integrated microelectrodes, hollow fibre permeation liquid membrane, as well as several methods outlined in section 12.11.5.1 (Partial Metal Concentrations in Sediment). Measurement techniques for the speciation of metals in aqueous solutions have been comprehensively reviewed by Tessier and Turner (1995), Unsworth et al. (2006) and Ekberg et al. (2011). If samples are to be submitted for analysis for speciation, extra care should be taken in sample collection, handling and storage. Such care should be taken because changes in factors such as the dissolved oxygen content of the sample could result in speciation of the sample changing significantly between the time of collection and the time of analysis. It is strongly recommended that the laboratory that will be completing the analyses be contacted in advance of sample collection to identify appropriate procedures.

A number of computer programs are available for the modelling of chemical speciation of metals in water, such as the following:

- MINTEQA2 (Allison et al. 1991)
- MINEQL⁺ (Schecher and McAvoy 1992, 1994)
- Windermere Humic Aqueous Model (WHAM) (Tipping 1994, 1998)
- VMINTEQ (Gustafson 2004)
- Chemical Equilibria in Soils and Solutions (CHESS) Model (Santore et al. 1998; Meyer et al. 1999)

- MINEQL+ in the BLM by McGeer et al. (2000) (Paquin et al. 2002a, 2002b)
- BIOCHEM ORCHESTRA (Vink and Meeussen 2007)
- TRANSPEC (Bhavsar et al. 2008)
- PHREEQCI (Parkhurst and Appelo 2000)
- FITEQL (Herbelin and Westall 1999)
- TICKET (Farley et al. 2008)

Since metal speciation is dependent on a number of factors, all of these programs require data on: pH, alkalinity, hardness, major cations and anions, ionic strength, total dissolved solids, DOC, total dissolved metal, sulphide, unique anthropogenic inputs such as EDTA, and concentrations of aluminum and iron (III) (EVS 1997). These programs can be used to predict the forms and concentrations of metals in effluent or water, and to predict toxicity (EVS 1997). Potential sources of error are described in EVS 1997.

12.10.3 Measurement of Reagents & Reagent By-products used in Processing

A wide range of chemical reagents are used in ore processing, and these reagents, or the by-products of the decomposition of these reagents, may occur in mine effluent. As a result, reagents and reagent by-products may occur in receiving environments. Analyses for reagents and reagent by-products may be helpful in cases where observed effects cannot be attributed to metals and other parameters monitored regularly as part of the EEM program at a site. Several particular reagents and by-products that were discussed during the 2009 Investigation of Cause Workshop for Metal Mining (Environment Canada 2012), are discussed briefly below.

Cyanide species may be present in effluents or as contaminated solid tailings. The toxicity of cyanide is related to its speciation. The free cyanide form (HCN, CN⁻) is classified as the most toxic because of its high metabolic inhibition potential, whereas the metal-cyanide complexes (e.g., [Fe(CN)₆]³⁻, [Fe(CN)₆]⁴⁻) are considered less toxic (Shifrin et al. 1996). Zagury (2012) showed that the more reactive cyanide species initially associated with solid tailings degrade primarily due to volatilization, leaching and bacterial degradation. The higher proportion of stable cyanide species observed in aged tailings probably results from early dissociation of weak to moderately strong complexes, possibly during weathering (e.g., breakdown in the presence of UV light).

Xanthates are commonly used as collectors of sulphide ores through flotation. Xanthates were reported in effluents and their receiving waters in concentrations up to 4.0 mg/L. These concentrations are sufficient to cause potential toxicity, given that xanthates exhibited toxicity (as measured by IC₂₅) ranging from 0.5 mg/L to 3 mg/L (Vigneault et al. 2012).

Thiosalts (thiosulphate, trithionate, tetrathionate, and other polythionates) are generated as a result of the flotation of sulphidic ore. Vigneault et al. (2012) reported that

thiosulphate was generally more toxic than tetrathionate. Sensitivity to thiosalts ranged from no responses for rainbow trout to an IC_{25} of 59.4 mg S_2O_3 / L for *C. dubia* (reproduction). The acute toxicity of thiosulphate to *D. magna* (lethality) was also greater than tetrathionate ($EC_{50} \sim 300$ and 750 mg/L, respectively) (Vigneault et al. 2012). Thiosulphate has been reported in mine effluent concentrations of 700 mg/L, thus in sufficient quantity to cause toxicity (Vigneault et al. 2012).

Selenium is mobilized to nearby watersheds during the mining and smelting of copper, lead, mercury, silver, uranium, and zinc ores. Selenium is a teratogen at elevated levels. Determining the aqueous concentration of selenium in effluents, wastewater and receiving water may determine whether selenium is present at elevated concentrations. Selenium biogeochemistry is complex, however, and more in-depth studies of selenium toxicity in fish tissue may yield more valuable results (Palace 2012).

Other process reagents and wastewater treatments likely to be discharged in concentrations sufficient to cause toxicity include Magnafloc, Nalmet, and lime (Vigneault et al. 2012).

12.10.4 Monitoring of Flow and Loadings in the Exposure Area

During IOC study, measuring water flow in the exposure area could help with data interpretation and understanding of the dilution ratio, mixing, mass balance, fate and effects of contaminants and causal relationships. It may be useful to evaluate the degree of exposure of communities to the mining effluent and its contaminants over a longer period of time and in a variety of conditions.

Assessments of loading rates are essential to mass balance calculations. Assuming complete mixing, the principal statement for mass balance at an industrial outfall is:

$$\begin{aligned} \text{Mass or loading rate of substance upstream} + \text{Mass rate added by outfall} \\ = \text{Mass rate of substance immediately downstream of the} \\ \text{outfall} \end{aligned}$$

Since the loading rate is the product of flow and concentration, the mass balance is given by:

$$Q_0 s_0 + Q_e s_e = Q s$$

Where Q_0 , Q_e and Q = flow rates upstream, effluent discharge rate, and downstream, respectively; and
 s_0 , s_e and s = concentration upstream, in effluent, and downstream, respectively.

A similar statement can be made for the balance of flows:

$$Q_0 + Q_e = Q$$

Upstream conditions of flow and concentrations are often known or can be readily measured. Effluent characteristics are usually known in greater detail. Solving for the downstream concentration, s , yields:

$$s = \frac{Q_o s_o + Q_e s_e}{Q}$$

The downstream concentration is therefore dependent on the upstream and downstream flows and the concentration of upstream and effluent chemical inputs. If the upstream concentration is zero, the downstream concentration, s , will equal the effluent concentration reduced by the ratio of effluent flow to total river flow. This is commonly known as a dilution effect.

The above equations apply to the simplest scenario immediately downstream of the outfall. Mass balance computations become more complex as one moves further downstream of the outfall and contributing factors of tributaries, multiple point sources, distributed sources, as well as dispersion, chemical settling, and chemical decay effects are considered. Useful references for formulating appropriate mass balance equations include Thomann and Mueller (1987), McCutcheon and French (1989), and Henderson-Sellers and French (1991).

12.11 Tools to be Considered for Sediment Analysis

12.11.1 Sediment Monitoring as a Tool for Investigation of Cause

At sites where there are effects in the benthic invertebrate community, there are a number of sediment monitoring techniques that may help in understanding the nature and cause of effects (Parker and Dumaresq 2002). The techniques described below are recommended. Not all of these techniques need to be used; possibly one or few techniques should be applied to investigate the cause of an effect.

12.11.2 Sediment Mass Transport

In certain stream systems, an understanding of sediment mass transport characteristics can be important, particularly to understanding the origin and fate of sediments. The morphology of alluvial stream and river channels with mobile beds is believed to be determined by storm events. Thus, the pool-and-riffle sequence characteristics of such

riverbeds, and the corresponding sediment particle size distributions within the channels, may have resident-times that are short relative to the EEM program. In these circumstances, appropriate characterization of stream channel morphology, and in particular of its more dynamic elements, may be important information in determining the cause of effects on benthic invertebrate communities. It may be important to know, for example, whether annual deposition and subsequent removal of fine-grained material is a regular feature of the study area, and whether ecologically significant movement of relatively coarse bed material (gravel and cobbles) is taking place on time scales of a few years or less.

Much of the evidence relating channel morphology to sediment transport and basin hydrology has been gathered in the United States from river basins that yield relatively large amounts of suspended matter. In glaciated regions typical of much of Canada, alluvial sediments in stream channels dominated by material of glacial origin may be more stable in relation to storm events. Further, in the Canadian Shield in particular, the abundance of lakes and wetlands, particularly in first- and second-order catchments, may help to buffer the effects of storm events. Thus, comparison of results with results of studies in the United States may be of limited value.

Direct observations of suspended sediment transport and bedload movement are difficult, because the key events are of short duration and occur at times of stream flows that prevent direct measurement due to hazardous conditions. However, there are a number of qualitative indicators of the dynamic nature of river channels that can be employed. Photographic records of peak run-off events that involve bankfull discharge (e.g., with a return period of the order of a year and greater) can be used to provide qualitative evidence of the importance of suspended sediment transport and deposition. Another method is to identify boulders or cobbles with colour codes, and periodically observing the extent of their displacement. It is suggested that the mine operator use a combination of methods to describe the physical setting of such dynamic river systems.

12.11.3 Sediment Depositional Rate & Sediment Dating for Historical Trends

An understanding of the sediment depositional rate can be important in understanding influences on sediment chemistry, and in particular the relative influence of a mining operation.

Depositional rates may be estimated quantitatively using sediment traps, but representative sampling may be difficult, and it may not be possible, using this method, to account for the impact of storms or other significant hydrological events. Because of these difficulties, direct measurement of sediment deposition rates should be reserved for exceptional circumstances. Long-term depositional rates can also be estimated quantitatively using sediment coring.

The relative influence of the mining operation can also be estimated in a more qualitative

manner in gravel-bed streams or in lake sediments dominated by soft organic matter. In such cases, the presence of fine mineral-derived sediments would likely indicate some influence from the mine. In other environment types, this method may not be effective, since the deposition of natural mineral-derived fine sediments would mask any influence from the mine. See Håkanson and Jansson (1983) for guidance on lake sedimentology.

Precise dating of sediments, combined with an inventory of the remains of certain organisms and plant material (e.g., diatoms, zooplankton, insects), provide a chronology of changes that often can be linked to the period of anthropogenic influence (Frey 1998). Isotope dating of sediment cores has been used to assess geochronology of year over year (Weech et al. 2012) (and multiple years to decades) sediment deposition rates to tease out confounding factors, such as historical effects. Isotopes such as carbon-14, uranium-234/238, lead-210, cesium-137 and beryllium-10 can be assessed (Cohen 2003; Ritchie and McHenry 1990; Zapata 2003; Bierman and Nicols 2004; Mabit et al. 2008). See Frey (1998) and Cohen (2003) for detailed information on paleolimnology.

12.11.4 Sediment Coring

Core samplers are often used to collect sediment profiles for the determination of the vertical distribution of sediment characteristics (Parker and Dumaresq 2002). Corers are also generally preferred where maintaining the integrity of the sediment profile is essential, because they are considered to be the least disruptive. Core samplers should be used where it is important to maintain an oxygen-free environment below the surficial sediment, to minimize oxidation (EPS 1/RM/29: Environment Canada 1994). A range of sediment coring devices is available. Although core samplers have the advantage of collecting minimally disturbed, intact sediment samples from surficial sediments (upper 15–30 cm) and deep sediments (> 30 cm deep), there are few that function efficiently in substrates with large amounts of sand, gravel, clay or till. Note that the Environment Canada document *Guidance Document on Collection and Preparation of Sediments for Physicochemical Characterization and Biological Testing* (EPS 1/RM/29: Environment Canada 1994), and ASTM (1992), contain extensive guidance on sediment coring, including sediment coring procedures, advantages and disadvantages of sediment corers, and transport and manipulation of collected samples.

12.11.5 Sediment Chemistry

12.11.5.1 Partial Metal Concentrations in Sediment (Partial Extraction), Sequentially Extracted Metals, SEM/AVS Ratios

Although total metal concentrations may not be directly related to biological availability and toxicity, many sediment quality guidelines are currently based on total metal concentrations (Parker and Dumaresq 2002). A variety of methods have been used to predict the biological effects of metals from metal-contaminated sediments. These include the normalization of sediments for particle size, organic content, or extractable

fraction of metals using AVS (acid volatile sulphides) and SEM (simultaneously extracted metals) (Parker and Dumaresq 2002).

It is generally thought that a particular chemical form of an element determines its behaviour, biological availability and potential toxicity, rather than the total concentration in sediments. Specific chemical forms can be measured in these ways:

- a) by direct instrument techniques;
- b) directly by sequential digestion of sediments; or
- c) by predicting levels through thermodynamic modelling.

Direct instrument techniques include: X-ray photoelectron spectrometry; scanning electron microscopy / X-ray microanalysis; secondary ion mass spectrometry; and Auger electron spectrometry (Parker and Dumaresq 2002). These methods have been applied to geochemical studies and for mineral exploration.

The relative strength of association between metals and particles can be assessed by single or sequential extraction or sediment-digestion methods. Weak acids or chelating agents (e.g., EDTA) and reducing agents may be used to differentiate between different chemical forms. Sediment fractions can be operationally defined (e.g., ferromanganese oxyhydrides) depending on the digestion method used. The recent AVS concept assumes that metal concentrations in pore water of anoxic sediments are controlled by sulphides. AVS are extracted by the cold-acid purge and trap technique. SEMs represent the portion of total metals released during AVS dissolution. The SEM/AVS ratio is sometimes used to characterize metal availability. When the SEM fraction exceeds the AVS fraction (e.g., SEM/AVS ratio > 1), the free metal may be present in the pore water at levels adequate to cause acute toxicity. However, toxicity cannot be predicted; only non-toxicity can be predicted (Parker and Dumaresq 2002).

The limitation of these analyses is that anoxic sediment samples must be carefully collected and stored to prevent oxidation. A description of sediment collection methods can be found in Warren et al. (1998).

12.11.5.2 C/N Ratio for Marine Sediment

Effects on the benthic invertebrate community may occur as a result of organic enrichment in sediments. To determine if organic enrichment is contributing to effects, a combination of measurement techniques should be used in the marine environment. The measurement of TOC provides an indication of organic enrichment. Measuring sediment Eh (redox) provides an indication of the oxygen conditions in marine sediments. Measuring sulphides in the sediment provides an indication as to whether the breakdown of organic sediment material is occurring (Environment Canada 1994).

Measuring the carbon to nitrogen ratio (C/N ratio) in marine sediments should provide an indication of the source of the organic enrichment. If the organic enrichment is a result of

land-based sources (e.g., municipal sewage, and pulp and paper effluent), the C/N ratio will be higher (Hargrave et al. 1995). If the organic enrichment is a result of the natural source such as the breakdown of marine aquatic plants, the C/N ratio will be much lower. Therefore, if a benthic invertebrate community study indicates an effect on the benthic invertebrate community, and there is evidence that the effect could be due to organic enrichment (elevated TOC, elevated Eh), determining the marine sediment C/N ratios can help identify the source of the organic loading to that ecosystem.

12.12 Sediment Pore Water Analysis

Fine-grained surface sediments in lakes typically contain 90–95% water (Adams 1991). Some of this water is bound to the crystalline lattice of minerals in the sediments, but most of the water simply occupies the space between sediment particles. This water is referred to as pore or interstitial water. The intimate association of this water with the surface of sediment particles results in reactions between the particles and the water that approach equilibrium. The partitioning of contaminants in sediments between the particulate and water phases depends to a large extent on the amount of organic carbon, sediment particle size, the chemical form of the contaminants, and the physiochemical environment (e.g., pH, temperature, redox potential, sorption/desorption properties of sediments, or the equilibrium between the solid and liquid phases) (Parker and Dumaresq 2002). The dynamics of these processes are not well understood; however, it is generally assumed that concentrations of most substances in the pore water approach equilibrium with the solid phase and its associated contaminants, and that metals in pore water largely represent the biologically available fraction in sediments (Parker and Dumaresq 2002). Consequently, pore water has been collected for toxicity testing to approximate the relative toxicity of contaminated sediment, and/or to assess contaminant levels.

The nature of sediments at the study site can largely influence the usefulness of pore water measurements. Sediments that are either very coarse-grained, or hard, compacted clays, will not likely have pore waters that are significantly contaminated (Burton and Pitt 2002). Therefore, sampling of pore waters should be restricted to sediments ranging from sandy to non-compacted clays.

If sediments are anoxic (most depositional sediments are below 2 cm in depth), all steps involved in sample collection and processing should be conducted in an inert atmosphere or with limited exposure to prevent oxidation and subsequent sorption/precipitation of reduced metal species if metal speciation is of interest. When anoxic sediments are exposed to air, volatile sulphides may also be lost, which may increase the availability (and toxicity) of sulphide-bound metals. Finally, pore water samples undergo rapid chemical changes, giving a storage life of only hours to days. A common device for sampling sediment pore water is the dialysis cell, also known as a peeper (Doig and Liber 2000). Field collection using peepers or suction devices are the most accurate methods to obtain representative samples, because it is less likely to alter the *in situ* chemistry of the pore water and is recommended for geochemical investigations (Burton and Pitt 2002; US EPA 2001). Laboratory methods that allow for extraction of greater volumes of water

are preferred when samples are being collected for toxicity testing, including centrifugation, pressurization or suction. The use and advantages of minipeepers for laboratory sediment toxicity tests are discussed in Doig and Liber (2002).

For information on field (*in situ*) and laboratory methods for collection of pore water, see Environment Canada's *Guidance Document on Collection and Preparation of Sediments for Physiochemical Characterization and Biological Testing* (EPS 1/RM/29), Environment Canada (1994), US EPA (2001), Doig and Liber (2002), and Burton and Pitt (2002).

For toxicity tests on pore water, see section 12.9.6.

12.13 References

Adams DD. 1991. Sampling Sediment Pore Water, pp. 203-202, in: CRC Handbook of Techniques for Aquatic Sediments Sampling, Mudroch A and MacKnight SD (eds.) CRC Press, Inc. Boca Raton, FL, 210 p.

Adams WU, Kimerle RA and Mosher RG. 1985. Aquatic Safety Assessment of Chemicals Sorbed to Sediments, pp. 429-453, in: Aquatic Toxicology and Hazard Assessment, Cardwell RD, Purdy R and Bahner RC (eds.), American Society for Testing and Materials, Philadelphia, PA.

Adams WJ, Blust R, Borgmann U, Brix KV, DeForest DK, Green AS, Meyer JS, McGeer JC, Paquin PR, Rainbow PS and Wood CM. 2011. Utility of tissue residues for predicting effects of metals on aquatic organisms. *Integr. Environ. Assess. Manag.* 2011: 7(1):75-98.

Alden III RW. 1992. Uncertainty and Sediment Quality Assessments: I. Confidence limits for the triad. *Env. Tox. Chem.* 11: 637-644.

Allison JD, Brown DS and Novo-Gradac KJ. 1991. MINTEQA2/PRODEFA2, A geochemical assessment model for environmental systems: version 3.0 user's manual. U.S. Environmental Protection Agency, Athens, GA. EPA/600/3-91/021.

Aloi JE. 1990. A critical review of freshwater periphyton field methods. *Canadian Journal of Fisheries and Aquatic Sciences* 47(3): pp. 656-670.

American Public Health Association (APHA), American Water Works Association (AWWA), and the Water Environment Federation (WEF). 2001. Standard Methods for the Examination of Water and Wastewater, Method 10200H (Chlorophyll), 2001.

Ankley GT, Schubauer-Berigan MK and Dierkes JR. 1991. Predicting the Toxicity of Bulk Sediments to Aquatic Organisms with Aqueous Test Fractions: Pore Water Versus Elutriate, *Environ. Toxicol. Chem.* 10: 1359-1366.

Aquatic Effects Technology Evaluation (AETE) Program. 1997. Technical Evaluation of Monitoring Methods using Macrophytes, Phytoplankton, and Periphyton to Assess the Impacts of Mine Effluents on the Aquatic Environment. AETE Project 2.3.2.

Aquatic Toxicity Workshop (ATW). October 2-5, 2011, Winnipeg, Manitoba, Canada.

Austin A. 1983. Evaluation of changes in a large oligotrophic wilderness park lake exposed to mine tailings effluent for 14 years: The periphyton. *Natur. Can.* 110:119-134.

ASTM (American Society for Testing and Materials). 1992. E 1391-90, Standard Guide for Collection, Storage, Characterization and Manipulation of Sediments for

Toxicological Testing, pp. 1134-1153, in: 1992 Annual Book of ASTM Standards, Vol. II.04, Section 11, Philadelphia, PA.

ASTM (American Society for Testing and Materials). 1997. Standard test methods for measuring the toxicity of sediment-associated contaminants with fresh water invertebrates (Annex 3, *Hexagenia spp.*, and Annex 4, *Tubifex tubifex*). Volume 11.05, Standard E1706-95b.

ASTM (American Society for Testing and Materials). 2010. Standard Test Method for Measuring the Toxicity of Sediment-Associated Contaminants with Freshwater Invertebrates. ASTM Method E 1706-05(2010).

Austin A. 1993. Evaluation of changes in a large oligotrophic wilderness park lake exposed to mine tailing effluent for 14 years: The periphyton. *Natur. Can.* 110: 119-134.

Austin A and Deniseger J. 1985. Periphyton community changes along a heavy metal gradient in a long narrow lake. *Environ. Exp. Bot.* 25: 41-52.

Austin A, Deniseger J and Clark MJR. 1985. Lake algal populations and physico-chemical changes after 14 years input of metallic mining wastes. *Water Res.* 19: 299-308.

Azim ME, Verdegem MCJ, van Dam AA and Beveridge MCM. 2005. Periphyton. Ecology, Exploitation and Management. CABI Publishing, Cambridge, MA.

Bandler C, Baron CL and Palace VP. 2012. An update on the use of caged fish for Environmental Effects Monitoring, in: Environment Canada. 2012. Environmental Effects Monitoring Investigation of Cause Workshop for Metal Mining: Proceedings, Gatineau, QC.

Bedard D, Hayton A and Persaud D. 1992. Ontario Ministry of Environment laboratory sediment biological testing protocol. Water Resources Branch, Ontario Ministry of Environment, Toronto, ON, 26 pp.

Beyers DW. 1998. Causal inference in environmental impact studies. *J. N. Am. Benthol. Soc.* 17: 367-373.

Bhavsar SP, Gandhi N and Diamond ML. 2008. Extension of coupled multispecies metal transport and speciation (TRANSPEC) model to soil. *Chemosphere*: 914-924.

Bierman PR and Nichols KK. 2004. Rock to sediment - slope to sea with ^{10}Be - rates of landscape change. *Annual Review of Earth and Planetary Sciences*; 32 Palo Alto: Annual Reviews Inc., 2004, 215-255.

Biggs BJF and Kilroy C. 2000. Stream Periphyton Monitoring Manual. Prepared for the New Zealand Ministry of Environment. National Institute of Water and Atmospheric Research, Christchurch, NZ.

Borgmann U, Norwood WP and Babirad IM. 1991. Relationship between chronic toxicity and bioaccumulation of cadmium in *Hyalella azteca*. Can. J. Fish. Aquat. Sci. 48: 1055–1060.

Borgmann U and Norwood WP. 1997a. Toxicity and accumulation of zinc and copper in *Hyalella azteca* exposed to metal-spiked sediments. Can. J. Fish. Aquat. Sci. 54: 1046–1054.

Borgmann U and Norwood WP. 1997b. Identification of the toxic agent in metal-contaminated sediments from Manitouwadge Lake, Ontario, using toxicity-accumulation relationships in *Hyalella azteca*. Can. J. Fish. Aquat. Sci. 54: 1055–1063.

Borgmann U and Norwood WP. 1999. Assessing the toxicity of lead in sediments to *Hyalella azteca*: the significance of bioaccumulation and dissolved metal. Can. J. Fish. Aquat. Sci. 56: 1494–1503.

Borgmann U. 2000. Methods for assessing the toxicological significance of metals in aquatic ecosystems: Bioaccumulation-toxicity relationships, water concentrations, and sediment spiking approaches. Aquat. Ecosyst. Health Management 3(3): 277–289.

Borgmann U and Norwood WP. 1999a. Predicting the toxicity of lead in sediments to *Hyalella azteca*: the significance of bioaccumulation and dissolved metal. Can. J. Fish. Aquat. Sci. 56: 1494–1503.

Borgmann U and Norwood WP. 1999b. Sediment toxicity testing using large water-sediment ratios: an alternative to water renewal. Environ. Poll. 106: 333–339.

Borgmann U. 2002. Toxicity test methods and observations using the freshwater amphipod, *Hyalella*. Environment Canada, National Water Research Institute, Burlington/Saskatoon, NWRI Contribution No. 02-332.

Borgmann U, Norwood WP, Reynoldson TB and Rosa F. 2001. Identifying cause in sediment assessments: Bioavailability and the sediment quality triad. Can. J. Fish. Aquat. Sci. 58: 950–960.

Borgmann U, Norwood WP and Dixon DG. 2004. Re-evaluation of metal bioaccumulation and chronic toxicity in *Hyalella azteca* using saturation curves and the biotic ligand model. Environ. Pollut. 131(3): 469–484.

Borgmann U, Grapentine L, Norwood WP, Bird G, Dixon DG and Lindeman D. 2005a. Sediment toxicity testing with the freshwater amphipod *Hyaella azteca*: relevance and application. *Chemosphere* 61 (11): 1740-1743.

Borgmann U, Ingersoll CG, Mathyk S and Lennie-Misgeld P. 2005b. Draft biological test method: Test for survival in 10-day water-only exposures using the freshwater amphipod, *H. azteca*, with emphasis on detection of toxicity due to ammonia.

Borgmann U, Ingersoll CG, Mathyk S and Lennie-Misgeld P. 2005c. Draft biological test method: Test for survival in 10-day water-only exposures using the freshwater amphipod, *H. azteca*, with emphasis on detection of toxicity due to ammonia.

Borgmann U, Couillard Y and Grapentine LC. 2007. Relative contribution of food and water to 27 metals and metalloids accumulated by caged *Hyaella azteca* in two rivers affected by metal mining. *Environmental Pollution* 145: 753-765.

Bosker T and Munkittrick K. 2009. Often overlooked: Biological QA/QC. *Integrated Environmental Assessment and Management* 5(3): 489-91.

Brodeur JC, Sherwood G and Hontela A. 1997. Impaired cortisol secretion in yellow perch (*Perca flavescens*) from lakes contaminated by heavy metals: in vivo and in vitro assessment. *Can. J. Fish. Aquat. Sci.* 54: 2752–2758.

Brodeur JC, Daniel C, Ricard AC and Hontela A. 1998. *In vitro* response to ACTH of the interrenal tissue of rainbow trout (*Oncorhynchus mykiss*) exposed to cadmium. *Aquat. Toxicol.* 42: 103–113.

Burton GA Jr. (ed.). 1992. *Sediment Toxicity Assessment*, Lewis Publishers Inc. Chelsea, MI, 457 pp.

Burton GA. 1998. Assessing aquatic ecosystems using pore water and sediment chemistry. Report prepared for AETE program, CANMET, Natural Resources Canada, Ottawa, Ontario.

Burton GA and Pitt RE. 2002. *Stormwater Effects Handbook. A Toolbox for Watershed Managers, Scientists, and Engineers*. CRC Press, Boca Raton, Florida. 875 pp.

Cairns MA, Nebeker AV, Gakstatter JH and Griffis WL. 1984. Toxicity of Copper-spiked Sediments to Freshwater Invertebrates, *Environ. Toxicol. Chem.* 3: 435-445.

Campbell P and Torgerson T. 1980. Maintenance of iron meromixis by iron redeposition in a rapidly flushed monimolimnion. *Can. J. Fish. Aquat. Sci.* 37(8): 1303-1313.

Campbell PGC. 1995. Interactions between trace metals and organisms: a critique of the free-ion activity model. In *Metal Speciation and Bioavailability in Aquatic Systems*. Edited by Tessier A and Turner DJ. Wiley & Sons, Chichester, UK, pp. 45-102.

Campbell PGC. 2012. Ecotoxicology of trace metals in the aquatic environment, in: Environment Canada. 2012. Environmental Effects Monitoring Investigation of Cause Workshop for Metal Mining: Proceedings, Gatineau, QC.

Campbell PGC, Kraemer LD, Giguère A, Hare L and Hontela A. 2008. Subcellular distribution of cadmium and nickel in chronically exposed wild fish: inferences regarding metal detoxification strategies and implications for setting water quality guidelines for dissolved metals. *Human Ecol. Risk Assess.* 14: 290-316.

Campbell PGC and Hare L. 2009. Metal detoxification in freshwater animals. Roles of metallothioneins. In *Metallothioneins and Related Chelators*. Edited by Sigel A, Sigel H and Sigel RKO. Royal Society of Chemistry, Cambridge, UK, pp. 239-277.

Canadian Environmental Protection Act, 1999. S.C., 1999, c. 33.

Canfield TJ, Kemble NE, Brumbaugh WG, Dwyer FJ, Ingersoll CG and Fairchild JF. 1994. Use of Benthic Invertebrate Community Structure and the Sediment Quality Triad to Evaluate Metal-Contaminated Sediment in the Upper Clark Fork River, Montana. *Environmental Toxicology and Chemistry* 13(12): 1999-2012.

Carr RS and Chapman DC. 1992. Comparison of Solid-phase and Pore-water Approaches for Assessing the Quality of Marine and Estuarine Sediments. *Chem. Ecol.* 7: 19-30.

Cattaneo A and Kalff J. 1980. The relative contribution of aquatic macrophytes and their epiphytes to the production of macrophyte beds. *Limnol. Oceanogr.* 25: 280-289.

[CCME] Canadian Council of Ministers of the Environment. 1999. Canadian environmental quality guidelines. Chapter 4: Canadian water quality guidelines for the protection of aquatic life. Hull (QC): Canadian Council of Ministers of the Environment. Available from: <http://ceqg-rcqe.ccme.ca/>

[CCME] Canadian Council of Ministers of the Environment. 2007. A protocol for the derivation of water quality guidelines for the protection of aquatic life. Canadian Council of Ministers of the Environment, Winnipeg, MN.

Chambers PA, Guy MA, Roberts ES, Charlton MN, Kent R, Gagnon C, Grove G and Foster N. 2001. Nutrients and their impact on the Canadian environment. Ottawa (ON): Agriculture and Agri-Food Canada, Environment Canada, Fisheries and Oceans Canada, Health Canada and Natural Resources Canada. 241 p.

Chambers PA, Culp JM, Glozier NE, Cash KJ, Wrona FJ, Norton L. 2006. Northern Rivers Ecosystem Initiative: Nutrients and Dissolved Oxygen—issues and impacts. *Environmental Monitoring and Assessment* 113(1-3): 117-142.

Chapman PM. 1992. Sediment Quality Triad Approach. US EPA Sediment Classification Methods Compendium. Office of Science and Technology, Washington, DC. 10-1-10-18.

Chapman PM and Hollert H. 2006. Should the Sediment Quality Triad become a tetrad, a pentad, or possibly even a hexad? *Journal of Soils and Sediments* 6(1): 4-8.

Chessman BC. 1985. Artificial-substratum periphyton and water quality in the lower La Trobe River, Victoria. *Aust. J. Mar. Freshwater. Res.* 36:855-871.

Clements WH. 1991. Community responses of stream organisms to heavy metals: a review of observational and experimental approaches, in: *Metal Ecotoxicology. Concepts and applications*. Edited by MC Newman, AW McIntosh, Lewis Publ. Chelsea, Michigan, pp. 363-391.

Clements WH. 2004. Small-scale experiments support causal relationships between metal contamination and macroinvertebrate community responses. *Ecological Applications* 14: 954-957.

Clements WH and Kiffney DM. 1994. Integrated laboratory and field approach for assessing impacts of heavy metals at the Arkansas River, Colorado. *Environ. Toxicol. Chem.* 13: 397-404.

Cohen AS. *Paleolimnology: The history and Evolution of Lake Systems*. 2003. Oxford University Press. New York, NY, 509 pp.

Costa PM, Diniz MS, Caeiro S, Lobo J, Martins M, Ferreira AM, Caetano M, Vale C, DelValls TA and Costa MH. 2009. Histological biomarkers in liver and gills of juvenile *Solea senegalensis* exposed to contaminated estuarine sediments: a weighted indices approach. *Aquatic Toxicology* 92 (3): 202-212.

Couillard Y. 1997. With the collaboration of L. St-Cyr. Technical evaluation of metallothionein as a biomarker for the mining industry. Technical report for the Aquatic Effects Technology Evaluation (AETE) program, AETE Project 2.2.1, prepared for Natural Resources Canada, Canada Centre for Mineral and Energy Technology (CANMET). 191 pp.

Couillard Y, Giguère A, Campbell PGC, Perceval O, Pinel-Alloul B and Hare L. 1999. Field evaluation of the use of metallothionein as a biomarker for metal contamination and toxic effects in the freshwater bivalve *Pyganodon grandis*: Subcellular metal partitioning. SETAC 20th Annual Meeting, 14-18 November 1999, Philadelphia, PA.

Couillard Y, Grapentine LC, Borgmann U, Doyle P and Masson S. 2008. The amphipod *Hyalella azteca* as a biomonitor in field deployment studies for metal mining. *Environmental Pollution* 156: 1314-1324.

Courtenay LA. 2002. Quantifying impacts of pulp mill effluent on fish in Canadian marine and estuarine environments: problems and progress. *Water Qual. Res. J. Can.* 37: 79-99.

Courtney LA and Clements WH. 2002. Assessing the influence of water and substratum quality on benthic macroinvertebrate communities in a metal-polluted stream: an experimental approach. *Freshwater Biology* 47(9): 1766-1779.

Crawford JK, Luoma SN. 1993. *Guidelines for studies of contaminants in biological tissues for the National Water– Quality Assessment Program*. U.S. Geological Service Open-File Rep. No. 92-494. Denver, CO.

Culp JM. 1999. Weight-of-Evidence Assessments for Large Rivers: Linking Community Bioassays with Field and Laboratory Studies. Workshop on Integrated Approaches for Interpreting Environmental Effects Monitoring Data. DOE EEM/1999/1.

Culp JM, Podemski CL, Cash KJ and Lowell RL. 1996. Utility of field-based artificial streams for assessing effluent effects on riverine ecosystems. *J. Aquat. Ecosystem Health* 5: 117-124.

Culp JM, Lowell RB and Cash KJ. 2000. Integrating *in situ* community experiments with field studies to generate weight-of-evidence risk assessments for large rivers. *Environ. Toxicol. Chem.* 19(4): 1167-1173.

Deaver E and Rodgers JH Jr. 1996. Measuring bioavailable copper using anodic stripping voltammetry. *Environ. Toxicol. Chem.* 15: 1925-1996.

Deutsch Forschungsgemeinschaft (German Research Council). 1998. Proposals for Safeguarding Good Scientific Practice. Wiley-VCH, Verlag GmbH, D-69469 Weinheim (Federal Republic of Germany). 85 pp.

Di Toro DM. 1989. A Review of the Data Supporting the Equilibrium Partitioning Approach to Establishing Sediment Quality Criteria, report to the National Research Council (as cited in Baudo 1990).

Di Toro DM, Allen HE, Bergman HL, Meyer JS, Paquin PR and Santore RC. 2000. The Biotic Ligand Model: A Computational Approach for Assessing the Ecological Effects of Metals in Aquatic Systems, published by the International Copper Association, Ltd., Environmental Program, as part of its series Copper in the Environment and Health.

Di Toro DM, Allen HE, Bergman HL, Meyer JS, Paquin PR and Santore RC. 2001. A Biotic Ligand Model of the Acute Toxicity of Metals. I. Technical Basis, Environmental Toxicology and Chemistry 20(10): 2383-2396.

Dodds WK. 2006. Eutrophication and trophic state in rivers and streams. Limnology and Oceanography 51 (1, part 2): 671-680.

Dodds WK and Welch EB. 2000. Establishing nutrient criteria in streams. J. N. Am. Benthol. Soc. 19(1): 186-196.

Doig L and Liber K. 2000. Dialysis minipeeper for measuring pore-water metal concentrations in laboratory sediment toxicity and bioavailability tests. Environmental Toxicology and Chemistry 19 (12): 2882-2889.

Ekberg C and Ödegaard-Jensen A. 2011. Uncertainties in chemical modelling of solubility, speciation and sorption. Accreditation and Quality Assurance 16 (4/5), Heidelberg: Springer Berlin, 2011, 207-214.

Environment Canada. 1990 (amended 1996). Biological Test Method: Acute Lethality Test Using *Daphnia* spp., Environmental Protection Series Report, EPS 1/RM/11, Ottawa, Ontario, 57 pp.

Environment Canada. 1992. Biological Test Method: Fertilization Assay Using Echinoids (Sea Urchins and Sand Dollars), Environmental Protection Series Report, EPS 1/RM/27, Ottawa, Ontario, 97 pp.

Environment Canada. 1994. Guidance Document on Collection and Preparation of Sediments for Physiochemical Characterization and Biological Testing, Environmental Protection Series Report EPS 1/RM/29. Ottawa.
<http://www.ec.gc.ca/Publications/default.asp?lang=En&xml=8F61B510-EFAA-4041-8FFF-BC7B99CC6B8B>

Environment Canada. 1996. Biological Test Method: Acute Lethality Test Using *Daphnia* spp., Environmental Protection Series Report, EPS 1/RM/11, July 1990 with May 1996 amendments. Ottawa, Ontario, 75 pp.

Environment Canada. 1997a. Test for Growth and Survival in Sediment Using the Freshwater Amphipod *Hyalella azteca*, Environmental Protection Series Report, EPS 1/RM/33, Ottawa, Ontario, 123 pp.

Environment Canada. 1997b. Test for Growth and Survival in Sediment Using Larvae of Freshwater Midges (*Chironomus tentans*, *Chironomus riparius*), Environmental Protection Series Report, EPS 1/RM/32, Ottawa, Ontario, 131 pp.

Environment Canada. 2007. Criteria and Guidance for Determination of Pronounced Eutrophication. National Environmental Effects Monitoring Office, Environment Canada.

Environment Canada. 2010. Pulp and Paper Environmental Effects Monitoring Technical Guidance Document—Chapter 11 Investigation of Cause and Investigation of Solutions. Environment Canada, 40 pp.

Environment Canada. 2012. Environmental Effects Monitoring Investigation of Cause Workshop for Metal Mining: Proceedings, Gatineau, Canada, ISBN: 978-1-100-19968-9.

Environment Canada. 2012. Second National Assessment of Environmental Effects Monitoring Data From Metal Mines subjected to the Metal Mining Effluent Regulations. Environment Canada, 66 pp.

ESG International Inc. (ESG). 1999. AETE Synthesis Report of Selected Technologies for Cost-Effective Environmental Monitoring of Mine Effluent Impacts in Canada, AETE Project 4.1.4, March 1999, CANMET, Natural Resources Canada, Ottawa, Ontario.

EVS Environmental Consultants. 1997. Technical evaluation: water quality and biological effects. Report for AETE program, CANMET, Natural Resources Canada, Ottawa, Ontario.

Fähræus-van Ree GE and Payne JF. 2005. Endocrine disruption in the pituitary of white sucker (*Catostomus commersoni*) caged in a lake contaminated with iron-ore mine tailings. *Hydrobiologia* 532: 221-224.

Farley KJ, Rader KJ and Miller BE. 2008. Tableau Input Coupled Kinetic Equilibrium Transport (TICKET) Model. *Environmental Science & Technology* 42(3): 838-847.

Fox GA. 1991. Practical causal inference for ecoepidemiologists. *J. Toxicol. Environ. Health* 33: 359-379.

Frey DG. 1998. What is paleolimnology? *J. Paleolimnology* 1: 5-8.

Gibbons WN and Munkittrick KR. 1994. A sentinel monitoring framework for identifying fish population responses to industrial discharges. *J. Aquat. Eco. Health* 3: 227-237.

Giesy JP and Hoke RA. 1990. Freshwater Sediment Quality Criteria: Toxicity Bioassessment, pp. 265-348, in: *Sediments: Chemistry and Toxicity of In-place Pollutants*, Baudo R, Giesy JP and Muntau H (eds.), Lewis Publishers, Inc. Chelsea, MI, 405 p.

Giguère A, Couillard Y, Perceval O, Campbell PGC, Pellerin-Massicotte J and Pinel-Alloul B. 1999. Field evaluation of the use of metallothionein as a biomarker for metal

contamination and toxic effects in the freshwater bivalve *Pyganodon grandis*: Responses at the organism level. SETAC 20th Annual Meeting, 14-18 November 1999, Philadelphia, PA.

Gilbertson M. 1997. Advances in forensic toxicology for establishing causality between Great Lakes epizootics and specific persistent toxic chemicals. Environ. Toxicol. Chem. 16: 1771-1778.

Gillis PL, Dixon DG, Reynoldson TB and Diener LC. 1999. Metallothionein as a biomarker for trace metal exposure and effects in *Tubifex tubifex* and *Chironomus riparius*. SETAC 20th Annual Meeting, 14-18 November 1999, Philadelphia, PA.

Grapentine LC. 2012. *In situ* measurements of toxicity and contaminant bioaccumulation with caged amphipods exposed to water and sediment for Investigation of Cause in metal mining Environmental Effects Monitoring studies in: Environment Canada. 2012. Environmental Effects Monitoring Investigation of Cause Workshop for Metal Mining: Proceedings, Gatineau, QC.

Gustafson JP. 2004. Visual MINTEQ, version 2.30; a Windows version of MINTEQA2, version 4.0. <http://www.lwr.kth.se/english/OurSoftware/Vminteq/>

Håkanson L and Jansson M. 1983. Principles of Lake Sedimentology, Springer-Verlag, Berlin, GD, 316 pp.

Hall GEM. 1998. Cost-effective protocols for the collection, filtration and preservation of surface waters for detection of metals and metalloids at ppb and ppt levels. Report for the AETE program, CANMET, Natural Resources Canada, Ottawa, Ontario.

Hargrave BT, Phillips GA, Doucette LI, White MJ, Milligan TG, Wildish DJ and Cranston RE. 1995. Biogeochemical observations to assess benthic impacts of organic enrichment from marine aquaculture in the Western Isles region of the Bay of Fundy, Can. Tech. Rep. Fish. Aquat. Sci. No. 2062, 159 pp.

Harkey GA, Landrum PF and Kaine SJ. 1994. Comparison of Whole-sediment, Elutriate and Pore-water Exposures for Use in Assessing Sediment-associated Organic Contaminants in Bioassays, Environ. Toxicol. Chem. 13(8): 1315-1329.

Haslam SM. 1982. A proposed method for monitoring river pollution using macrophytes. Environ. Technol. Letters 3: 19-34.

Hellawell JM. 1986. Biological indicators of freshwater pollution and environmental management. Elsevier Applied Science Publishers, London, 546 pp.

Henderson-Sellers B and French RH. 1991. Water Quality Modelling Volume IV: Decision Support Techniques for Lakes and Reservoirs. CRC Press, Boca Raton,

Florida.

Henry CJ, Higgins KF and Buhl KJ. 1994. Acute toxicity and hazard assessment of Rodeo®, X-77 Spreader®, and Chem-Trol® to aquatic invertebrates. Arch. Environ. Contam. Toxicol. 27: 392-399.

Herbelin AL and Westall JC. 1999. FITEQL, A Computer Program for Determination of Chemical Equilibrium Constants from Experimental Data. Report 99-01, Version 4.0. Oregon State University, Corvallis, OR.

Hewitt ML and Servos MR. 2001. An Overview of Substances Present in Canadian Aquatic Environments Associated with Endocrine Disruption. Water Qual. Res. J. Canada 36(2): 191–213.

Hewitt ML, Dube MG, Culp JM, MacLatchy DL and Munkittrick KR. 2003. A proposed framework for investigation of cause for environmental effects monitoring. Human and Ecological Risk Assessment 9(1): 195-211.

Hinton DE, David E, Darrel JL and Adams SM. 1990. Integrative histopathological approaches to detecting effects of environmental stressors on fishes. American Fisheries Society. Symposium: Biological indicators of stress in fish No. 8, pp. 51-66. 1990. FR 36(1).

Hinton DE, Kullman SW, Bencic DC, Hardman RC, Chen PJ, Carney M and Volz DC. 2005. Resolving mechanisms of toxicity while pursuing ecotoxicological relevance. Marine Pollution Bull. 51: 635-648.

Hodson PV. 1990. Indicators of ecosystem health at the species level and the example of selenium affects on fish. Environmental Monitoring and Assessment 15(3): 241-255.

Holm J, Palace VP, Siwik P, Sterling G, Evans RE, Baron CL, Werner J and Wautier K. 2005. Developmental effects of bioaccumulated selenium in eggs and larvae of two salmonid species. Environ. Toxicol. Chem. 24: 2373-2381.

Hontela A, Ramussen JB, Audet C and Chevalier G. 1992. Impaired cortisol stress response in fish from environments polluted by PAHs, PCB, and mercury, Arch. Environ. Contam. Toxicol. 22: 278–283.

Ingersoll CG, Ivey CD, Brunson EL, Hardesty DK and Kemble NE. 2000. Evaluation of toxicity: Whole-sediment versus overlying-water exposures with amphipod *H. azteca*. Environmental Toxicology and Chemistry 19: 2906-2910.

International Agency for Research on Cancer. 2008. Code of Good Scientific Practice. World Health Organization, Geneva, Switzerland, 11 pp.

Kelly M. 1988. Mining and the freshwater environment. Elsevier Applied Science, London, UK, 231 pp.

Knezovich JP and Harrison FL. 1987. The Bioavailability of Sediment-sorbed Organic Chemicals: A Review, Water Air Soil Pollut. 32: 233-245.

Knezovich JP and Harrison FL. 1988. The Bioavailability of Sediment-sorbed Chlorobenzenes to Larvae of the Midge, *Chironomus decorus*, Ecotoxicol. Environ. Safety 15: 226-241.

Kutka FJ and Richards C. 1996. Relating diatom assemblage structure to stream habitat quality. J.N. Am. Benthol. Soc. 15: 469-480.

Laflamme JS, Couillard Y, Campbell PGC and Hontela A. 1998. Impaired physiological response to stress and high levels of metallothionein in perch (*Perca flavescens*) from lakes contaminated by heavy metals. 25th Annual Aquatic Toxicity Workshop, 18- 21 October 1998, Quebec City, QC.

Langmuir D, Chrostowski P, Vigneault B and Chaney R. Issue Paper on the Environmental Chemistry of Metals. Submitted to: U.S. Environmental Protection Agency Risk Assessment Forum. Submitted by: ERG 110 Hartwell Avenue Lexington, MA.

Ledo H, Rivas Z, Gutierrez J, Gutierrez E, Ojeda J and Avila H. 2004. Baseline of Ca, Mg, Fe, Mn, and Al concentrations in Catatumbo River Surficial sediments. Water, Air, and Soil Pollution 155: 117–135.

Liber D and Liber K. 2000. Dialysis minipeeper for measuring pore-water metal concentrations in laboratory sediment toxicity and bioavailability tests. Environ. Toxicol. Chem. 19 (1): 2882-2889.

Liber K, Doig LE and White-Sobey SL. 2011. Toxicity of uranium, molybdenum, nickel, and arsenic to *Hyaella azteca* and *Chironomus dilutus* in water-only and spiked-sediment toxicity tests. Ecotox. Environ. Saf. 74: 1171-1179.

Lowe RL and Pank Y. 1996. Benthic algal communities as biological monitors, in: Algal ecology. Freshwater benthic ecosystem. Edited by Stevenson RJ, Bothwell ML and Lowe RL. Academic Press, San Diego, CA.

Lowell RB, Culp JM and Dube MG. 2000. A weight-of evidence approach for northern river risk assessment: integrating the effects of multiple stressors. Environ. Toxicol. Chem. 19(4): 1182-1190.

Lowell RB, Tessier C, Walker SL, Willsie A, Bowerman M and Gautron D. 2007. National Assessment of Phase 1 Data from the Metal Mining Environmental Effects

Monitoring Program. National Environmental Effects Monitoring Office, Environment Canada, Gatineau, QC, 45 pp.

Luoma SN and Rainbow PS. 2008. Metal contamination in aquatic environments: science and lateral management. Cambridge University Press, Cambridge, UK.

Mabit L, Benmansour M and Walling DE. 2008. Comparative advantages and limitations of the fallout radionuclides ^{137}Cs , ^{210}Pb and ^7Be for assessing soil erosion and sedimentation. *Journal of Environmental Radioactivity* 99 (12): 1799-1807.

Martel PH, Kovacs TG, O'Connor BL and Voss RH. 1997. Source and identity of compounds in a thermomechanical pulp mill effluent inducing hepatic mixed-function oxidase activity in fish. *Environ Toxicol. Chem.* 16: 2375-83.

McCutcheon SC and French RH. 1989. *Water Quality Modelling Volume I: Transport and Surface Exchange in Rivers*. CRC Press, Boca Raton, Florida.

McGeer JC, Playle RC, Wood CM and Galvez F. 2000. A Physiologically Based Biotic Ligand Model for Predicting the Acute Toxicity of Waterborne Silver to Rainbow Trout in Freshwaters. *Environmental Science and Technology* 34: 4199-4207.

McGeer J, Henningsen G, Lanno R, Fisher N, Sappington K and Drexler J. 2004. Issue paper on the bioavailability and bioaccumulation of metals. Available online at <http://cfpub1.epa.gov/ncea/raf/recordisplay.cfm?deid=86119>.

McGeer JC, Clifford M, Janssen CR and De Schamphelaere KAC. 2010. Modeling the toxicity of metals to aquatic biota using the biotic ligand approach, pp. 205-231 in: *Essential Reviews in Experimental Biology Vol. 2. Surface Chemistry, Bioavailability and Metal Homeostasis in Aquatic Organisms: an Integrated Approach*. Edited by Bury NR and Handy RD. SEB Press, London, UK.

McGeer J, Clifford M, Ng T and Wood C. 2012. Using bioaccumulation models for predicting dissolved metal toxicity, in: Environment Canada. 2012. *Environmental Effects Monitoring Investigation of Cause Workshop for Metal Mining: Proceedings*, Gatineau, QC.

Meador J, Parkerton T and McElroy A. 2011. Tissue residue approach to toxicity assessment: A step forward. Society of Environmental Toxicology and Chemistry.

Meyer JS, Santore RC, Bobbitt JP, DeBrey LD, Boese CJ, Paquin PR, Allen HE, Bergman HL and Di Toro DM. 1999. Binding of nickel and copper to fish gills predicts toxicity but free ion activity does not. *Environ. Sci. Tech.* 33: 913-916.

[MOEE] Ontario Ministry of Environment and Energy. 1997. Sediment and biological assessment of Canagigue Creek at the Uniroyal Chemical Ltd. Plant, Elmira, Ontario. 1995-1996. Ministry of Environment and Energy, 37 pp. + appendices.

Muir DCG, Kenny DF, Grift N, Robinson RD, Titman RD and Murkin HR. 1991. Fate and acute toxicity of bromoxynil esters in an experimental prairie wetland. *Environ. Toxicol. Chem.* 10: 395-406.

Mulliss RM, Revitt DM and Shutes RBE. 1996. A statistical approach for the assessment of the toxic influences on *Gammarus pulex* (Amphipoda) and *Asellus aquaticus* (Isopoda) exposed to urban discharges. *Wat. Res.* 30: 1237-1243.

Munkittrick KR, Portt CB, Van Der Kraak GJ, Smith IR and Rokosh DA. 1991. Impact of bleached kraft mill effluent on population characteristics, liver MFO activity, and serum steroid levels of a Lake Superior white sucker (*Catostomus commersoni*) population. *Can. J. Fish Aquat. Sci.* 48: 1371-1380.

Munkittrick KR, Van Der Kraak GJ, McMaster ME, Portt CB, van den Heuvel MR and Servos MR. 1994. Survey of receiving-water environmental impacts associated with discharges from pulp mills. 2. Gonad size, liver size, hepatic EROD activity and plasma sex steroid levels in white sucker. *Environ. Toxicol. Chem.* 13: 1089-1101.

Munkittrick KR, McMaster ME, Van Der Kraak G, Portt C, Gibbons WN, Farwell A and Gray M. 2000. Development of Methods for Effects-Driven Cumulative Effects Assessment Using Fish Populations: Moose River Project. The Society of Environmental Toxicology and Chemistry, Pensacola, FL, 236 pp.

Nebeker AV, Cairns MA, Gakstatter JH, Malueg KW, Schuytema GS and Krawczyk DR. 1984. Biological Methods for Determining Toxicity of Contaminated Freshwater Sediments to Invertebrates, *Environ. Toxicol. Chem.* 3: 617-630.

Niyogi S and Wood CM. 2004. The Biotic Ligand Model, a flexible tool for developing site-specific water quality guidelines for metals. *Environ. Sci. Technol.* 6177-6192.

North America Metals Council. 2008. Selenium Tissue Thresholds: Tissue Selection Criteria, Threshold Development Endpoints, and Potential to Predict Population or Community Effects in the Field. 178 pp. <http://www.namec.org/docs/00043675.PDF>

Norwood WP, Borgmann U and Dixon DG. 2007. Chronic toxicity of arsenic, cobalt, chromium and manganese to *Hyalella azteca* in relation to exposure and bioaccumulation. *Environmental Pollution* 147: 262-272.

Novak LJ and Holtze LE. 2012. Toxicity Reduction Evaluations (TREs) as a Tool for Investigation of Cause—Mining Case Studies, in: Environment Canada. Environmental

Effects Monitoring Investigation of Cause Workshop for Metal Mining: Proceedings, Gatineau, QC.

Nowierski M, Dixon DG and Borgmann U. 2005. Effects of water chemistry on the bioavailability of metals in sediment to *Hyalella azteca*: Implications for sediment quality guidelines. Arch. Environ. Contam. Toxicol. 49: 322–332.

[OECD] Organisation for Economic Co-operation and Development, Global Science Forum. 2007. Best Practices for Ensuring Scientific Integrity and Preventing Misconduct. Organisation for Economic Co-operation and Development, Paris, France, 13 pp.

Palace VP, Doebel C, Baron CL, Evans RE, Wauiter KG, Klaverkamp JF, Werner J and Kollar S. 2005. Caging small bodied fish as an alternative method for environmental effects monitoring (EEM). Can. Wat. Qual. Res. J. 40: 328-333.

Palace VP. 2012. Investigating Selenium Toxicity Using an Environmental Effects Monitoring Approach, in: Environment Canada. 2012. Environmental Effects Monitoring Investigation of Cause Workshop for Metal Mining: Proceedings, Gatineau, QC.

Paquin PR, Di Toro DM, Santore RC, Trivedi D and Wu KB. 1999. A Biotic Ligand Model of the Acute Toxicity of Metals. III. Application to Fish and *Daphnia* Exposure to Silver, Section 3 in Integrated Approach to Assessing the Bioavailability and Toxicity of Metals in Surface Waters and Sediments, a submission to the EPA Science Advisory Board, Office of Water, Office of Research and Development, Washington, DC, pp. 3-59 to 3-102. US EPA-822-E-99-001.

Paquin PR, Gorsuch JW, Apte S, Batley GE, Bowles KC, Campbell PGC, Delos CG, Di Toro DM, Dwyer RL, Galvez F, Gensemer RW, Goss GG, Hogstrand C, Janssen CR, McGeer JC, Naddy RB, Playle RC, Santore RC, Schneider U, Stubblefield WA, Wood CM, Wu KB. 2002a. The Biotic Ligand Model: A Historical Overview, Special Issue: The Biotic Ligand Model for Metals—Current Research, Future Directions, Regulatory Implications, Comparative Biochemistry and Physiology, Part C 133(1-2): 3-35.

Paquin PR, Zoltay V, Winfield RP, Wu KB, Mathew R, Santore R, Di Toro DM. 2002b. Extension of the Biotic Ligand Model of Acute Toxicity to a Physiologically-based Model of the Survival Time of Rainbow Trout (*Oncorhynchus mykiss*) Exposed to Silver, Special Issue: The Biotic Ligand Model for Metals—Current Research, Future Directions, Regulatory Implications, Comparative Biochemistry and Physiology, Part C 133(1-2): 305-343.

Parker R and Dumaresq C. 2002. Effluent characterization, water quality monitoring, and sediment monitoring in the metal mining EEM program. Water Qual. Res. J. Canada 37(1): 219-228.

Parkhurst DL and Appelo CA. 2000: User's guide to PHREEQC (Version 2)—A computer program for speciation, batch-reaction, one-dimensional transport, and inverse geochemical calculations. U.S. Geological Survey, Water-Resources Investigations Report 99-4259, 112 pp.

Parrott JL. 2005. Overview of methodology and endpoints in fathead minnow lifecycle tests assessing pulp mill effluents. *Water Qual. Res. J. Can.* 40(3): 334-346.

Parrott JL, Bennie DT. 2009. Lifecycle exposure of Fathead minnows to a mixture of six common pharmaceuticals and triclosan. *J. Toxicol. Environ. Health, Part B*, 9: 297-317.

Parrott JP, Hewitt LM and McMaster ME. 2012. Fathead Minnow Lifecycle Assays for Assessment of Complex Effluents, in: Environment Canada. 2012. Environmental Effects Monitoring Investigation of Cause Workshop for Metal Mining: Proceedings, Gatineau, QC.

Patrick WH Jr, Gambrell RP and Khalid RA. 1977. Physicochemical Factors Regulating Solubility and Bioavailability of Toxic Heavy Metals in Contaminated Dredged Sediment, *J. Environ. Sci. Health A12*: 475-492.

Pearson TH, Rosenberg R. 1978. Macrobenthic succession in relation to organic enrichment and pollution of the marine environment. *Oceanography and Marine Biology Annual Review* 163: 229-311.

Peplow D, Edmonds R. 2005. The effects of mine waste contamination at multiple levels of biological organization. *Ecological Engineering* 24 (1/2): 101-119.

Perceval O, Couillard Y, Giguere A, Pinel-Alloul B and Campbell PGC. 1999. Field evaluation of the use of metallothionein as a biomarker for metal contamination and toxic effects in the freshwater bivalve *Pyganodon grandis*: Responses at the population level. SETAC 20th Annual Meeting, 14-18 November 1999, Philadelphia, PA.

Pereira P, de Pablo H, Vale C, Rosa-Santos F and Cesário R. 2009. Metal and nutrient dynamics in a eutrophic coastal lagoon (Óbidos, Portugal): the importance of observations at different time scales. *Environ. Monit. Assess.* 158: 405-418.

Phillips DJH and Rainbow PS. 1993. Biomonitoring of trace aquatic contaminants. Elsevier Applied Science Publishers, London, UK.

Pierron F, Bourret V, St-Cyr J, Campbell PGC, Bernatchez L and Couture P. 2009. Transcriptional responses to environmental metal exposure in wild yellow perch (*Perca flavescens*) collected in lakes with differing environmental metal concentrations (Cd, Cu, Ni). *Ecotoxicol.* 18: 620-631.

Pinel-Alloul B, Campbell PGC, Hare L, Couillard Y, Giguere A and Perceval O. 1999. Field evaluation of the use of metallothionein as a biomarker for metal contamination and toxic effects in the freshwater bivalve *Pyganodon grandis*: Overview. SETAC 20th Annual Meeting, 14-18 November 1999, Philadelphia, PA.

Pollard AI and Yuan L. 2006. Community response patterns: evaluating benthic invertebrate composition in metal-polluted streams. *Ecological Applications* 16 (2): 645-655.

Rickwood CJ, Dubé MG, Hewitt M, Kovacs TG and MacLachy DL. 2006. Use of paired fathead minnow (*Pimephales promelas*) reproductive test. Part 2: Source identification of biological effects at a bleached kraft pulp mill. *Env. Toxicol. Chem.* 25: 1847-1856.

Ridal JJ, Yanch LE, Fowlie AR, Razavi NR, Delongchamp TM, Choy ES, Fathi M, Hodson PV, Campbell L, Blais JM, Hickey MBC, Yumvihoze E and Lean DRS. 2010. Potential causes of enhanced transfer of mercury to St. Lawrence River Biota: implications for sediment management strategies at Cornwall, Ontario, Canada. *Hydrobiologia*, 2010: 81-98.

Ridley-Thomas CI, Austin A, Lucey WP and Clark MJR. 1989. Variability in the determination of ash free dry weight for periphyton communities: a call for a standard method. *Water Research* 23(6): 667-670.

Ritchie JC and McHenry JR. 1990. Application of radioactive fallout cesium-137 for measuring soil erosion and sediment accumulation rates and patterns: a review. *Journal of Environmental Quality* 19 (2): 215-233.

Roch M, Nordin RN, Austin A, McKean CJP, Deniseger J, Kathman RD, McCarter JA and Clark MJR. 1985. The effects of heavy metal contamination on the aquatic biota of Buttle Lake and the Campbell River drainage (Canada). *Arch. Environ. Contam. Toxicol.* 14: 347-362.

Sánchez España J, López Pamo E, Santofimia Pastor E and Diez Ercilla M. 2008. The acidic mine pit lakes of the Iberian Pyrite Belt: an approach to their physical limnology and hydrogeochemistry. *Applied Geochemistry* 23 (5): 1260-1287.

Santore RC, McGrath J, Brix K, Paquin PR and Di Toro DM. 1998. Use of a biotic ligand model to calculate site-specific water effect ratios for metals. SETAC 1998 Annual Conference, Charlotte, NC.

Santore RC, Di Toro DM, Paquin PR, Allen HE and Meyer JS. 2001. A Biotic Ligand Model of the Acute Toxicity of Metals. II. Application to Acute Copper Toxicity in Freshwater Fish and *Daphnia*, *Environmental Toxicology and Chemistry* 20(10): 2397-2402.

Santore RC, Mathew R, Paquin PR and Di Toro DM. 2002. Development of a Biotic Ligand Model of Acute Toxicity for Zinc, Special Issue: The Biotic Ligand Model for Metals—Current Research, Future Directions, Regulatory Implications, Comparative Biochemistry and Physiology, Part C 133(1-2): 271-285.

Sappington KG, Bridges TS, Bradbury SP, Erickson RJ, Hendriks AJ, Lanno RP, Meador JP, Mount DR, Salazar MH and Spry DJ. 2011. Application of the tissue residue approach in ecological risk assessment, Integrated Environmental Assessment and Management Special Issue: Tissue Residue Approach Special Series 7(1): 116–140.

Schecher WD and McAvoy DC. 1992. MINEQL+: a software environment for chemical equilibrium modelling. Comput. Environ. Urban Systems 16: 65-76.

Schecher WD and McAvoy DC. 1994. MINEQL+: A Chemical Equilibrium Program for Personal Computers (Version 3.01), Hallowell, Maine: Environmental Research Software, 107 pp.

Schuytema GS, Nelson PO, Malueg KW, Nebeker AV, Krawczyk DF, Ratcliff AK and Gakstatter JH. 1984. Toxicity of Cadmium in Water and Sediment to *Daphnia magna*. Environ. Toxicol. Chem. 3: 293-308.

Shaner SW and Knight AW. 1985. The Role of Alkalinity in the Mortality of *Daphnia magna* in Bioassays of Sediment-bound Copper, Comp. Biochem. Physiol. 82C: 273-277.

Sharpe RL, Machtans H, Crowe J, Smith P, Patrick H, Chapman PM, Connell R and Daniels E. 2012. Con Mine: Investigation of Cause Study on fish livers – Challenges to designing a new Investigation of Cause study, in: Environment Canada. 2012. Environmental Effects Monitoring Investigation of Cause Workshop for Metal Mining: Proceedings, Gatineau, QC.

Shifrin NS, Beck BD, Gauthier TD, Chapnick SD and Goodman G. (1996) Chemistry, toxicology and human health risk of cyanide compounds in soils at former manufactured gas plant sites. Regul. Toxicol. Pharmacol. 23: 106-116.

Shobanov NA, Kiknadze II and Butler MG. 1999. Palearctic and Nearctic *Chironomus (Camptochironomus) tentans Fabricius* are different species (Diptera, Chironomidae). Ent. Scand. 30: 311-322.

Shuhaimi-Othman M, Pascoe D, Borgmann U and Norwood WP. 2006. Reduced metals concentrations of water, sediment and *Hyalella azteca* from lakes in the vicinity of the Sudbury metal smelters, Ontario, Canada. Environmental Monitoring and Assessment 117 (1/3): 27-44.

Small AM, Adey WH, Lutz SM, Reese EG and Roberts DL. 1996. A macrophyte-based rapid biosurvey of stream water quality: restoration at the watershed scale. *Restoration Ecol.* 4: 124-145.

[SETAC] Society of Environmental Toxicology and Chemistry. SETAC Technical Issue Paper. Sound Science. SETAC, Pensacola, FL, 1999.

Sortkjaer O. 1984. Macrophytes and macrophyte communities as test systems in ecotoxicological studies of aquatic systems. *Ecol. Bull.* 36: 75-80.

St. Cyr L and Campbell PGC. 2000. Bioavailability of sediment-bound metals for *Vallisneria americana Michx*, a submerged aquatic plant, in the St. Lawrence River. *Canadian Journal of Fisheries and Aquatic Sciences* 57: 1330-1341.

Stumm W and Morgan JJ. 1996. *Aquatic chemistry: chemical equilibria and rates in natural waters* (3rd ed.) Wiley Interscience, New York, NY.

Suter GW II. 1993. *Ecological Risk Assessment*. Lewis Publishers, Boca Raton, FL.

Szarek-Gwiazda E and Zurik R. 2006. Distribution of Trace Elements in Meromictic Pit Lake. *Water, Air & Soil Pollution* 174(1-4): 181-196.

Taylor LN, McFarlane WJ, Pyle GG, Couture P and McDonald DG. 2004. Use of performance indicators in evaluating chronic metal exposure in wild yellow perch (*Perca flavescens*). *Aquatic Toxicology* 67: 371-385.

Taylor LN, Novak K, Holtze K, Ali N and Scroggins R. 2012. Separating current effluent quality from historical contamination using a laboratory based monitoring tool for investigation of cause, in: Environment Canada. 2012. Environmental Effects Monitoring Investigation of Cause Workshop for Metal Mining: Proceedings, Gatineau, QC.

Tessier A, Couillard Y, Campbell PGC and Auclair JCA. 1993. Modeling of Cd partitioning in oxic lake sediments and Cd concentrations in the freshwater bivalve *Anodonta grandis*. *Limnol. Oceanogr.* 38: 1-17.

Tessier A and Turner DR (ed.) 1995. *Metal speciation and bioavailability in aquatic systems*. John Wiley and Sons Ltd., Chichester, UK.

Thomann RT and Mueller JA. 1987. *Principles of Surface Water Quality Modelling and Control*. Harper Collins Publishers Inc., New York, NY.

Thomas GP. 2012. Use and application of Benthic Transplant Devices (BTDs) in EEM investigations, in: Environment Canada. 2012. Environmental Effects Monitoring Investigation of Cause Workshop for Metal Mining: Proceedings, Gatineau, QC.

Tipping E. 1998. Humic ion-binding Model VI: an improved description of the interactions of protons and metal ions with humic substances. *Aq. Geochem.* 4: 3-48.

Unsworth ER, Warnken KW, Zhang H, Davison W, Black F, Buffle J, Cao J, Cleven R, Galceran J, Gunkel P, Kalis E, Kistler D, Herman P, van Leeuwen HP, Martin M, Noël S, Nur Y, Odzak N, Puy J, van Riemsdijk W, Sigg L, Temminghoff E, Tercier-Waeber ML, Toepperwien S, Town RM, Weng L and Xue H). 2006. Model predictions of metal speciation in freshwaters compared to measurements by *in situ* techniques. *Environmental Science & Technology* 40(6): 1942-1949.

[US EPA] United States Environmental Protection Agency. 1989. Generalized methodology for conducting industrial toxicity reduction evaluations. EPA-600/2-88/070.

[US EPA] United States Environmental Protection Agency. 1991a. Methods for aquatic toxicity identification evaluations: Phase I toxicity characterization procedures. EPA-600/6-91/003.

[US EPA] United States Environmental Protection Agency. 1991b. Toxicity identification evaluation: characterization of chronically toxic effluents, Phase I. EPA-600/6-91/005.

[US EPA] United States Environmental Protection Agency. 1991c. Sediment toxicity identification evaluation: Phase I (Characterization), Phase II (Identification) and Phase III (Confirmation) modifications of effluent procedures. EPA-600/6-91/007.

[US EPA] United States Environmental Protection Agency. 1993a. Methods for aquatic toxicity identification evaluations: Phase II toxicity identification procedures for samples exhibiting acute and chronic toxicity. EPA-600/R-92/080.

[US EPA] United States Environmental Protection Agency. 1993b. Methods for aquatic toxicity identification evaluations: Phase III toxicity confirmation procedures for samples exhibiting acute and chronic toxicity. EPA-600/R-92/081.

[US EPA] United States Environmental Protection Agency. 1996. Marine toxicity identification evaluation (TIE): Phase I guidance document. Duluth, MN: Environmental Research Laboratory. EPA/600/R-96/054.

[US EPA] United States Environmental Protection Agency. 1997. Sediment Toxicity Evaluation (TIE) Phases I, II, and III, Guidance Document. EPA/600/R-07/080.

[US EPA] United States Environmental Protection Agency. 2001. Methods for Collection, Storage and Manipulation of Sediments for Chemical and Toxicological Analyses: Technical Manual. EPA-823-B-01-002.

[US EPA] United States Environmental Protection Agency. 2003. Biotic Ligand Model: Technical Support Document for its Application to the Evaluation of Water Quality Criteria for Copper Office of Science and Technology. Health and Ecological Criteria Division, Washington, DC. EPA Number 822R03027.

[US EPA] United States Environmental Protection Agency. 2004. Draft Aquatic Life Water Quality Criteria for Selenium—2004. EPA-822-D-04-001. Office of Water, Washington, DC.

[US EPA] United States Environmental Protection Agency. 2007. Framework for Metals Risk Assessment. Office of the Science Advisor, Risk Assessment Forum. Washington, DC: EPA/120/R-07/001.

van de Guchte C, Maas-Diepeveen JL. 1988. Screening Sediments for Toxicity: A Water-concentration Related Problem, in: Proceedings of the 14th Annual Aquatic Toxicity Workshop, November 1-4, 1987, Toronto, ON, Can. Tech. Rep. Fish. Aquat. Sci. 1607: 81-91.

Van Geest JL, Poirier DG, Solomon KR and Sibley PK. 2011. A comparison of the bioaccumulation potential of three freshwater organisms exposed to sediment-associated contaminants under laboratory conditions. Environ. Toxicol. Chem. 30: 939–949.

Van Geest JL, Poirier DG, Sibley PK and Solomon KR. 2011. Validation of Ontario's new laboratory-based bioaccumulation methods with *in situ* field data. Environ. Toxicol. Chem. 30: 950–958.

van Wijngaarden RPA, van den Brink PJ, Crum SJH, Voshaar JHO, Brock TCM, Leeuwangh P. 1996. Effects of the insecticide Dursban® 4E (active ingredient chlorpyrifos) in outdoor experimental ditches: I. Comparison of short-term toxicity between the laboratory and the field. Environ. Toxicol. Chem. 15: 1133-1142.

Vigneault B, Desforges M, McGeer, J. 2012. Mining Reagents and By-Products (e.g. Thiosalts) As Potential Toxicants in Mine Effluents, in: Environment Canada. 2012. Environmental Effects Monitoring Investigation of Cause Workshop for Metal Mining: Proceedings, Gatineau, QC.

Vijver MG, Van Gestel CAM, Lanno RP, Van Straalen NM and Peijnenburg WJGM. 2004. Internal metal sequestration and its ecotoxicological relevance: a review. Environ. Sci. Technol. 38: 4705-4712.

Vink JPM and Meeussen JCL. 2007. BIOCHEM-ORCHESTRA: a tool for evaluating chemical speciation and ecotoxicological impacts of heavy metals on river flood plain systems. Environmental Pollution 148 (3): 833-841.

Walker PA, Kille P, Hurley A, Bury NR and Hogstrand C. 2008. An in vitro method to assess toxicity of waterborne metals to fish. *Toxicol. Appl. Pharmacol.* 230: 67-77.

Wang D, Couillard Y, Campbell PGC and Jolicoeur P. 1999. Changes in subcellular metal partitioning in the gills of freshwater bivalves (*Pyganodon grandis*) living along an environmental cadmium gradient. *Can. J. Fish. Aquat. Sci.* 56: 774-784.

Wang F, Goulet RR and Chapman PM. 2004. Testing sediment biological effects with the freshwater amphipod *Hyaella azteca*: the gap between laboratory and nature. *Chemosphere* 57: 1713-1724.

Warren LA, Tessier A and Hare L. 1998. Modelling cadmium accumulation by benthic invertebrates *in situ*: The relative contributions of sediment and overlying water reservoirs to organism cadmium concentrations. *Limnol. Oceanogr.* 43: 1442-1454.

Warwick RM and Clarke KR. 1991. A Comparison of Some Methods for Analysing Changes in Benthic Community. *J. Mar. Biol. Ass. U.K.* 71: 225-244.

Weber CI. 1973. Biological field and laboratory methods for measuring the quality of surface waters and effluents. U.S. Environmental Protection Agency Report 670/4/73/001.

Weber LP, Higgins PS, Carlson RI and Janz DM. 2003. Development and validation of methods for measuring multiple biochemical indices of condition in juvenile fishes. *Journal of Fish Biology* 63: 637-658.

Weech S, Orr P, Russel C, Fedat L and Yaschyshyn D. 2012. Investigation of Cause of Effects on Benthic Invertebrates at the Kidd Metallurgical Site, in: Environment Canada. 2012. Environmental Effects Monitoring Investigation of Cause Workshop for Metal Mining: Proceedings, Gatineau, QC.

Weitzel RL, Sanocki SL and Holecek H. 1979. Sample replication of periphyton collected from artificial substrates, in: *Methods and Measurements of Periphyton Communities: a Review*. Edited by RL Weitzel. ASTM STP 690, Philadelphia, PA, pp. 90-115.

Werner I, Clar SL and Hinton DE. 2003. Biomarkers aid understanding of aquatic organism responses to environmental stressors. *California Agriculture* 57 (4): 110-115.

Whitton BA. 1984. Algae as monitors of heavy metals in freshwaters, in: *Algae as Ecological Indicators*. Edited by LE Shubert, Academic Press, London, UK, pp. 257-280.

Whitton BA, Say PJ and Wher JD. 1981. Use of plants to monitor heavy metals in rivers, in: *Heavy metals in Northern England: environmental and biological aspects*. Edited by

PJ Say and BA Whitton. Department of Botany, University of Durham, Durham, UK, pp. 135-145.

Wolf JC and Wolfe MJ. 2005. A Brief Overview of nonneoplastic hepatic toxicity in fish. *Toxicologic Pathology* 33: 75–85.

Yan ND. 1979. Phytoplankton community of an acidified, heavy metal contaminated lake near Sudbury, Ontario: 1973-1977. *Water, Air, and Soil Poll.* 11:43-45.

Zagury GJ. 2012. Cyanide speciation and fate in gold mine tailings: A case study, in: Environment Canada. 2012. Environmental Effects Monitoring Investigation of Cause Workshop for Metal Mining: Proceedings, Gatineau, QC.

Zapata F. 2003. The use of environmental radionuclides as tracers in soil erosion and sedimentation investigations: recent advances and future developments. *Soil & Tillage Research* 69 (1/2):3-13.

Table of Contents

13.	Report on Historical Information.....	1
13.1	Overview.....	1
13.2	Recommendations for the Facility’s Review of Historical Information.....	1
13.3	Recommendations for Historical Biological Monitoring Report	1
13.4	References.....	2

13. Report on Historical Information

13.1 Overview

This chapter is intended to provide guidance to mines that have collected “historical” information, i.e., information on biological monitoring studies conducted prior to becoming subject to the *Metal Mining Effluent Regulations* (MMER). See subsection 14(b) in Schedule 5 of the MMER for the provisions related to historical information and Chapter 1 for additional information (sections 1.3.2.2, 1.4.2 and 1.4.5).

13.2 Recommendations for the Facility’s Review of Historical Information

In order to determine the relevance of historical biological monitoring studies, a review of historical information or data collected for another regulatory agency should be undertaken by the facility. The AQUAMIN Working Group (1994) provided guidance to evaluate the relevance of historical studies and one of their main points was that the methods used in the historical monitoring studies should be validated protocols in the peer-reviewed scientific literature or at least be scientifically defensible. Chapter 12 provides updated guidance on good scientific practices and sound science. In addition, in order to determine the relevance of the historical study to the metal mining environmental effects monitoring program, the mine should refer to the appropriate chapters within this guidance document (e.g., study design, fish population survey, fish tissue, benthic invertebrate community survey, data analysis).

13.3 Recommendations for the Historical Biological Monitoring Report

Along with details on how the historical monitoring results were used to determine if the effluent was causing an effect on fish populations, fish tissue or the benthic invertebrate community, the report should consider the following information:

- study design information, site characterization, description of sampling areas and stations;
- potentially confounding or influencing factors to be considered;
- biological community descriptors;
- field and laboratory methods used;
- supporting environmental variables measured;
- data assessment and interpretation;
- quality assurance and quality control procedures;
- changes in mine operating conditions subsequent to the historical studies (e.g., loadings of deleterious substances), location of structures potentially affecting the aquatic environment (e.g., locations of dams, bridges, discharge points) or any

- other event that has had the potential to modify the effects of the effluent on fish, fish tissue or benthic invertebrates. Examples of such events are: change in ore processed, change in ore processing protocols or treatment of effluent, change in hydrological factors, presence of new confounding factors; and,
- a summary of biological monitoring data from the ecoregion found in the primary literature, government reports, and reports from other mines or industries as additional supporting information.

For additional information on the points listed above, refer to the appropriate chapters within this guidance document.

13.4 References

AQUAMIN – Assessment of the Aquatic Effects of Mining in Canada. 1994. *Criteria to Evaluate Information for AQUAMIN*. Working Document. Prepared for the AQUAMIN Steering Group by AQUAMIN Working Group 1.

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