

**Metal Mining
Environmental Effects Monitoring
Review Team Report**

August, 2007

Foreword

In December 2005, Environment Canada initiated the review of the Metal Mining Environmental Effects Monitoring (EEM) Program. The EEM program for metal mines under the 2002 *Metal Mining Effluent Regulations* (MMER) is the second of such federally regulated monitoring programs, the first of which was an EEM program for pulp and paper mills introduced in 1992. For both of these EEM programs, there has been a commitment to continuous improvement based on experiences and lessons learned through a process of engaging stakeholders and scientists, to ensure the EEM program is effective, efficient and based on the most current scientific knowledge. During the development of the Metal Mining EEM Program (1997-2000), Environment Canada committed to reviewing the overall effectiveness of the program approximately three years after its commencement, upon completion of the first phase of biological monitoring. To meet this commitment, Environment Canada established the Metal Mining EEM Review Team which consisted of a group of experts from government, industry, and environmental and aboriginal groups to undertake this program review based on experiences learned from the first phase of monitoring.

The EEM program review process brought together many of the stakeholder representatives that had been involved in the original development of the EEM program, affording a wealth of knowledge and experiences within the team. The stakeholder representatives that participated in this process were chosen based on their knowledge and expertise with the EEM program.

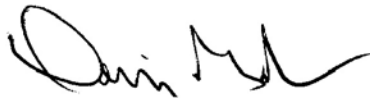
The Review Team held monthly teleconferences and four face-to-face meetings from December 2005 to February 2007. At the first meeting in January of 2006, the Review Team identified and prioritized the issues for review. Many of these issues had been identified in the previous year through the National EEM Team via the regional coordinators who implement the EEM program. A detailed workplan was developed to address these issues during the planned year-long exercise and the issues were then addressed one-by-one. Based on the outcome of these discussions, a record of decisions was developed. The list of issues, workplan, and record of decisions were reviewed at each of the face-to-face meetings. Three subgroups of experts were established early in the process to address specific issues and report back to the Review Team. They addressed fish, sublethal toxicity, and stakeholder involvement.

We would like to take this opportunity to gratefully acknowledge the commitment and dedication of all members of the Review Team to improving the EEM program. The expertise and contribution each member brought to the team and the insight into finding solutions to the issues identified in this process will serve to improve the EEM program.

This report is the culmination of more than a year's process to review the Metal Mining EEM Program. It provides an overview of the program review and resulting recommendations for consideration by Environment Canada. While the team was able to reach common ground on issues related to most components of the EEM program, it could not reach a consensus on the application of sublethal toxicity testing (SLT) within the program. However, the team expressed a need for change on the issue of SLT recognizing that the status quo was not acceptable. The range of opinions and suggestions for moving forward on this issue are articulated in this report for objective consideration by Environment Canada.

The report contains 42 recommendations which encompass improvements to the guidance for conducting EEM studies, proposed amendments to the MMER, and improvements to the implementation of the EEM program. They are targeted at improving the efficiency, effectiveness, and scientific defensibility of the EEM program.

On behalf of the Metal Mining EEM Review Team, we respectfully submit this report on EEM program improvements to Environment Canada for consideration. We look forward to your response.



Deni Gautron
Co-chair
Environment Canada, National EEM Office



Robert Prairie
Co-chair
Xstrata Zinc Inc.

Executive Summary

The *Metal Mining Effluent Regulations* (MMER), which were promulgated in 2002, require metal mines to undertake environmental effects monitoring (EEM) to ensure the adequate protection of all receiving aquatic environments by assessing effects on fish, fish habitat and the usability of fisheries resources.

During the development of the Metal Mining EEM Program, Environment Canada committed to a multi-stakeholder review after the first phase of the biological monitoring program to ensure that the program is effective, efficient and based on the most current scientific knowledge. Environment Canada initiated its review of the Metal Mining EEM Program in December 2005 by creating the Metal Mining EEM Review Team. The purpose of the Review Team was to review the experiences and results of the EEM Program from the first phase of metal mining EEM studies and based on this review to provide recommendations to Environment Canada for improving the program. The team consisted of a group of experts from government, the mining industry, and representatives from the aboriginal and environmental communities.

During the one year review, the team met regularly to discuss the main issues identified, which were grouped and are presented in this report in the following categories: program implementation, effluent and water quality, sublethal toxicity testing, cross-cutting science issues, fish survey, usability of fisheries resources, and the benthic invertebrate community survey.

Three subgroups were established to address specific issues raised for the fish survey, sublethal toxicity testing and stakeholder involvement. The Review Team was able to reach common ground on issues related to most components of the program. Consensus could not be reached regarding the application of sublethal toxicity testing; however, the range of opinions and suggestions are provided within this report for consideration.

This report summarizes the discussions and work completed by the Metal Mining EEM Review Team. The recommendations listed below are provided for Environment Canada's consideration, and if implemented, will improve the Metal Mining EEM Program.

Program implementation

Recommendation 1: Environment Canada should encourage mine operators to engage the public in EEM through existing mechanisms such as public liaison committees or other appropriate external outreach activities.

Recommendation 2: The work of the Stakeholder Involvement Subgroup should continue.

Recommendation 3: Efforts to improve communication between industry and government on EEM process issues should be continued.

Recommendation 4: Environment Canada should ask Science Committee members to prepare a list of their current and future mining-related research with the understanding that future research priorities may change after the national assessment is completed.

Recommendation 5: Environment Canada should review the TAP process and consult with all stakeholders to determine if it is working effectively and efficiently.

Recommendation 6: Environment Canada should conduct a national internal review of the level of detail for feedback and types of recommendations provided to mines on study designs and interpretative reports. Based on this review, Environment Canada should develop service standards, including timelines and required level of detail, for feedback to mines.

Recommendation 7: The current approach of updating electronic reporting systems for EEM as needed should be continued. Environment Canada should facilitate the process for mines, consultants, and data users to inform Environment Canada of electronic reporting concerns and problems in a timely manner so that modifications to the electronic reporting can be made before the next round of reporting.

Recommendation 8: The Guidance Document should be clarified to convey the message that the guidance is intended to be flexible and to provide examples of study designs to meet the EEM regulatory requirements.

Recommendation 9: The Guidance Document should be revised to indicate that mines with EEM effects that are consistent both in magnitude and geographic extent with those anticipated through the federal Environmental Assessment (EA) process, are not subjected to investigation of cause (IOC) for those designated effects.

Recommendation 10: The role of EEM monitoring as part of EA follow-up and the advantage of using EEM monitoring techniques and endpoints during EA baseline monitoring should be brought to the attention of EA practitioners and mines. Industry and Environment Canada need to improve awareness of the linkages between the two programs.

Effluent characterization and water quality survey

Recommendation 11: The Regulatory Information Submission System (RISS) should be modified to include: a) electronic submission of optional site-specific variables; b) method detection limits; c) rationale for selection of final discharge points (FDPs) such that electronic submission fulfills reporting requirements.

Recommendation 12: A yearly written report submission should be maintained until electronic reporting of data reflects adequately the content of the report.

Recommendation 13: Once the revised data reporting requirements have been finalized and the reporting issues (completeness and consistency) within RISS have been resolved, a process should be initiated to revise the requirements of records maintained on-site at the mine. These records would focus on methodological aspects of the characterization as well as QA/QC details and would be recorded and maintained at the mine for not less than 5 years.

Recommendation 14: Mines should be encouraged to use, as a complement to the Guidance Document, the document called Further Guidance for the Annual Effluent and Water Quality Monitoring Report for the Metal Mining EEM Program that was published in January 2005. It is also recommended that the information be integrated in the Guidance Document in the next revision.

Recommendation 15: Both electrical conductivity and selenium should be added to the list of required effluent and water quality variables under Schedule 5, Part 1 of the MMER.

Recommendation 16: Thallium should be added to the list of optional, site-specific variables in the Guidance Document.

Recommendation 17: The MMER should be amended to permit the exclusion of Radium 226 measurement from water quality monitoring if the mine has satisfied the MMER provisions for the reduction of frequency of effluent monitoring of radium 226.

Recommendation 18: The electronic reporting should be modified to accommodate the changes indicated in Section 5.2.2 *Relevance of variables measured in effluent characterization* following the amendment to the MMER.

Recommendation 19: Environment Canada should revise the guidance documents to indicate that when water quality exposure concentrations exceed reference concentrations by a factor of 2 or more, the extent and magnitude for which this condition exists should be estimated in the study design or biological interpretative report instead of in the annual effluent and water quality report.

Recommendation 20: Regional coordinators should verify that steps were taken by mines to ensure that method detection limits (MDLs) recommended in the Guidance for the Sampling and Analysis of Metal Mining Effluents are attained when reporting results. If recommended MDLs are not attained consistently in the future, an amendment to the MMER to include mandatory MDLs should be considered in the next review of the EEM program.

Sublethal toxicity

Recommendation 21: Technical improvements proposed by the Sublethal Toxicity Subgroup to the Regulatory Information Submission System (RISS) should be implemented.

Recommendation 22: Environment Canada should facilitate a broad discussion on the utility and application of sublethal toxicity tests within EEM and then Environment Canada should clearly articulate the goals for sublethal toxicity tests and how results will be used.

Recommendation 23: As the status quo is considered unacceptable, Environment Canada should consider changes to sublethal toxicity testing within the EEM program.

Recommendation 24: Environment Canada should consider an amendment to modify the wording of Schedule 5, s. 17(g) of the MMER to remove the requirement to determine if there is

a correlation between sublethal toxicity data and fish population, fish tissues, and benthic invertebrate community survey results.

Cross cutting science issues

Recommendation 25: The mesocosm guidance should be revised to include trophic transfer methodologies (i.e., exposure from food sources as well as water). The EEM program should continue to track the use of mesocosms and science evolution as information becomes available.

Recommendation 26: Regional EEM Offices, with the aid of the National EEM Office, should facilitate as required assistance from the Science Committee in designing site specific investigation of cause (IOC) studies. The National EEM Office should track IOC issues, resulting study designs, and associated guidance to better assist mines in future IOC.

Recommendation 27: Relevant literature should be cited in the Guidance Document to provide further information and examples of how study designs can address confounding factors.

Recommendation 28: A text box should be added to the electronic reporting to explain how the information will be used to encourage mines to report confounding factors accurately.

Recommendation 29: To assist mines in designing and conducting EEM studies in areas where there is historical contamination, the Guidance Document should be expanded to describe a range of tools that could be used to distinguish effects due to historical contamination, as well as the benefits and limitations of each tool.

Recommendation 30: The Guidance Document should be revised to indicate that the use of 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.

Recommendation 31: After the next monitoring phase, the issue of designing EEM studies in wetlands and the need for specific guidance should be reassessed.

Recommendation 32: Critical effect sizes should be developed following the next phase of EEM.

Recommendation 33: The work of the Science Committee should be better communicated to all stakeholders including the listing of related publications for the past 10 years, which should be available on the EEM Web site.

Recommendation 34: Communications between all stakeholders involved in EEM should continue beyond the mandate of the Metal Mining EEM Review Team. This should include annual stakeholder meetings or workshops to get an update and discuss new science and other issues as they arise. Periodic teleconferences of key stakeholders could also take place to ensure communications and coordination.

Fish Survey

Recommendation 35: The Guidance Document should be revised to include references for ageing techniques and structures to use, as well as to indicate that fish difficult to age, such as stickleback, should be avoided in lethal surveys.

Recommendation 36: Guidance on non-lethal sampling should be revised as needed following the review of the non-lethal studies by the Science Committee as well as additional information gained in future phases of monitoring.

Recommendation 37: The guidance on small bodied fish sampling (or study design) should be updated with new references as studies are published in the peer-reviewed literature.

Recommendation 38: The Guidance Document should clarify that the comparison of data from mixed fishing techniques should be avoided, where possible. If multiple fish techniques are used, then the same techniques must be used between exposure and reference areas and there must not be any size discrepancy introduced into the data set.

Recommendation 39: An amendment to the MMER is proposed to define an effect on the usability of fisheries resources consistent with the Health Canada fish consumption guideline. In addition, it is proposed that this definition be further modified to indicate that exposure area concentrations should be statistically significantly greater than the reference area concentrations.

Recommendation 40: The Guidance Document should be modified following the amendment to the MMER to indicate the new Health Canada fish consumption guideline as well as provide statistical guidance for conducting a one-tail test to determine if the concentration in the exposure area is statistically significantly greater than the reference area.

Recommendation 41: Assuming that commercial laboratories can meet this recommendation, the detection limit for mercury in effluent in the Guidance Document for Sampling and Analysis and the EEM Guidance Document should be revised to 0.01 µg/L in order to detect with confidence concentrations exceeding 0.1 µg/L.

Benthic invertebrate survey

Recommendation 42: The calculation method for Bray-Curtis index should be reviewed by experts from the Science Committee, based on advances reported in peer-reviewed scientific literature.

Acknowledgments

We would like to acknowledge the hard work and dedication of all the members of the Review Team. There were three subgroups formed during the Review Team process; stakeholder involvement, sublethal toxicity and fish. The Stakeholder Involvement Subgroup was chaired by Elizabeth Gardiner (Mining Association of Canada) and included Charles Dumaresq (Environment Canada), Alan Penn (Cree Regional Authority), Judy Parkman (Recycling Organization Against Rubbish), Christine Brereton (INCO), and Robert Christie (Pictou Harbour Environmental Protection). The Sublethal Toxicity Subgroup was chaired by Jim McGeer (Natural Resources Canada and Wilfrid Laurier University) and included Rick Lowell (Environment Canada), Bernard Vigneault (Natural Resources Canada), Lisa Taylor (Environment Canada), Rick Scroggins (Environment Canada) and Arden Rosaasen (AREVA Resources Canada Inc.). The Fish Subgroup was chaired by Mark McMaster (Environment Canada) and included Vince Palace (Fisheries and Oceans Canada), Malcolm McKee (Canadian Nuclear Safety Commission), Raymond Chabot (Environment Canada – regional representative), Christine Brereton (INCO) and Michelle Bowerman (Environment Canada). Thanks to the dedication and hard work of the chairs and members of the subgroups. All of these individuals strived to review and discuss the issues identified by the Review Team and recommend solutions within very tight timelines.

We would also like to thank the regional EEM Coordinators (Sue Ellen Maher, Raymond Chabot, Nardia Ali, Jenny Ferone and Mike Hagen) and their staff who responded to numerous requests for information and data on very short timelines.

A special note of thanks to the many research scientists, including Mark McMaster, Vince Palace, Joseph Culp, Monique Dubé, Bernard Vigneault, Jim McGeer, Kelly Munkittrick and Nancy Glozier, who found time from their busy schedules to participate in meetings, present new research, and contribute their scientific expertise to help further the sound scientific basis behind EEM.

Table of Contents

Foreward	iii
Executive Summary	iv
Acknowledgments	ix
1 Introduction	1
2 Background.....	2
2.1 Program Overview	2
2.2 Effluent Characterization and Water Quality Studies	4
2.2.1 Effluent characterization and water quality monitoring.....	4
2.2.2 Sublethal toxicity testing.....	5
2.3 Biological Monitoring Studies	6
2.3.1 Fish monitoring	6
2.3.2 Benthic invertebrate community survey	7
2.3.3 Use of fisheries resources.....	8
2.4 Program Administration.....	8
3 Metal Mining EEM Review Team	9
3.1 Mandate and Work of the Metal Mining EEM Review Team.....	9
3.1.1 Purpose of the Metal Mining EEM Review Team Report.....	10
4 National Assessment of the Results of Phase 1 of the Metal Mining EEM Program.....	11
5 Program Review Issues and Recommendations.....	11
5.1 Program Implementation.....	11
5.1.1 Public involvement in EEM	11
5.1.2 Government-industry communication.....	12
5.1.3 National consistency	13
5.1.4 Information management: database and electronic reporting	13
5.1.5 Flexibility in implementation of guidance	13
5.1.6 EEM and environmental assessment.....	14
5.2 Effluent Characterization and Water Quality Survey	15
5.2.1 Reporting.....	15
5.2.2 Relevance of variables measured in effluent characterization	16
5.2.3 Guidance for estimating geographic extent of effects for water quality	17
5.2.4 Use of appropriate detection limits	17
5.3 Sublethal Toxicity	18
5.3.1 Review of sublethal toxicity data	18
5.3.2 Correlation of sublethal toxicity data with field responses	20
5.4 Cross Cutting Science Issues.....	21
5.4.1 Alternatives - review of mesocosms in EEM.....	21
5.4.2 Guidance for investigation of cause	21

5.4.3 Confounding factors	22
5.4.4 Historical contamination	22
5.4.5 Wetlands	23
5.4.6 Reference areas	23
5.4.7 Establishing effect sizes	24
5.4.8 Ensuring scientific evolution.....	24
5.5 Fish Survey.....	25
5.5.1 Ageing techniques for small bodied fish.....	25
5.5.2 Non-lethal vs lethal surveys	25
5.5.3 Guidance on small bodied fish	26
5.5.4 Interpretation of significant interactions	26
5.5.5 Analysis carried out with fish caught through different methods	27
5.5.6 Absence of fish.....	27
5.6 Fish Tissue Analysis (Use of Fisheries Resources)	27
5.6.1 Fish consumption guideline	28
5.6.2 Appropriate detection level for mercury in effluent.....	28
5.7 Benthic Invertebrate Survey.....	29
5.7.1 Statistical approach – univariate vs multivariate.....	29
5.7.2 Bray-Curtis index	29
6 Conclusion.....	30
7 References	31
Appendix A: Metal Mining EEM Review Team Terms of Reference.....	33
Appendix B: List of Issues	40
Appendix C: Barriers to Public Involvement in EEM	41
Appendix D: Sublethal Toxicity Subgroup Report.....	44

List of Tables

Table 1 Variables measured in effluent characterization and water quality monitoring studies. ...	5
Table 2 Types of responses and indicators for the lethal and non-lethal fish survey.	7
Table 3 Core indicators for the benthic invertebrate community survey.....	8

List of Figures

Figure 1 Sequence of events in the Metal Mining Environmental Effects Monitoring Program. ...	3
-----------------------------------------------------------------------------------------------	---

1 Introduction

In 2002, the federal government promulgated the *Metal Mining Effluent Regulations* (MMER) under the *Fisheries Act*. The MMER replaced the *Metal Mining Liquid Effluent Regulations*, which had been in place since 1977. The MMER regulate the quality of effluent discharged by mines producing base metals, precious metals, iron ore, uranium, and other metals. The MMER include limits on pH and concentrations of arsenic, copper, cyanide, lead, nickel, zinc, total suspended solids and radium 226. The Regulations also require that effluent be non-acutely lethal to rainbow trout. Mines are required to conduct Environmental Effects Monitoring (EEM) to evaluate the effects of mining effluent on the aquatic environment, specifically fish, fish habitat, and the use of fisheries resources. Results from the EEM program will provide a comprehensive overview of effects of metal mining on water quality and aquatic organisms. Amendments to the MMER, were published in October of 2006.

The requirement for EEM was included as a component of the MMER because the Canadian metal mining industry operates in many different types of ore deposits, uses a variety of mining and milling processes, and discharges effluents to a variety of receiving environments. For these reasons, it was unknown whether national minimum effluent quality limits alone would ensure adequate protection of all receiving environments. Thus, all mines subject to the MMER are required to conduct an EEM program to evaluate the effects of metal mine effluents on fish, fish habitat, and the usability of fisheries resources. The EEM program is a scientific monitoring approach that can be used to help determine the effects in aquatic ecosystems caused by metal mine effluent and the effectiveness of environmental protection measures.

There are two key components to the EEM program:

- Effluent and Water Quality Monitoring Studies include effluent characterization, water quality monitoring and sublethal toxicity testing with an annual reporting requirement. These studies are used as supporting information to help interpret biological data,
- Biological Monitoring Studies include a fish survey, a benthic invertebrate community survey to assess impacts on fish habitat, and a survey of mercury tissue levels in fish to assess impacts on the usability of fisheries resources.

Development of the EEM program for metal mining under the MMER was based on recommendations from the multi-stakeholder Metal Mining EEM Working Group (1997 – 1999), which was comprised of government, industry, and representatives from the aboriginal and environmental communities. During the development of the program, Environment Canada committed to reviewing the overall effectiveness of the program, approximately three years after its commencement, upon completion of the first phase of biological monitoring. To meet this commitment, Environment Canada established the Metal Mining EEM Review Team to undertake this program review based on experiences learned from the first phase of monitoring. The Review Team consisted of a group of experts from government, industry, environmental and aboriginal communities. The role of the Review Team was to provide Environment Canada with recommendations to improve the effectiveness, efficiency, and scientific defensibility of the EEM program.

The review commenced in December 2005, approximately six months before completion of the first phase of monitoring.

2 Background

2.1 Program Overview

The EEM program is designed to be nationally consistent in its approach, while allowing technical flexibility to respond to site-specific issues. The program is founded on a tiered or phase approach to monitoring, which allows for more extensive and more frequent monitoring at sites where there are effects and less monitoring at sites where no effects are found. Each phase of monitoring builds upon the information obtained from previous studies.

Biological monitoring is conducted in the water bodies into which mine effluent is discharged, or in water bodies downstream from the effluent discharge point. These are termed “exposure areas” and are also known as “the receiving environment”. To provide a basis for the comparison of results, biological monitoring is also conducted on ecologically similar water bodies that are not exposed to the mine effluent. These are termed “reference areas.”

Each phase of biological monitoring for the EEM program begins with the preparation of a study design which is submitted to Environment Canada for review before field studies are undertaken. Field studies include site characterization to identify the key physical and chemical characteristics of the exposure and reference areas, and biological monitoring to determine if there are effects on fish populations, benthic invertebrate communities, and fish usability. Following the completion of the field studies, an interpretive report is prepared and submitted to Environment Canada, summarizing the results of the monitoring, and providing a discussion and interpretation of these results, as well as a conclusion regarding the next steps for monitoring at the site.

Key to the phased character of the program is that the monitoring requirements are based on the results of previous biological monitoring studies. Thus, if biological effects are identified in the first phase of monitoring, then the objective of the next study is to confirm the presence of those effects. If the effects are confirmed, then the objective of the subsequent phase of monitoring would be to determine the magnitude and geographic extent of the effects. If the magnitude and the geographic extent of an effect have been determined and the cause of the effect is not known, the monitoring will move into a phase known as investigation of cause. Conversely, if no biological effects are identified, then biological monitoring continues on a less frequent basis. These phases of the EEM program, and the associated monitoring frequencies, are illustrated in Figure 1.

Note that it is important to consider how effects are defined in the EEM program. Specific definitions are provided in the MMER, but, essentially, an effect is a statistically significant difference in fish population or benthic invertebrate community indicators taken in an exposure area and reference area (or along a gradient of effluent exposure), or an exceedence of a fish tissue consumption guideline for mercury in fish exposed to the effluent.

It is also important to note that the EEM program was designed to allow mines to build upon the results of monitoring conducted before the EEM requirements came into force. About half of the mines subject to the MMER had historical monitoring data. Mines that did not have historical monitoring data were required to submit study designs to Environment Canada within 12 months of the MMER coming into force, and had to submit their first interpretive reports by June 2005. Mines with historical data were required to submit that data, and then had until June 2006 to submit their interpretive reports for the first phase of EEM.

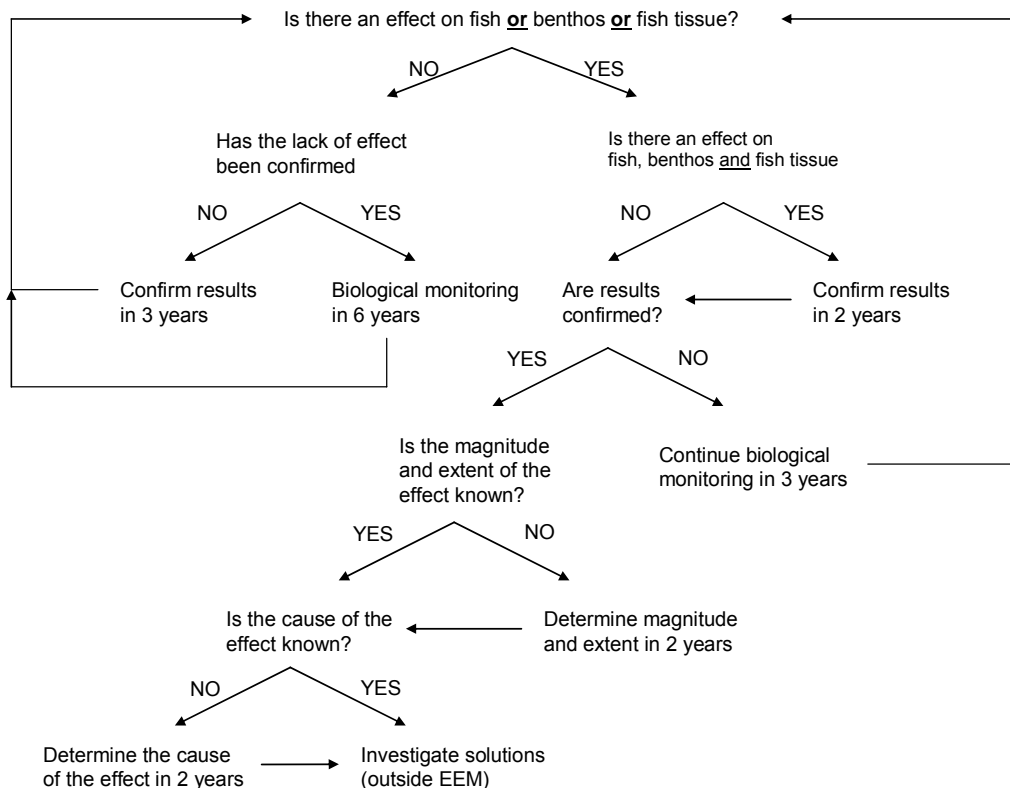


Figure 1 Sequence of events in the Metal Mining Environmental Effects Monitoring Program.

2.2 Effluent Characterization and Water Quality Monitoring Studies

Effluent characterization and water quality monitoring studies consist of three key components:

- effluent characterization;
- water quality monitoring; and
- sublethal toxicity testing.

2.2.1 Effluent characterization and water quality monitoring

The objective of effluent characterization and water quality monitoring is to answer the question “what is the estimated mine-related change in contaminant concentrations in the exposure area?”

Data generated from effluent characterization and water quality monitoring are used to:

- monitor changes in mining operations 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 samples of effluent to provide information on the concentrations of contaminants in the mine effluent. Samples are analyzed for pH and the substances regulated under the MMER, termed “deleterious substances” as well as other potential contaminants and supporting parameters (Table 1). Samples are collected four times per calendar year.

Water quality monitoring is conducted four times a year by collecting and analyzing samples of water from the exposure and reference areas. In addition, samples of water are collected from sampling areas in receiving environments where biological monitoring is completed, at the same time that biological monitoring is conducted. Samples for water quality monitoring are analyzed for the same parameters measured for effluent characterization (Table 1) as well as dissolved oxygen and temperature.

Table 1 Variables measured in effluent characterization and water quality monitoring.

Regulated Variables^{1,2}	Required Additional Variables^{2,4}	Required Supporting Variables	Recommended Site-Specific Variables⁹
Arsenic	Aluminium	Alkalinity ^{3,5,7}	Fluoride
Copper	Cadmium	Total hardness ^{3,5,7}	Manganese
Lead	Iron	Dissolved oxygen ⁵	Selenium
Nickel	Mercury ⁶	Temperature ⁵	Uranium
Zinc	Molybdenum	Salinity ^{5, 7, 8}	Total phosphorus
Radium 226	Ammonia		Conductivity
Total cyanide	Nitrate		Calcium
Total suspended solids			Chloride
pH ³			Magnesium
			Potassium
			Sodium
			Sulphate
			Total thiosalts
			Water depth ⁵
			Optical depth or transparency ⁵
			Dissolved organic carbon
			Total organic carbon

- 1) List of parameters regulated (deleterious substances and pH) as per Schedule 3 of the MMER; concentration limits specified in the regulation.
- 2) All concentrations are total values; dissolved concentrations may also be reported; effluent loading will also be calculated and reported.
- 3) Required for all mines for effluent characterization, but for water quality monitoring pH is only required for mines depositing to freshwater.
- 4) List of parameters required for effluent characterization and water quality monitoring as per Schedule 5 of the MMER.
- 5) Parameters measured for water quality monitoring in receiving waters (exposure and reference areas).
- 6) Analyses of mercury may be discontinued if the concentration of total mercury in effluent is less than 0.1 µg/L in 12 consecutive samples of effluent.
- 7) For mines depositing to estuarine waters.
- 8) For mines depositing to marine waters.
- 9) 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.

2.2.2 Sublethal toxicity testing

The purpose of sublethal toxicity testing is to answer the question “is there evidence from sublethal toxicity testing that mine effluent may affect fish, invertebrates and aquatic plants?” Acceptable sublethal toxicity data generated through the testing of an effluent from a specific discharge location over time can provide an indication of the degree of variability in effluent quality and temporal or seasonal trends.

As outlined in the EEM Guidance Document (2002), there are two potential uses of sublethal toxicity tests in the EEM program:

1. To estimate the potential for exposure area impacts and to estimate the potential geographic extent of the sublethal response using the results from toxicity tests conducted on mining effluent.
2. To measure changes in effluent quality over time as a result of changes in water treatment, site runoff management, acid rock drainage (ARD) mitigation measures or mine process changes.

To answer the above question mines conduct a suite of four freshwater sublethal toxicity tests (fish, invertebrate, plant and algae) or three main sublethal toxicity tests (fish, invertebrate and algae), depending on the receiving environment type. Tests are conducted and reported in accordance with standard test protocols developed by Environment Canada. These tests were required twice per year for the first three years after mines are subject to the EEM requirements of the MMER, and are then required to conduct these tests once a year thereafter.

2.3 Biological Monitoring Studies

As described in Section 2.1 biological monitoring is conducted every two, three or six years depending upon the results of previous monitoring. Effects of effluent on fish are assessed through comparison of adult fish populations exposed to effluent with unexposed adult fish populations or along a gradient of decreasing effluent concentrations. Effects on fish habitat are assessed through comparison of benthic invertebrate communities from areas exposed and unexposed to effluent or along a gradient of decreasing effluent concentrations. Measurements of mercury in fish tissue assess effects on the usability of fisheries resources.

2.3.1 Fish monitoring

Fish monitoring for the EEM program consists of monitoring the fish to determine if there are differences in the growth, reproduction, survival or condition of fish populations, in order to determine whether or not the mine effluent is having an effect on fish.

Fish monitoring is carried out by means of an adult fish survey, which compares differences in survival (age structure), energy use (growth and reproduction) and energy storage (condition) of fish exposed to effluent discharge with those of unexposed fish collected from a reference area or from areas of gradually increasing distance downstream of a discharge.

Adult fish surveys may consist of the lethal sampling of fish, or non-lethal sampling. For lethal sampling, the Guidance Document recommends that 20 adult males and 20 adult females from two species be sampled. For non-lethal fish surveys, 100 individuals are recommended. More samples are required for non-lethal surveys since there are fewer parameters that can be measured when the fish are not killed during the survey.

There are five core endpoints required for the fish survey (Table 2). These endpoints are regarded as the principle measures that integrate the performance of the fish associated with their environmental conditions. These approaches have been under development for more than a decade (Munkittrick and Dixon 1989 a,b; Munkittrick 1992; Gibbons and Munkittrick 1994; Munkittrick et al. 2000).

Table 2 Types of responses and indicators for the lethal and non-lethal fish survey.

a) Type of Response	Lethal Endpoint (20 adult males and 20 adult females; 2 species)	Non-Lethal Endpoint (100 individuals)
Survival	Age	Relative age class strength
		Length frequency distribution
Energy Use	Weight at age	Weight at age
		Size of young of year (YOY)
	Relative gonad size	Relative abundance of YOY
Energy Storage	Condition	Condition
	Relative liver size	

2.3.2 Benthic invertebrate community survey

In addition to having the potential to affect fish directly, metal mine effluents may affect fish indirectly through changes in their habitat. To determine if mine effluent discharge is affecting the fish habitat, mines are required to conduct a benthic invertebrate community survey. Benthic invertebrates are small organisms, commonly insect larvae, which live on the bottom of lakes and streams. As an important source of food for various fish species, they are an important component of fish habitat.

The benthic invertebrate community survey is a standard monitoring tool that has been widely applied in environmental management as an indicator of the condition of fish habitat. Statistical differences in four core indices (Table 3) between exposure and reference areas are used to quantify effects on the benthic invertebrate community structure and function. These indices describe a range of responses, including changes in productivity, species composition and biodiversity, which may result from changes in environmental conditions due to factors such as the discharge of mine effluent (Glozier et al. 2002). The Bray-Curtis index summarizes the overall community differences between the reference and the exposed areas (Environment Canada 2002).

Table 3 Core indicators for the benthic invertebrate community survey.

Measurement	Indicator
Total density	number of animals per unit area
Taxon Richness	number of taxa or kinds of animals
Simpson's Evenness	measure of how evenly the animals are distributed among taxa
Bray-Curtis Index	measure of overall community composition differences between the reference and the exposed areas

2.3.3 Use of fisheries resources

The third component of biological monitoring is an assessment of whether or not mining effluents have affected the use of fisheries resources. For the metal mining EEM program, concentrations of mercury in fish tissue are used as an indicator of fish usability, since elevated levels of mercury in fish tissue can make fish unsafe for human consumption.

To assess effects of fish usability, mines are required to conduct a fish tissue analysis for mercury if this metal is found in the effluent at a concentration greater or equal to 0.1 µg/L. A fish tissue study is conducted in both an exposure and a reference area. An effect on fish usability is defined as measurements of total mercury that exceed 0.45 mg/kg wet weight in fish taken in an exposure area and that are statistically different from the measurements of total mercury in fish taken in a reference area.

2.4 Program Administration

The Metal Mining EEM Program is a complex scientific program, involving detailed legal requirements, as well as extensive technical guidance to practitioners to help them meet the legal requirements. Clear communications between all involved in the program are key to its success.

The EEM program was established by Environment Canada, and the Department plays an important role in facilitating communications between all involved in the program, as well as facilitating the ongoing development of the program. To provide a leadership and facilitation role, Environment Canada has established the following:

National EEM Office: consisting of staff from Environment Canada headquarters, it provides leadership for the regulated EEM programs for both the pulp and paper and metal mining sectors and ensures the scientific evolution of the programs. Two of the many roles of the National EEM Office are to develop EEM Guidance Documents on how to conduct EEM studies and to conduct national analysis of the EEM data.

Science Committee: the Science Committee, which is comprised of scientific experts in all aspects of the program, serves to ensure that the EEM program continues to evolve with our

scientific understanding, and offers technical expertise needed for effective design and implementation of the program.

National EEM Team: the National EEM Office works closely with the regional Environment Canada staff, the Science Committee and other government representatives through the National EEM Team to ensure that EEM programs are implemented in a meaningful, cost-effective, and nationally consistent manner.

Technical Advisory Panels (TAPs): TAPs have been formed in regions to provide technical and scientific advice on EEM study designs and interpretative reports and are comprised of regional coordinators, and representatives from provincial and other government departments.

3 Metal Mining EEM Review Team

3.1 Mandate and Work of the Metal Mining EEM Review Team

During the development of the Metal Mining EEM Program, Environment Canada committed to a multi-stakeholder review after the first phase of the program. In response to this commitment, the multi-stakeholder Metal Mining EEM Review Team was established in the fall of 2005.

The purpose of the Review Team was to review the experiences and results of the EEM program from the first phase of EEM studies conducted at Canadian metal mines and, based on this review, to provide recommendations to Environment Canada for improving the program. The Review Team brought together experts that were knowledgeable of the EEM program, as well as experts in the field of metal mining and effects on aquatic ecosystems. This team of experts included representatives from the federal government (Environment Canada, Fisheries and Oceans Canada, Natural Resources Canada, Canadian Nuclear Safety Commission), mining industry, and environmental and aboriginal communities.

The key tasks of the Review Team included:

- Identify questions to be addressed in the national assessment of the first phase of the metal mining EEM data (completed by the National EEM Office);
- Review all aspects of the EEM program to identify what worked and where problems were encountered;
- Identify and assess possible solutions to resolve the issues and problems identified;
- Review results of the national assessment; and
- Submit recommendations for improving the Metal Mining EEM Program to Environment Canada.

A list of members of the Metal Mining Review Team is included in the terms of reference of the Review Team in Appendix A.

At the first face-to-face meeting of the Review Team, issues for review were identified by members during a brainstorming process. The issues were then grouped into categories to be addressed. These categories included:

- program implementation;
- effluent characterization and water quality;
- sublethal toxicity testing;
- cross-cutting science issues;
- fish survey and fish usability survey; and
- benthic invertebrate community survey.

A workplan was then developed to address each of the issues through teleconferences, face-to-face meetings, or the work of subgroups. A comprehensive list of all issues considered by the Review Team is provided in Appendix B. Three subgroups were established to address in detail specific issues relating to the fish survey, sublethal toxicity testing, and stakeholder involvement in the EEM program. Review Team members were given the opportunity to bring forward new issues for discussion throughout the process. Further, at the first meeting, members were asked what types of data analysis should be done at a national level. Views and ideas were discussed through four meetings and monthly teleconferences between December 2005 and February 2007 and lead to the production of the report.

3.1.1 Purpose of the Metal Mining EEM Review Team Report

The purpose of the Metal Mining EEM Review Team Report is to summarize the work of the Review Team, and to present to Environment Canada the Review Team's conclusions and recommendations regarding changes to the EEM program. Recommendations will be considered by Environment Canada in the continuing development of the EEM program.

The recommendations presented in the Review Team Report are based primarily on the collective expertise and experiences of the Review Team members, as well as consideration of some of the national results of the first phase of the Metal Mining EEM Program and perspectives offered by the Science Committee. The Review Team did not conduct a detailed assessment of the results, nor has it seen the final results of the national assessment at the time of this report. This detailed national assessment is being conducted by Environment Canada and will be published separately, in an Environment Canada report.

In developing these conclusions and recommendations, the Review Team recognized certain limitations inherent in this review; in particular, the fact that the single data set provided by the first phase of the EEM program was not likely to provide sufficient information for substantial program changes. Nevertheless, the Review Team regards the recommendations in the report as important recommendations which, if implemented, would help to improve the program. In the future, when more data are available, a more in depth review could lead to further improvements to the program.

4 National Assessment of the Results of Phase 1 of the Metal Mining EEM Program

The national assessment of the results of phase 1 of the Metal Mining EEM Program was prepared by Environment Canada, separately but in parallel with the work of the Review Team. The assessment report will be peer reviewed prior to its release during 2007.

The purpose of the national assessment is to present and discuss, on a national, rather than site-specific or regional basis, the results of EEM biological data collected in 2004 and 2005 as part of the first phase of the Metal Mining EEM Program.

It is important to note that the activities of the Review Team, and the preparation of the national assessment, did not occur in isolation. The Review Team was given the opportunity to identify questions and specific analyses to be considered in the preparation of the national assessment. In addition, the Review Team had an opportunity to see preliminary results of the national assessment as they became available. This informed the ongoing discussions of the Review Team. However, it was not possible for the Review Team to conduct a detailed review of results from the national assessment as the final report was not completed.

The National Assessment will address, among other things, the following key questions:

- 1) What are the types and magnitude of effects of mine effluent on adult fish and benthic invertebrate communities?
- 2) How are the effects influenced by ore type, habitat, gender, species, and continuous versus intermittent discharges?
- 3) What are the effects of mine effluent on the usability of fisheries resources?
- 4) What are the ranges of effect sizes?

5 Program Review Issues and Recommendations

5.1 Program Implementation

5.1.1 Public involvement in EEM

As part of the review of the metal mining EEM program, the Stakeholder Involvement Subgroup was established to address the issue of public involvement in EEM, keeping in mind the three key questions of the EEM review:

- What's working?
- What's not working?
- How do we fix it?

Specifically, this Subgroup worked to improve the understanding of the current state of public involvement in EEM, and to identify means to facilitate increased public involvement.

In general, the Subgroup members agreed that the public involvement component of the program has not worked well to date.

The Subgroup is conscious of the needs of interested communities and organizations for practical information on the current status of the EEM program and for opportunities for constructive involvement in EEM. Subgroup members also recognize the need to have realistic expectations about the degree to which the public can be involved, given the specific, regulated and often tight timelines, as well as the highly technical nature and very specific focus of the program. The Subgroup developed a document (Appendix C) that identifies barriers to public involvement in EEM and discussions on this issue should continue. Future proposed activities for the Subgroup include:

- a review of Chapter 10 of the EEM Guidance Document on public involvement;
- consideration of the potential benefits of holding a multi-stakeholder workshop on public involvement in EEM;
- development of a document to help build capacity among the public with respect to the EEM program. This document would provide basic information on EEM to the public and describe how the public can become more involved in the program;
- development of material that specifically helps mine operators to engage the public. This material, promoting the potential benefits of public involvement in EEM, would describe ways in which the public can become more involved in EEM, and how a mine operator can increase public involvement.

Recommendation 1: Environment Canada should encourage mine operators to engage the public in EEM through existing mechanisms such as public liaison committees or other appropriate external outreach activities.

Recommendation 2: The work of the Stakeholder Involvement Subgroup should continue.

5.1.2 Government-industry communication

Some of the key areas in EEM where difficulties and inconsistencies in government-industry communication were identified as issues included: review of historical information (if applicable), study design review and approval, EEM field survey, and feedback on the EEM interpretive report.

Overall, the team agreed that communication among industry and government is good but continual efforts for improvement should be made. However, it was noted that the timeliness of feedback from authorities needs to be predictable, reasonable and consistent among regions. This latter issue is addressed under the section on national consistency.

Recommendation 3: Efforts to improve communication between industry and government on EEM process issues should be continued.

Recommendation 4: Environment Canada should ask Science Committee members to prepare a list of their current and future mining-related research with the understanding that future research priorities may change after the national assessment is completed.

5.1.3 National consistency

Industry raised concerns regarding national consistency particularly with respect to timeliness and content of the feedback to the mines, which was quite varied among regions both for study design and interpretive reports. Environment Canada has acknowledged that there are problems in this area, often related to resourcing disparities among regions as well as the complexity of dealing with some Technical Advisory Panels (TAPs). There was concern about the variability in TAP membership among regions and whether this affected the quality of feedback for study designs and interpretative reports. The group thought that TAP membership should be as broad as necessary, provided that the quality of feedback on study designs and interpretative reports was appropriate, manageable and acceptable. It was also noted that some regions choose to be more active than others, and that this can be dependent upon regional resources.

Recommendation 5: Environment Canada should review the TAP process and consult with all stakeholders to determine if it is working effectively and efficiently.

Recommendation 6: Environment Canada should conduct a national internal review of the level of detail for feedback and types of recommendations provided to mines on study designs and interpretive reports. Based on this review, Environment Canada should develop service standards, including timelines and required level of detail, for feedback to mines.

5.1.4 Information management: database and electronic reporting

The EEM program information management system includes the biological data entry software and associated national database for the biological monitoring data. Mines are required to submit their data associated with their interpretive reports electronically every two, three or six years, depending upon their reporting requirements, as determined by the results of previous monitoring phases. In addition, mines submit data from their effluent characterization, water quality monitoring, and sublethal toxicity testing annually, by March 31, using the Regulatory Information Submission System (RISS), which is also used for the quarterly submission of compliance data. Currently, the database and electronic reporting system are updated periodically as issues are identified. As this system appears to be working well, the team recommended that this system be continued.

Recommendation 7: The current approach of updating electronic reporting systems for EEM as needed should be continued. Environment Canada should facilitate the process for mines, consultants, and data users to inform Environment Canada of electronic reporting concerns and problems in a timely manner so that modifications to the electronic reporting can be made before the next round of reporting.

5.1.5 Flexibility in implementation of guidance

Concerns were raised that the Guidance Document for Metal Mining EEM does not include options for study designs to compare against baseline conditions for historical data. It was clarified that the guidance document is intended as a starting point for study designs and is intended to be flexible to allow creativity in design of studies to accommodate site-specific needs. This could include before-after control impact (BACI) designs, which are used to compare current conditions from effluent exposure against baseline conditions, prior to effluent discharge or distinguish effects from current discharge from historical contamination. Further discussions related to this issue acknowledged that a before-after design would be useful and could be suggested to assess other issues (i.e., to distinguish historical impacts from current discharge) but could not be required as mines would not be required to do EEM until they are subject to the MMER.

Recommendation 8: The Guidance Document should be clarified to convey the message that the guidance is intended to be flexible and to provide examples of study designs to meet the EEM regulatory requirements.

5.1.6 EEM and environmental assessment

Most new mines are required to conduct environmental assessments (EAs) before being able to obtain permits to begin construction and ultimately enter production. In general terms, the objective of EAs is to assess the potential impacts of a proposed project on the environment and human health. Information from EAs is used in deciding whether to proceed with proposed projects, and the recommendations in EAs may include specific remedial action to reduce or eliminate potential impacts, and to conduct follow-up monitoring. EAs generally include monitoring to determine baseline environmental conditions, prior to the beginning of the development of the proposed project. The baseline data is then used in the predictions of potential environmental effects of the development. Baseline data may also be used as a basis for comparison with the results of monitoring conducted after the development begins, to assess the actual versus the predicted impacts.

The Review Team discussed the linkages between EEM and EA. In particular, there are potential linkages between EEM and anticipated potential environmental impacts developed as part of the EA process. It was recognized that, in designing baseline monitoring studies for proposed mining developments, the future requirements with respect to EEM should be taken into consideration. While the scope of baseline studies for EAs frequently extends far beyond the scope of EEM, the results of future EEM work can be used to follow-up on predictions made during the EA. This can only happen however, if comparable monitoring methods and endpoints are used. Thus, there was agreement that, for proposed mining projects conducting baseline monitoring studies for EA, the aquatic component of these studies should assess, as a minimum, the endpoints used in the EEM program. In addition, where mines are required to conduct follow-up monitoring as part of the EA process, EEM could be used to satisfy some of these requirements, if comparable methods and endpoints were used in the baseline monitoring. Therefore, when the needs of both programs are taken into consideration, EEM results can provide important input to the EA process, and the need for overlapping monitoring requirements to satisfy the different programs can be avoided.

It was acknowledged that the objectives of EEM and follow-up monitoring conducted to satisfy EA requirements may not be as compatible once EEM enters the investigation of cause phase. In EA follow-up monitoring, the magnitude and extent of effects are assessed, but generally, the causes of observed effects are only investigated if the effects exceed the level of effects that were determined to be acceptable through the EA process. However, the assessment of whether effects are considered to be acceptable is considered to be outside the scope of EEM. Once the magnitude and extent of effects have been determined, then mines are required to investigate the causes of these effects as part of the EEM program, regardless of whether or not the EA may have determined that the effects observed would be considered acceptable.

Recommendation 9: The Guidance Document should be revised to indicate that mines with EEM effects that are consistent both in magnitude and geographic extent with those anticipated through the federal Environmental Assessment (EA) process are not subjected to investigation of cause (IOC) for those designated effects.

Recommendation 10: The role of EEM monitoring as part of EA follow-up and the advantage of using EEM monitoring techniques and endpoints during EA baseline monitoring should be brought to the attention of EA practitioners and mines. Industry and Environment Canada need to improve awareness of the linkages between the two programs.

5.2 Effluent Characterization and Water Quality Survey

5.2.1 Reporting

Concerns were raised about the need for an annual written report from each mine site and the variability in the format of reports on effluent characterization and water quality. EEM regional and national staff currently use written reports to check for any missing RISS EEM data and to verify the information provided. This is done because of concerns about the quality of effluent characterization and water quality EEM data in RISS. This data verification is important for regional and national analyses of the data to ensure confidence in these analyses. The Review Team agreed that mines should adhere to a consistent format for their paper reports. However, this format should be harmonized with other agencies to which these reports are submitted. The Review Team also noted that mines monitor a number of optional, site-specific variables. Although these data are provided in the written reports, the data are not available electronically and, as such, not readily accessible.

Recommendation 11: The Regulatory Information Submission System (RISS) should be modified to include: a) electronic submission of optional site-specific variables; b) method detection limits; and c) rationale for selection of final discharge points (FDPs) such that electronic submission fulfills reporting requirements.

Recommendation 12: A yearly written report submission should be maintained until electronic reporting of data reflects adequately the content of the report.

Recommendation 13: Once the revised data reporting requirements have been finalized and the reporting issues (completeness and consistency) within RISS have been resolved, a process should be initiated to revise the requirements of records maintained on-site at the mine. These records would focus on methodological aspects of the characterization as well as QA/QC details and would be recorded and maintained at the mine for not less than 5 years.

Recommendation 14: Mines should be encouraged to use, as a complement to the Guidance Document, the document called Further Guidance for the Annual Effluent and Water Quality Monitoring Report for the Metal Mining EEM Program that was published in January 2005. It is also recommended that the information be integrated in the Guidance Document in the next revision.

5.2.2 Relevance of variables measured in effluent characterization

The Review Team revisited the list of variables for effluent characterization and water quality monitoring. Early in the review process, several variables were identified for potential addition to the list of variables monitored for EEM including electrical conductivity, thallium, selenium and rubidium. There were no variables proposed for removal from the list of required variables for monitoring.

Selenium toxicity in freshwater fish associated with metal mining was presented by Dr. Vince Palace (Fisheries and Oceans Canada) with supporting background information provided by Dr. Glen Bird (formerly with the Canadian Nuclear Safety Commission). Although selenium is an essential dietary nutrient, there is a very narrow range between dietary requirements and toxic thresholds, which can occur at two to seven fold above background (Maier and Knight 1994, Lemly 1993, 1996). Selenium bioconcentrates in lower trophic organisms (Besser et al. 1989, 1993) and small increases in waterborne selenium can lead to reproductive effects in fish that acquire elevated selenium from contaminated food webs (Skorupa 1998). Selenium does not result in sudden fish kills at environmental concentrations, but instead exerts teratogenic effects (increased rates of developmental deformities in early life stages) in fish and other animals that lay eggs (i.e. oviparous) such as birds and reptiles (Lemly 1996). Some metal mines have shown elevated levels of selenium in their receiving environments, which could cause teratogenic effects.

Conductivity is a measure of the water's ability to conduct an electrical current and is used to measure the amount of dissolved ions in the water. Conductivity is used to identify the location and concentration of an effluent plume in freshwater environments and is used by almost all mines in their plume delineation studies.

Although it was suggested that metal mining is a source of thallium to aquatic environments, the Review Team found that there was insufficient evidence presented on the toxic effects and the contribution of metal mining effluents to thallium releases to justify inclusion of this variable in the EEM program. It was noted that this variable could be considered as an optional variable on a site-specific basis.

Although rubidium was considered for addition to the list of monitored variables, there was insufficient evidence on either sources, or toxicity of rubidium to justify including this variable in the EEM program at this time.

The MMER include provisions (subsection 13(2)) to reduce the frequency of monitoring of radium 226 in effluent if the concentration of radium 226 is less than 0.037 Bq/L in 10 consecutive tests. It was suggested during the EEM review that if radium 226 concentrations in effluent are low, then there is no need to measure radium 226 as part of water quality monitoring.

Recommendation 15: Both electrical conductivity and selenium should be added to the list of required effluent and water quality variables under Schedule 5, Part 1 of the MMER.

Recommendation 16: Thallium should be added to the list of optional, site-specific variables in the Guidance Document.

Recommendation 17: The MMER should be amended to permit the exclusion of Radium 226 measurement from water quality monitoring if the mine has satisfied the MMER provisions for the reduction of frequency of effluent monitoring of radium 226.

Recommendation 18: The electronic reporting should be modified to accommodate the changes indicated in section 5.2.2 *Relevance of variables measured in effluent characterization* following the amendment to the MMER.

5.2.3 Guidance for estimating geographic extent of effects for water quality

Both the Metal Mining EEM Guidance Document and the Further Guidance for the Annual Effluent and Water Quality Monitoring Report for the Metal Mining EEM Program recommend a comparison of water quality data be conducted between the exposure and reference areas. In addition, for variables showing concentrations with differences that are greater in the exposure area by a factor of two, it is recommended that the mine estimate the geographic extent for which this condition exists, based on expanded water quality monitoring or modelling. It is currently recommended that the estimate be reported in the annual effluent and water quality monitoring report. The intent of this comparison is to support the biological monitoring results and to help determine appropriate sampling sites. As such, it has been suggested that this comparison would be more appropriate for the biological interpretive report or study design rather than the annual effluent and water quality monitoring report.

Recommendation 19: Environment Canada should revise the guidance documents to indicate that when water quality exposure concentrations exceed reference concentrations by a factor of 2 or more, the extent and magnitude for which this condition exists should be estimated in the study design or biological interpretative report instead of in the annual effluent and water quality report.

5.2.4 Use of appropriate detection limits

In the review of the effluent characterization and water quality data, it was found that detection limits reported by mines are nationally inconsistent, vary over several orders of magnitude, and frequently exceed method detection limits (MDLs) recommended in the Guidance for the Sampling and Analysis of Metal Mining Effluents. Without national consistency in MDLs, these data are not very useful for monitoring changes in effluent or receiving environment conditions, variability in effluent quality, temporal or seasonal trends, or in interpreting results from the biological monitoring and sublethal toxicity testing.

Recommendation 20: Regional coordinators should verify that steps were taken by mines to ensure that method detection limits (MDLs) recommended in the Guidance for the Sampling and Analysis of Metal Mining Effluents are attained when reporting results. If recommended MDLs are not attained consistently in the future, an amendment to the MMER to include mandatory MDLs should be considered in the next review of the EEM program.

5.3 Sublethal Toxicity

5.3.1 Review of sublethal toxicity data

The Sublethal Toxicity Subgroup was formed and given the mandate to:

- present the Review Team with recommendations regarding the review of sublethal toxicity data as part of the national EEM review; and
- base the review on the uses of sublethal toxicity testing as identified in the Guidance Document.

The development of recommendations for review of the sublethal toxicity data was based on the objectives for sublethal toxicity testing as outlined in Section 7 of the Metal Mining EEM Guidance Document (Environment Canada 2002). The two potential uses of sublethal toxicity testing described in the Guidance Document are to:

- 1) measure changes in effluent quality over time as a result of changes in water treatment, site runoff management, acid rock drainage mitigation measures or mine process changes; and
- 2) estimate the potential for exposure area impacts and to estimate the potential geographic extent of the sublethal response using the results from toxicity tests conducted on mining effluent.

The subgroup quickly concluded that it would not be possible for the Review Team to evaluate “changes in water treatment, site runoff management, acid rock drainage mitigation measures or mine process changes” since the reporting of this information is not required under the MMER. As such, it would not be possible to evaluate whether sublethal toxicity testing could measure changes in effluent quality over time as a result of these specifically mentioned factors. The subgroup did, however, make general recommendations on the use of sublethal toxicity testing in relation to effluent quality.

To identify appropriate methodologies to test the second proposed use of sublethal toxicity data (i.e. to estimate exposure area impacts), a team of consultants, led by Dr. Bruce Kilgour of

Stantec Ltd. was commissioned to conduct a literature review and recommend a scientifically defensible methodology to assess the relationship between sublethal toxicity data and field responses (Stantec Consulting Ltd. 2006) (attached to Appendix D). The report (Stantec Consulting Ltd. 2006) found that:

“There is presently inadequate data to demonstrate that sublethal toxicity tests are predictive of receiving-environment effects. Sublethal toxicity at the end-of-pipe often co-occurs with receiving-environment biological effects, but mainly when effluents dominate the receiving environment. Where effects co-occur, there tend to be inconsistencies in the endpoints that correlate. Most (one exception) studies that evaluated the relationship between sublethal toxicity and receiving-environment effects were observational studies conducted on receiving environments that had historical impacts that confounded the interpretation of correlation.”

The report provides methodologies for analytical approaches with:

“the understanding that any conclusions related to concordance between sublethal toxicity and receiving environment effects could be limited because of the confounding effects of historical contamination.”

The Sublethal Toxicity Subgroup provided to the Metal Mining EEM Review Team a report summarizing their discussions. Eight recommendations, outlined in Appendix D, reflect the consensus of the members. The subgroup did not rank the recommendations in terms of priority/importance. The Metal Mining EEM Review Team recognized that they are all valid but that six of them were dependent on the two key recommendations which focus on the need to clearly articulate what sublethal toxicity would be used for:

- the EEM program should facilitate a discussion on the utility and application of sublethal toxicity tests within EEM and ideally, should involve all stakeholder groups; and
- the EEM program must clearly identify the goals (clear questions and hypotheses) for sublethal toxicity testing and how results will be applied to meet those goals. The application of sublethal toxicity must be clearly described, based on sound science.

With all the information available (the preliminary national assessment of sublethal toxicity data, the Stantec report, and the Sublethal Toxicity Subgroup Report), and the subsequent discussions, members of the Review Team (like the Sublethal Toxicity Subgroup) held a wide range of views regarding the utility and application of sublethal toxicity testing within EEM. However, three areas of consensus were identified by the Review Team:

- 1) Regulatory agencies requesting the submission of information (sublethal toxicity testing or otherwise) must clearly articulate to industry and all other stakeholders how the information will be used.
- 2) National assessment of sublethal toxicity testing against stated objectives within the guidance document could not be accomplished as these stated objectives are not clearly articulated. The ability to assess sublethal toxicity testing against objectives on a site specific and or regional basis was not evaluated.

- 3) The wide range of sublethal tests available can contribute to studies that investigate the cause of effects in the field.

To characterize the range of views within the Review Team regarding the utility and application of sublethal toxicity testing in the EEM program, four options were developed for consideration by Environment Canada.

- 1) Establish a multi-stakeholder group to review how routine, ongoing sublethal toxicity testing can contribute to the objectives of the EEM program. In the interim, sublethal toxicity testing requirements will remain status quo within the MMER.
- 2) Establish a multi-stakeholder group to review how routine, ongoing sublethal toxicity testing can contribute to the objectives of the EEM program. In the interim, sublethal toxicity testing requirements will be removed from the MMER.
- 3) Establish a multi-stakeholder group to review how routine, ongoing sublethal toxicity testing can contribute to the objectives of the EEM program. In the interim, do a smaller number of tests more frequently (to be defined by the group).
- 4) Recognizing the debate has been ongoing for many years, routine, regulated sublethal toxicity testing requirement is removed from the MMER as there is no confidence that a multi-stakeholder group would be able to advance the issue until new science is available.

Each of the options received some level of support from members, therefore no consensus was reached. A wide range of opinions was expressed on the utility of sublethal toxicity testing within the EEM program, the type and the number of tests required, and the frequency of testing. However, it was clear that maintaining the status quo is not an acceptable way forward and that there is a need for change on sublethal toxicity testing requirements within EEM. The Review Team members also recognized that analyses of the current sublethal toxicity testing database (3 years of data) could make valuable contributions to further discussions and encouraged activities (by Environment Canada and/or independent experts) to facilitate and share such analyses.

Recommendation 21: Technical improvements proposed by the Sublethal Toxicity Subgroup to the Regulatory Information Submission System (RISS) should be implemented.

Recommendation 22: Environment Canada should facilitate a broad discussion on the utility and application of sublethal toxicity tests within EEM and then Environment Canada should clearly articulate the goals for sublethal toxicity tests and how results will be used.

Recommendation 23: As the status quo is considered unacceptable, Environment Canada should consider changes to sublethal toxicity testing within the EEM program.

5.3.2 Correlation of sublethal toxicity data with field responses

Under the MMER, Schedule 5, s. 17(g) requires mines to undertake a comparison of results of the fish population, fish tissue, and benthic invertebrate community surveys and the results of the sublethal toxicity testing to determine if there is a correlation. It was noted that consultants have

raised concerns regarding this section since the word “correlation” implies that an analysis be conducted that is highly debatable, with methodologies that are not clearly identified. It was agreed by the Review Team that the word “correlation” should be removed from this section.

Recommendation 24: Environment Canada should consider an amendment to modify the wording of Schedule 5, s. 17(g) of the MMER to remove the requirement to determine if there is a correlation between sublethal toxicity data and fish population, fish tissue, and benthic invertebrate community survey results.

5.4 Cross Cutting Science Issues

5.4.1 Alternatives - review of mesocosms in EEM

There may be situations where field surveys may not be appropriate due to site specific reasons, such as confounded receiving environments and unsafe sampling conditions. If standard EEM field surveys cannot be conducted, use of scientifically defensible, cost-effective, and technically feasible alternative monitoring methods can be considered. Artificial stream systems or mesocosms are one of the recommended alternatives for the EEM programs (Environment Canada 2002).

An example of the use of a mesocosm as an alternative monitoring approach by three mines in a joint study, which were confounded with both historical contamination and an upstream municipal wastewater discharge, was presented to the Review Team. In this joint study, a 21-day fathead minnow test was used to examine the impact of water-only versus water plus food-borne (trophic transfer) exposure. The study results showed no significant effect in the EEM endpoints, but significant decreases to other reproductive endpoints such as egg production and spawning events were observed. However, results from the trophic transfer exposure, which is more representative of field conditions, showed that there was a significant increase in egg production and spawning events. Discussion of these results by the Review Team concluded that revisions to the guidance document were needed to include exposure from trophic transfer since the current guidance only considers water exposure.

Recommendation 25: The mesocosm guidance should be revised to include trophic transfer methodologies (i.e., exposure from food sources as well as water). The EEM program should continue to track the use of mesocosms and science evolution as information becomes available.

5.4.2 Guidance for investigation of cause

Once a mine-related effect is confirmed and the extent and magnitude of the effect is known, the mine can then proceed to studies to investigate the cause of the effect. Guidance for investigation of cause (IOC) specific to mines was identified as an issue early in the review process. While IOC guidance is currently available for the Pulp and Paper EEM Program, it is not applicable to metal mining. A survey of regions, which was undertaken to determine how many mines would be initiating IOC for the next EEM phase, found that very few mines, if any, would be heading into IOC at this time. For this reason, it was recommended that any mines needing assistance for

designing IOC studies should inform their regional coordinator(s), who, in turn, could request assistance as needed in designing an appropriate study from the Science Committee, facilitated by the National EEM Office.

Recommendation 26: Regional EEM Offices, with the aid of the National EEM Office, should facilitate as required assistance from the Science Committee in designing site specific investigation of cause (IOC) studies. The National EEM Office should track IOC issues, resulting study designs, and associated guidance to better assist mines in future IOC.

5.4.3 Confounding factors

Confounding factors refer to any anthropogenic, natural or other factors that are not related to the mine effluent under study but may reasonably be expected to contribute to an observed effect (Environment Canada 2002). Common confounding factors, such as historical contamination or upstream discharges from other industry or municipal sources that are encountered in EEM studies are typically addressed through innovative study designs to avoid the influence of confounding factors. The Review Team discussed the need for additional options in the guidance document to deal with confounding factors; however, they acknowledged that confounders vary from site to site. As well, the team was concerned that too much guidance would stifle the creativity that mines have shown in study designs. Published research articles were noted as an additional potential source of information. Further, once common confounders are identified through the regional reviews and national assessment, the Science Committee can expand its research to address identified issues.

Recommendation 27: Relevant literature should be cited in the Guidance Document to provide further information and examples of how study designs can address confounding factors.

Recommendation 28: A text box should be added to the electronic reporting to explain how the information will be used to encourage mines to report confounding factors accurately.

5.4.4 Historical contamination

Some mines discharge into areas that may contain historical contamination. Historical contamination is defined as contamination from previous mining activities or from other non-mining activities. The accumulation in biotic or abiotic environmental compartments due to releases from the current operation is not considered historical contamination, unless that contamination occurred prior to the promulgation of the MMER. The challenge for mines conducting EEM is to distinguish effects related to historical contamination, which is recognized as the most significant confounding factor at some sites. The Review Team identified a range of tools to expand the guidance document to help distinguish historical effects from existing effects (if any) including paleolimnology, sediment transplant, artificial substrates, sediment triads, sublethal toxicity, mesocosms, and before-after study designs to distinguish effects of current discharge from historical contamination. It is noted that the baseline monitoring study design for new or proposed mines could not be required under the MMER as the mine would not yet be subject to the regulation. However, it can be suggested as an option to aid the mine in distinguishing a new discharge from past contamination.

Recommendation 29: To assist mines in designing and conducting EEM studies in areas where there is historical contamination, the Guidance Document should be expanded to describe a range of tools that could be used to distinguish effects due to historical contamination, as well as the benefits and limitations of each tool.

5.4.5 Wetlands

During the early stages of the review, guidance for sampling in wetlands was identified as an issue since wetlands are part of the receiving environment of a number of mines. 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, which are adapted to a wet environment. Wetlands include bogs, fens, marshes, swamps and shallow waters (usually two metres deep or less) (Government of Canada, 1991). For some mines, constructed wetlands are used for water treatment purposes, but these are not considered here. Other mines may discharge into a water body that, in turn, flows into a natural wetland. For the purposes of EEM studies, the Environment Canada policy for these mines was to sample downstream of wetlands. In consideration of developing guidance for sampling in wetlands, concerns were raised that sampling to conduct EEM studies could cause stress to wetlands. In addition, the team noted that EEM study design and monitoring practices must be consistent with the Federal Policy on Wetland Conservation (Government of Canada 1991), particularly with regard to the goals of maintenance of the functions of wetlands and the no net loss of wetland functions on federal lands. Through discussions with regional coordinators, the team found that all existing mines under the MMER that currently discharge into wetlands were able to design their EEM studies around or downstream of the wetland. This issue can be re-evaluated once effects are confirmed in the next phase of monitoring for existing mines (or at sites for new mines coming under the MMER), and the number of mines proceeding to extent and magnitude monitoring studies that may encounter wetlands is known.

Recommendation 30: The Guidance Document should be revised to indicate that the use of 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.

Recommendation 31: After the next monitoring phase, the issue of designing EEM studies in wetlands and the need for specific guidance should be reassessed.

5.4.6 Reference areas

Although the issue of finding suitable reference sites was raised as an issue early in the review process, it was determined, through consultation with the regions, to be a problem at only a few sites. Therefore, the Review Team concluded that the existing guidance on selection of suitable reference areas and issues as they arise could continue to be dealt with in the study design phase on a site-specific basis through consultation with the regional EEM coordinators or TAPs.

5.4.7 Establishing effect sizes

Various approaches for defining critical effect sizes or effect sizes of concern for the Metal Mining EEM Program for fish and benthic invertebrates were presented. In the Metal Mining EEM Program, critical effect sizes will be used to determine when a mine should undertake further monitoring to assess the spatial extent and magnitude of measured effects. It was pointed out that statistical difference alone should not be used to make decisions regarding the need for more extensive monitoring, due to its sensitivity to study design and sample size. Some approaches that have previously been used for the development of critical effect sizes include:

- using ± 2 standard deviations (SD),
- examining natural variability,
- “noise” due to data variability among reference sites,
- incidence of rare responses, and
- differences beyond a fixed number (a threshold that would trigger a regulatory response, a published number that defines an impact at the population or community level, or best professional judgment).

For the Pulp and Paper EEM program, these various methodologies have supported the use of about ± 20 -30% for relative gonad and liver weight and $\pm 10\%$ for condition for fish and ± 2 SD for benthic invertebrates as the critical effect sizes.

It was suggested that development of critical effect sizes be postponed until the next phase since some of the approaches proposed for their development are based on the data distribution (i.e. need confirmation and additional data from the subsequent monitoring phases). It was agreed by the Review Team that critical effect sizes for the Metal Mining EEM Program are not needed at this stage since the next phase of monitoring will be used to confirm if effects are consistent between phases; such confirmation helps to increase confidence that those effects (or absence of effects) are real and not just an exception. Critical effect sizes, however, will be needed at the end of the next phase to determine which mines should conduct studies to determine the magnitude and extent of effects.

Recommendation 32: Critical effect sizes should be developed following the next phase of EEM.

5.4.8 Ensuring scientific evolution

The work of the Science Committee and its relationship to the EEM program was discussed. It was acknowledged that, although the work of the Science Committee is communicated at various scientific meetings, such as the Aquatic Toxicity Workshop, these meetings are not always attended by all stakeholders in the program. It was proposed that advances in science of relevance to EEM should be made better accessible to stakeholders and that the National EEM Office act as a “clearinghouse” for EEM-related research and other information.

In addition, it was noted that there was a need to keep the lines of communication open with all stakeholders to discuss scientific and other issues as they arise. It was proposed that the Review Team continue to meet to discuss issues as they arise 1-2 times per year.

Recommendation 33: The work of the Science Committee should be better communicated to all stakeholders including the listing of related publications for the past 10 years, which should be available on the EEM Web site.

Recommendation 34: Communications between all stakeholders involved in EEM should continue beyond the mandate of the Metal Mining EEM Review Team. This should include annual stakeholder meetings or workshops to get an update and discuss new science and other issues as they arise. Periodic teleconferences of key stakeholders could also take place to ensure communications and coordination.

5.5 Fish Survey

A subgroup of fish experts was established to discuss the fish issues that were identified early in the Metal Mining EEM Program Review and proposed solutions to address these issues. The fish issues identified by the Review Team included:

- fish ageing techniques;
- non-lethal vs lethal surveys;
- guidance for small bodied fish;
- interpretations of significant interactions;
- analysis carried out with fish caught by different methods; and
- absence of fish species.

5.5.1 Ageing techniques for small bodied fish

Some concern was raised about the difficulty in ageing some small bodied fish used in the Metal Mining EEM Program. In the review of the interpretive reports from the first phase, it was found that ageing fish was not as extensive a problem as was initially thought. Five mines could not determine the age of brook stickleback and one mine could not age the threespine stickleback because the annuli were not clearly visible. The experts on the fish subgroup agreed that stickleback is difficult to age and would be more suitable for a non-lethal design where a size (length) distribution could be used to determine the various age classes in a population. It was also suggested that additional guidance could be developed to clarify which structures/techniques should be used to age different fish species as well as including appropriate references in the guidance document.

Recommendation 35: The Guidance Document should be revised to include references for ageing techniques and structures to use, as well as to indicate that fish difficult to age, such as stickleback, should be avoided in lethal surveys.

5.5.2 Non-lethal vs lethal surveys

The Metal Mining EEM Review Team had requested that non-lethal and lethal study designs be compared to determine whether the required measurement endpoints detected similar types of responses between the two different study designs. There was some discussion about whether there was sufficient data to provide a worthwhile analysis since only eleven mines carried out both types of designs. It was also suggested that these two types of studies could not be compared unless they used the same fish species for both types of studies and only five of the eleven used the same species. As such, it was concluded that insufficient data were available to do the comparison to see if EEM endpoints detected similar responses.

Further to this topic, it was proposed that the data about sampling success, species sampled, and sampling period for non-lethal and lethal studies be reviewed by the fish experts and recommendations made for improvements to the guidance on non-lethal fish studies.

Recommendation 36: Guidance on non-lethal sampling should be revised as needed following the review of the non-lethal studies by the Science Committee as well as additional information gained in future phases of monitoring.

5.5.3 Guidance on small bodied fish

The Metal Mining EEM Review Team had suggested that the guidance on small bodied fish needed to be improved. The fish subgroup indicated that science on using small bodied fish species in surveys is evolving constantly. A number of studies are being conducted on small bodied fish and improved guidance should be achieved over time.

Recommendation 37: The guidance on small bodied fish sampling (or study design) should be updated with new references as studies are published in the peer-reviewed literature.

5.5.4 Interpretation of significant interactions

A significant interaction can occur in the ANCOVA (analysis of covariance) statistical analyses used for a number of the fish endpoint site comparisons. It occurs when the relationship between, for example, body weight and gonad weight, is different between the exposed and reference fish (i.e., the slopes are not parallel). When this occurs, the statistical analysis stops and, currently, no effect size is estimated. A significant interaction represents a statistically significant difference and has been defined by the program as an effect. In some cases, a significant interaction could indicate a more serious effect than endpoints that show statistically significant differences but have no significant interaction.

At the present time, experts from the EEM Science Committee are developing a statistical tool to determine an effect size of concern when there is a significant interaction in these endpoints. This tool will calculate the maximum magnitude of difference between the two regression lines and will use this as the magnitude of effect, which will allow endpoints that have interactions to be more fully analyzed and incorporated into the national assessments of the data. The new statistical tool is being developed and, once finalized, the implications for the Metal Mining EEM Program should be assessed. The National EEM Office is currently examining the magnitudes,

pattern and direction of the interactions found results of the first phase of the metal mining EEM program, with the goal of verifying the applicability of this new approach.

5.5.5 Analysis carried out with fish caught through different methods

During the analysis of the fish data for the National Assessment, it was found that there were incidences when fish caught through different techniques were pooled for analysis or fish were sampled using different techniques in the reference and exposure areas. Some fish capture methods are biased towards different sizes of fish. When multiple fishing techniques are combined it could introduce a size discrepancy, which could result in data that is of limited value.

This issue was raised with the Science Committee who advised that unless a mine could show that the fish were of equal size and variance, then pooling of fish caught by different methods should be avoided. The fish subgroup indicated that the incidence of pooling of fish caught by different methods should improve in the next round of sampling due to feedback from the regions and improved sampling proficiency at the mine sites.

Recommendation 38: The Guidance Document should clarify that the comparison of data from mixed fishing techniques should be avoided, where possible. If multiple fish techniques are used then the same techniques should be used between exposure and reference areas and there must not be any size discrepancy introduced into the data set.

5.5.6 Absence of fish

An issue was brought forward that there were poor catch numbers for the fish survey and some difficulty finding suitable reference areas. The fish subgroup discussed whether poor fishing success indicated an absence of fish or whether it was the result of poor site investigation leading to poor sampling design. In discussions with the Science Committee, it was noted that a similar problem had occurred for pulp and paper but the catch numbers improved in subsequent sampling as the consultants became more familiar with the study sites and capture techniques. In addition, there was some concern about how mines move downstream if they are not successful catching fish in an immediate downstream site. The fish subgroup's preference is that mines gradually move downstream until a sufficient number of fish is captured. The reference area should then be selected based on this downstream area where fishing was successful. The current guidance for fish sampling is adequate as mines and their consultants should continue to fish downstream until adequate numbers of fish are caught.

5.6 Fish Tissue Analysis (Use of Fisheries Resources)

Fish tissue analysis is an assessment of whether or not mining effluents have affected the use of fisheries resources. For the metal mining EEM program, concentrations of mercury in fish tissue are used as an indicator of fish usability, since elevated levels of mercury in fish tissue can make fish unsafe for human consumption.

There were two issues identified with respect to the fish tissue analysis: the fish consumption guideline and the detection limit used for total mercury in effluent characterization studies.

5.6.1 Fish consumption guideline

The MMER defines an effect on fish usability as measurements of total mercury in fish tissue that exceed 0.45 µg/g wet weight in fish tissue taken in an exposure area and that are statistically different from the measurements of total mercury in fish tissue taken in a reference area.

The existing Health Canada guideline for total mercury in fish tissue to protect human consumers is currently 0.5 µg/g. To provide national consistency among federal departments, this guideline is proposed as a more appropriate definition of an effect for mercury in fish tissue.

Recommendation 39: An amendment to the MMER is proposed to define an effect on the usability of fisheries resources consistent with the Health Canada fish consumption guideline. In addition, it is proposed that this definition be further modified to indicate that exposure area concentrations should be statistically significantly greater than the reference area concentrations.

Recommendation 40: The Guidance Document should be modified following the amendment to the MMER to indicate the new Health Canada fish consumption guideline as well as provide statistical guidance for conducting a one-tail test to determine if the concentration in the exposure area is statistically significantly greater than the reference area.

5.6.2 Appropriate detection level for mercury in effluent

The MMER requires a survey of mercury in fish tissue when mercury is present in the effluent at concentrations equal to or exceeding 0.1 µg/L. The Guidance Document for Sampling and Analysis currently recommends this same concentration (0.1 µg/L) as the method detection limit (MDL) for mercury in effluent (Environment Canada 2001). In order to detect the concentration with confidence, the MDL should be an order of magnitude (ten times) lower than the concentration specified for mercury in the MMER of 0.1 µg/L. An informal survey of private Canadian laboratories revealed that several laboratories offered method detection limits of 0.01 µg/L and the price per sample was dependent on the volume of each sample and the total number of samples with no evident correlation with detection limit. Two options were considered to address this concern: raise the mercury concentration specified in the MMER, or lower the recommended detection limit. It was agreed that the only option was to lower the method detection limit for mercury.

Recommendation 41: Assuming that commercial laboratories can meet this recommendation, the detection limit for mercury in effluent in the Guidance Document for Sampling and Analysis and the EEM Guidance Document should be revised to 0.01 µg/L in order to detect with confidence concentrations exceeding 0.1 µg/L.

5.7 Benthic Invertebrate Survey

The benthic invertebrate community survey is used to assess changes in fish habitat. This survey is a standard and broadly accepted monitoring approach in the scientific community. There were only two issues discussed by the Review Team concerning the benthic invertebrate community survey; the statistical approach used for data analyses and the Bray-Curtis index.

5.7.1 Statistical approach – univariate vs multivariate

An issue was raised with respect to the statistics that are used to analyze the data collected from the benthic invertebrate community survey. A few members of the Review Team suggested that multivariate statistics could be used at some mines to analyze the benthic invertebrate community data. The current guidance recommends that, at a minimum, univariate analysis be undertaken to provide information on the four core endpoints (and bivariate correlation for gradient designs). This entails ANOVA comparisons of exposure to reference areas for control/impact designs, and regression analyses for gradient designs. The guidance recommends that multivariate analysis may be undertaken on a site-specific basis by the mine to aid in interpretation of the data, but the results of site-specific multivariate analysis are not summarized at the national level. Multivariate analyses could include multidimensional scaling analyses, principle component analyses, and/or a number of other potential multivariate approaches. The specific multivariate approach chosen would depend on the site-specific goals of the analyses. It was agreed that the guidance is sufficient in this area and no further changes to the program are currently needed.

5.7.2 Bray-Curtis index

The Bray-Curtis index is one of the four endpoints for the benthic invertebrate survey, which summarizes the overall community differences between the reference and the exposed areas. The national analysis of the benthic survey results from the phase 1 of the Metal Mining EEM Program indicated that a large number of mines showed a high response in the Bray-Curtis index. Concern was expressed that, for this reason, this endpoint may be difficult to interpret. To address this concern, a research team (Joseph Culp, Nancy Glozier, Alan Willsie, and Rick Lowell) was established to examine the sensitivity of the Bray-Curtis index using two approaches. The first approach compares the Bray-Curtis results to univariate indices for benthic taxa that are sensitive to pollution (e.g., Ephemeroptera, Plecoptera, Trichoptera, Chironomids) that have been widely used in benthic studies. The comparison will consider possible discriminating factors such as mine type and habitat. The second complementary approach uses multivariate ordination techniques to determine which benthic families are responsible for the difference in benthic community structure. These multivariate results will be compared to Bray-Curtis values, including comparisons of statistical significance levels between the multivariate analyses and the Bray-Curtis comparison outlined in the Guidance Document (Environment Canada 2002) considering false positives and false negatives. The work of this team is ongoing.

In addition to examining the sensitivity of the Bray-Curtis index, Review Team members also expressed that the way the Bray-Curtis index is calculated should be reviewed by experts from the Science Committee.

Recommendation 42: The calculation method for the Bray-Curtis index should be reviewed by experts from the Science Committee, based on advances reported in peer-reviewed scientific literature.

6 Conclusion

The Metal Mining EEM Review Team provided a successful forum to review the first phase of the Metal Mining EEM Program. Although it was recognized that the Metal Mining EEM Program is working well, numerous issues were brought forward for discussion to improve the overall effectiveness of the program. The Review Team was able to reach consensus on nearly all of the issues discussed resulting in recommendations presented in this report for improvements to the guidance documents, amendments to the MMER, changes to the electronic reporting, and improvements to the EEM program implementation. Where opinions diverged, options are provided for consideration by Environment Canada in finding ways to improve the program.

The Review Team recommended that this process continue, to keep the lines of communication open with stakeholders and periodically provide an opportunity to improve the EEM program based on new science, new ideas and an opportunity for stakeholders to bring forward their concerns in a solution-oriented process. It is foreseeable to have other review processes in place in the future, for example in five years, to ensure that issues that arise during subsequent phases of the program are addressed.

7 References

- Besser, J.M., T.J. Canfield, and T.W. LaPoint. 1993. Bioaccumulation of organic and inorganic selenium in a laboratory food chain. *Environ. Toxicol. Chem.* 12:57-72.
- Besser, J.M., J.N. Huckins, E.E. Little, and T.W. LaPoint. 1989. Distribution and bioaccumulation of selenium in aquatic microcosms. *Environ. Pollut.* 62:1-12.
- Department of Fisheries and Oceans (2002) *Metal Mining Effluent Regulations*, Part II, *Canada Gazette*, Queen's Printer for Canada, (SOR/2002-222) June 19, 2002.
- Department of Fisheries and Oceans (2006) *Regulations Amending the Metal Mining Effluent Regulations*, Vol. 140, No. 21 *Canada Gazette*, Queen's Printer for Canada, (SOR/2006-239) October 3, 2006.
- Environment Canada 2001. Guidance document for the sampling and analysis of metal mining effluents. EPS 2/MM/5-April 2001.
- Environment Canada 2002. Metal Mining Guidance Documents for Aquatic Environmental Effects Monitoring. Environment Canada, National EEM Office, Gatineau, QC., Canada.
- Gibbons WN, Munkittrick KR. 1994. A Sentinel monitoring Framework for Identifying Fish Population Responses to Industrial Discharges. *J. Aquat. Ecosyst. Health* 3:327-237.
- Glozier, N.E., J.M. Culp, T.B. Reynoldson, R.C. Bailey, R.B. Lowell and L. Trudel. 2002. Assessing Metal Mine Effects using Benthic Invertebrates for Canada's Environmental Effects Program. *Water Qual. Res. J. Canada.* 37 (1) 251-278.
- Government of Canada. 1991. The federal policy on wetland conservation. Published by authority of the Minister of Environment. Ministry of Supply and Services Canada. Cat. No. CW66-116/1991E.
- Lemly AD. 1993. Guidelines for evaluating selenium data from aquatic monitoring and assessment studies. *Environ. Monit. Assess.* 28:83-100.
- Lemly, A.D. 1996. Selenium in aquatic organisms. pp. 427–445. *In:* W.N. Beyer, G.H. Heinz & A.W. Redmon-Norwood (ed.) *Environmental Contaminants in Wildlife*. CRC Press, Boca Raton, FL, U.S.A.
- Munkittrick KR. 1992. A Review and Evaluation of Study Design Considerations for Site-specificity in Assessing the Health of Fish Populations. *J. Aquat. Ecosyst. Health* 1:283-292.
- Munkittrick KR, Dixon DG. 1989a. An Holistic Approach to Ecosystem Health Assessment Using Fish Population Characteristics. *Hydrobiologia* 188/189:122-135.

Munkittrick KR, Dixon DG. 1989b. Use of White Sucker (*Catostomus commersoni*) Populations to Assess the Health of Aquatic Ecosystems Exposed to Low-level Contaminant Stress. Can. J. Fish. Aquat. Sci. 46:1455-1462.

Munkittrick KR, McMaster ME, Van Der Kraak G, Portt C, Gibbons WN, Farwell A, Gray M. 2000. Development of Methods for Effects-driven Cumulative Effects Assessment Using Fish Populations: Moose River Project. SETAC Technical Publication Series, SETAC Press, Pensacola, Fla.

Skorupa, J.P. 1998. Selenium poisoning of fish and wildlife in nature: lessons from twelve real-world experiences. In W.T. Frankenberger, Jr., and R.A. Engberg (eds.). Environmental Chemistry of Selenium. Marcel Dekker, New York. pp. 315-354.

Stantec Consulting Ltd. 2006. Using sublethal toxicity tests to predict receiving-environment effects. File No. 160940051. Prepared for Natural Resources Canada.

Subgroup on Sublethal Toxicity Testing (McGeer, G., Lowell, R., Rosaasen, A., Scroggins, R., Taylor, L., Vigneault, B.) 2006. Recommendations regarding the review of sublethal toxicity test data. Final Report submitted to Metal Mining Environmental Effects Monitoring Review Team.

Appendix A: Metal Mining EEM Review Team Terms of Reference

Purpose

The purpose of the multi-stakeholder Metal Mining EEM Review Team is to review the experiences and results of the EEM program from the first round of EEM studies by Canadian metal mines and, based on this review, to provide recommendations to Environment Canada for improving the program.

Background

The 2002 *Metal Mining Effluent Regulations* (MMER) under the *Fisheries Act* set out stringent limits for effluent quality from Canadian metal mines. Because the Canadian mining industry operates in many different types of ore deposits, uses a variety of mining and milling processes, and discharges effluents to a variety of receiving environments, the uniform effluent limits alone may not ensure adequate protection of all receiving environment. Thus, all mines subject to the MMER are required to conduct an Environmental Effects Monitoring (EEM) program to evaluate the effects of metal mine effluents on fish, fish habitat and the use of fisheries resources. EEM is a scientific monitoring approach that can be used to help determine the effects in aquatic ecosystems caused by metal mine effluent and the effectiveness of environmental protection measures.

Development of the EEM program for metal mines was based on recommendations from a series of multi-stakeholder groups. When the program was being developed, Environment Canada committed to reviewing the overall effectiveness of the program, approximately three years after its commencement. To meet this commitment, Environment Canada is establishing a group of experts to undertake this program review.

Guiding Premises

The work of the Metal Mining EEM (MMEEM) Review Team will be guided by the following premises:

- the EEM program is not designed to deliver research;
- the EEM program is scientifically defensible and cost-effective; and
- the EEM program delivery will be through regulatory requirements with associated guidance documents.

Tasks

1. Identify questions to be addressed in the national assessment of the first phase of the metal mining EEM data.
2. Review all aspects of the EEM program and identify what worked and where problems were encountered.
3. Identify and assess possible solutions to resolve the issues and problems identified.
4. Review results of the national assessment.

5. Submit recommendations for improving the metal mining EEM program to Environment Canada.

Deliverables

A written report submitted to Environment Canada detailing recommendations to improve the effectiveness, efficiency and scientific/technical defensibility of the metal mining EEM program along with economic considerations.

Anticipated Timelines

The program review will take approximately one year to complete, commencing in December 2005 with an anticipated completion date of December 2006.

Anticipated Meeting Frequency

It is anticipated that teleconferences will be held monthly and there will be approximately four face-to-face meetings (January, May, late September and late November/early December 2006).

Environment Canada Roles and Responsibilities

Environment Canada will consider all recommendations made by the group and will consult with all stakeholders groups in an open and transparent manner regarding any proposed changes to the program. The final authority for decisions regarding revisions to the MMER rests with the Ministers of Environment and Fisheries and Oceans and the Governor in Council.

Environment Canada (and others as appropriate) will conduct all data analysis required for the National Assessment and any other related analysis, as required. Environment Canada will provide secretariat services to facilitate the work plan and to ensure an accurate record of all teleconferences, meetings and decisions made.

Operating Principles

The MMEEM Review Team will strive for consensus and will conduct its affairs in a consultative, team-oriented and timely manner. Where opinions diverge, effort will be made to ensure all opinions are accurately recorded.

The Review Team will be timely and forthwith in the conduct of its affairs. Environment Canada will provide secretariat services to ensure an accurate record of its decision-making.

Expert Working Group Roles and Responsibilities

The role of the MMEEM Review Team is to provide to Environment Canada recommendations to improve the effectiveness, efficiency and scientific defensibility of the EEM program. These recommendations should represent the views of all members of the group. The Review Team has

the authority to establish subgroups as necessary, and to ensure that adequate and appropriate guidance is provided to Environment Canada.

The MMEEM Review Team will have the following responsibilities:

- a) establishing an overall work plan, in cooperation with Environment Canada;
- b) ensuring that the program review activities are coherent and focused;
- c) working with the secretariat to ensure the review is completed on schedule;
- d) managing the Working Group projects by:
 - a. setting terms of reference and mandates, including timelines and budget; and
 - b. reviewing work plans and deliverables.

Membership (Updated from the original list)

Members have been invited to participate based on their expertise and not their affiliation. Other experts may attend by invitation of the MMEEM Review Team.

Chairs

Kathleen Hedley (co-chair)	National EEM Office, Environment Canada (until January 2007)
Deni Gautron	National EEM Office, Environment Canada (co-chair starting January 2007)
Robert Prairie (co-chair)	Xstrata Zinc Canada

Environment Canada

Joseph Culp	Aquatic Ecosystems Impacts Research Division
Rick Lowell	National EEM Office
Charles Dumaesq	Mining and Minerals Section
Chris Doiron	Mining and Minerals Section (corresponding member only)
Lisa Taylor	Environmental Technology Centre
Rick Scroggins	Environmental Technology Centre (alternate)
Mark McMaster	Aquatic Ecosystems Impacts Research Division
Nardia Ali	Environment Canada - Ontario Region
Jenny Ferone	Environment Canada - Prairie and Northern Region (alternate)
Nancy Glozier	Prairie and Northern Region
Sherry Walker	National EEM Office (secretariat)
Michelle Bowerman	National EEM Office (secretariat, until October 2006)
Bonna Ring	National EEM Office (secretariat, October 2006-April 2007)

Natural Resources Canada

Jim McGeer	Natural Resources Canada and Wilfrid Laurier University
Bernard Vigneault	Natural Resources Canada

Canadian Nuclear Safety Commission

Glen Bird	Canadian Nuclear Safety Commission (until August 2006)
Malcolm McKee	Canadian Nuclear Safety Commission (starting August 2006)

Fisheries and Oceans Canada

Vince Palace	Freshwater Institute, Fisheries and Oceans Canada
Nuzrat Khan	Headquarters, Fisheries and Oceans Canada (until June 2006)
Patrice Leblanc	Headquarters, Fisheries and Oceans Canada (corresponding member only)

Industry

Elizabeth Gardiner	Mining Association of Canada
Kevin Himbeault	Cameco
Christine Brereton	CVRD Inco Ltd.
Arden Rosaasen	AREVA Resources Canada Inc.

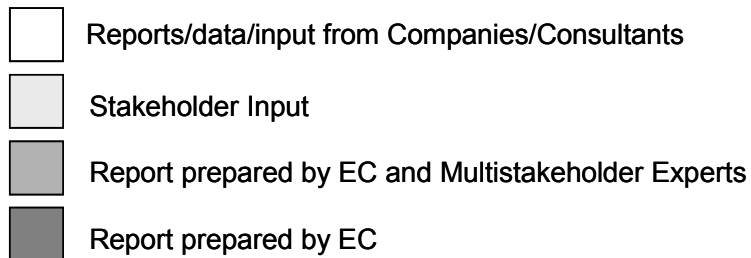
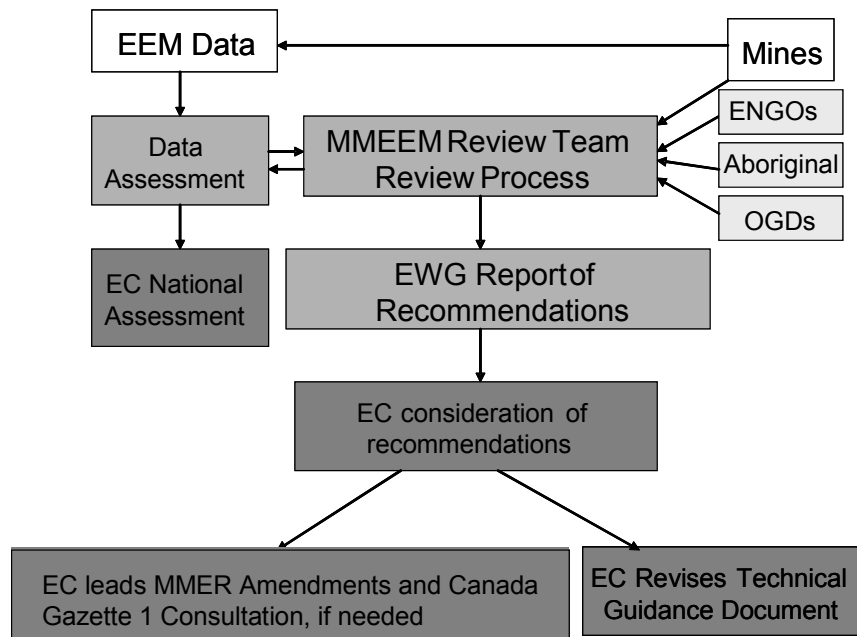
Environmental Non-Government Organizations

Robert Christie	Pictou Harbour Environmental Protection
Judy Parkman	Recycling Organization Against Rubbish

Aboriginal

Alan Penn	Cree Regional Authority
-----------	-------------------------

Attachment A. General Process Flow Chart for Metal Mining EEM Program Review



Attachment B. Key Questions and Preliminary Workplan

Key Questions:

- What worked and what were the main problems encountered?
- What are the main recommendations for improvements?
- Are the data collected valuable for assessing effects of mining effluent?
- What are the main preliminary conclusions of the first phase of monitoring with respect to potential effects of mine effluent?
- Other questions?

Topic	Description	Timelines (January 2006 – December 2006) (tentative)
EC National Analysis	Identify key questions (EWG)	January 31 & February 1, 2006
	Data Analysis	Ongoing
	Report Drafting	Fall 2006
	Final Report	Winter 2006 / 2007
Meetings and Teleconferences	Initial Teleconference	December 20, 2005
	Other Teleconferences	As needed
	1 st Face-to-Face Meeting	January 31 & February 1, 2006
	2 nd Face-to-Face Meeting	June 2006
	3 rd Face-to-Face Meeting	September 2006
Implementation	4 th Face-to-Face Meeting	December 2006
	Stakeholder involvement	January 31 & February 1, 2006
	Government-Industry communication	June 2006
	Information Management	January 31 & February 1, 2006
	Alternatives	June – September 2006
	Investigation of Cause / Investigation of Solutions	June – September 2006
	Others	As needed
Effluent Characterization and Water Quality Survey	Monitoring and Analytical Issues	February – March 2006
	Relevance of Parameters (Additions / Deletions)	February – March 2006
	Results	February – March 2006
Sublethal Toxicity Testing	Analytical Issues	March – May 2006
	Test Relevance (Additions/Deletions)	March – May 2006
	Results	March – May 2006
Study Design (fish and benthos)		June 2006
Fish Survey	Monitoring	July – October 2006
	Data analysis and interpretation	July – October 2006
	Results	July – October 2006
Benthic Invertebrate Community Survey	Monitoring	July – September 2006
	Data Analysis and Interpretation	July – September 2006
	Results	July – September 2006
Report of Recommendations to EC	Outline	September 2006
	1 st Draft	November 2006

	Review of 1 st Draft	November 2006
	2 nd Draft	December 2006
	Review of 2 nd Draft	December 2006
	Final Draft	December 2006
	Report Submission to EC	December 2006

Appendix B: List of Issues

The initial list of issues included:

- Program implementation
 - Stakeholder involvement (subgroup)
 - Government-industry Communication – feedback to mines from the regional offices
 - National consistency
 - Information management: Database and electronic reporting
 - Flexibility in the implementation of guidance
 - EEM and broader environmental assessment/risk management (are there areas of overlap)
- Effluent Characterization and Water Quality Survey
 - Reporting (needs, frequency, format)
 - Relevance of variables (additions/deletions)
 - Detection limits
 - Ra₂₂₆ (water quality monitoring when not present in effluent)
 - Guidance needed to estimate geographic extent when exposure concentration exceeds reference by a factor of 2 or more
- Review of sublethal toxicity (subgroup)
 - Objective review of current state of science concerning utility of sublethal toxicity data compared with field data
 - Discussions on analysis of sublethal toxicity test data as a measure of effluent quality
- Cross-Cutting Science Issues
 - Alternatives
 - mesocosm research and commercial availability
 - Guidance for investigation of cause
 - Confounding factors
 - Wetlands
 - Historical contamination
 - Reference areas
 - Ensuring scientific evolution
- Fish Survey Issues
 - Non-lethal vs lethal surveys
 - More guidance on small bodied fish
 - Fish ageing techniques
 - Presence/absence of fish species
 - Interpretations of significant interactions
- Fish tissue analysis (analyzing the usability of fish)
 - Appropriate detection limit in consideration of trigger level for mercury
 - Update fish consumption guideline
- Benthic Invertebrate Community Survey
 - Univariate vs multivariate statistical techniques
 - Techniques for benthos sampling in wetlands

Appendix C: Barriers to Public Involvement in EEM

The purpose of this document is to briefly summarize the barriers to effective public involvement in the EEM program. Barriers are described from the perspective of industry, the public and regulatory agencies.

This document is intended be a starting point for discussions of how to make public involvement in EEM more effective. It is intended to be a discussion document for the Metal Mining EEM Review Team as a whole, and in particular the Stakeholder Involvement Subgroup.

Barriers for Industry

- EEM program is new, and companies and their consultants are still “learning the ropes”.
- Many companies may lack the expertise needed to develop and implement public involvement mechanisms, and the association communications strategy.
- Many companies have their EEM work completed by consultants, yet the consultants lack the expertise to become involved in public involvement activities, and it is more appropriate for companies themselves to lead such activities. Thus, in this relationship between company and consultant, any opportunity for public involvement falls through the cracks.
- In many cases, company environmental staff is stretched thin, and may not feel that they have the time or resources to devote to public involvement.
- Some may not be aware of the range of activities that could be used to involve the public, as described in the EEM Guidance Document. Thus, they may think of consultations or public meetings to involve the public, and may reject these options for various reasons, without considering other mechanisms that may be quite appropriate, particularly as a first step in public involvement.
- EEM is a highly technical subject, and some companies may regard it as too technical to be discussed with the public, and may feel that it is too difficult to communicate the design and results of EEM studies.
- Some companies may feel that it is pre-mature to involve the public in EEM at this time.
- Some companies may be fearful of the public response to EEM results. They may feel that, with one data point only, the public may jump to conclusions.
- Some companies may regard public involvement as being of little or no value, and not worth the “hassle” association with such involvement.
- Because public involvement is not a legal requirement, some companies may not think it important.
- Because companies may be frustrated by all the regulations with which they have to comply, they may not be aware that they can actually accomplish a lot on a voluntary basis.
- Some companies may feel that information made available to the public will give them an unfair disadvantage over competition and may lead to additional environmental costs necessary to stay in business.

Barriers for the Public

- The EEM program is new, and members of the public may not yet be aware of the program.
- In many cases, concerned members of the public are also involved in other consultations that may lead to what has been termed “consultation fatigue”. As a result, they may not have the time or resources to devote to public involvement related to EEM.
- EEM is a highly technical subject, and some members of the public may be put off by this technical jargon, particularly if companies and their consultants do not attempt to provide material written in a more accessible manner.
- Language barriers may exist in some communities. For example, in communities where many people speak aboriginal languages, and some, particularly elders, may not be fluent in English or French, some materials may need to be provided in the language of the community, and translation services may need to be provided at any meetings in the community to ensure that all can participate. Similarly, in some cases, company management may require that reports be prepared on one of Canada’s official languages, while those in the communities may speak the other language. Thus, some materials from the company may need to be translated, and translation services may be required at any public meetings.
- Poor communication relating to opportunities for public involvement or a failure to provide information to the public in a timely manner may be perceived or interpreted by the public as misleading or demonstrating a lack of transparency, contributing to a lack of trust.
- Some members of the public may not be interested in the narrow scope of issues addressed by EEM, or may be frustrated by this scope. Their fundamental concerns about mines may not be addressed through the components of the EEM program. Communities are often most interested in what they can see, hear and smell — tangible effects. They may not understand what EEM information will really mean for them.
- Ineffective efforts by companies to initiate public involvement may result in interested members of the public not becoming engaged. In turn, the companies may equate a lack of response with a lack of interest, rather than pointing to the effectiveness of the mechanisms used to engage the public.
- Some members of the public may prefer not to know about the results of EEM. Fearing potential personal impacts of “bad news” such as potential impacts on real estate values and the stigma associated with living near a major polluter, they may prefer not to know about EEM and take a “no news is good news” approach.

Barriers for Regulatory Agencies

- EEM program is new, and staff of regulatory agencies are still “learning the ropes”
- In many cases, staff is stretched thin, and may not feel that they have the time or resources to devote to public involvement.
- Some may not be aware of the range of activities that could be used to involve the public, as described in the EEM Guidance Document. Thus, they may think of consultations or public

meetings to involve the public, and may reject these options for various reasons, without considering other mechanisms that may be quite appropriate, particularly as a first step in public involvement.

- EEM is a highly technical subject, and staff from regulatory agencies may regard it as too technical to be discussed with the public, and may feel that it is too difficult to communicate the design and results of EEM studies.
- Some staff may feel that it is premature to involve the public in EEM at this time.
- It may not occur to Regional Office staff that they should or could be working directly with the public along with the industry.
- Some staff may regard public involvement as being of little or no value, and not worth the “hassle” association with such involvement.

Other Considerations

- Once submitted to Environment Canada, EEM data and interpretive reports are public documents. Thus, they are available to the public upon request. Proactive public involvement can help to facilitate effective communications regarding the results presented in these documents, in the event that members of the public request copies of these documents.
- There should be a recognition by all parties that, just as EEM is complicated and site-specific, so is community engagement.
- Environment Canada could do more to explain the EEM program schedule to industry and the public.

Appendix D: Sublethal Toxicity Subgroup Report

Recommendations Regarding the Review of Sublethal Toxicity Test Data

Final Report:

Submitted to

**Metal Mining Environmental Effects
Monitoring Review Team**

by:

The Subgroup on Sublethal Toxicity Testing

Jim McGeer (chair)

Rick Lowell

Arden Rosaasen

Rick Scroggins

Lisa Taylor

Bernard Vigneault

December 21, 2006

Table of Contents

1. Introduction	2
2. Mandate:	2
3. Subgroup Discussions	2
3a. Discussions on: <i>measuring changes in effluent quality over time as a result of changes in water treatment, site runoff management, acid rock drainage (ARD) mitigation measures or mine process changes.</i>	3
Recommendation 1:	3
Recommendation 2:	4
3b. Discussions on: <i>estimating the potential for exposure area impacts and to estimate the potential geographic extent of the sublethal response using the results from toxicity tests conducted on mining effluent.</i>	5
Recommendation 4:	7
3c. Additional subgroup discussions.	7
Recommendation 5:	7
Recommendation 6:	7
Recommendation 7:	8
Recommendation 8:	9

1. Introduction

The 2002 Metal Mining Effluent Regulations (MMER) under the Fisheries Act set out stringent limits for effluent quality from Canadian metal mines. The MMER includes an Environmental Effects Monitoring (EEM) program to collect information on the impacts of mine effluents on Canadian aquatic ecosystems in order to help determine: (i) effects in aquatic ecosystems caused by industrial effluent; and (ii) the effectiveness of environmental protection measures.

Development of the EEM program for metal mines was developed based on recommendations from a series of multi-stakeholder groups. When the program was developed, Environment Canada committed to review the program, approximately three years after its commencement. To meet this commitment, Environment Canada has established the multi-stakeholder Metal Mining EEM Review Team (MMEEM Review Team). Their purpose is to review the experiences and results of the EEM program from the first round of EEM studies by Canadian metal mines and, based on this review, to provide recommendations to Environment Canada for improving the program. Within the context of this review, a sub-group was formed to present recommendations regarding the review of sublethal toxicity data. This report provides the recommendations from the subgroup along with associated information relevant to sublethal toxicity testing in the context of EEM.

2. Mandate:

The development of recommendations for review of the sublethal toxicity (SLT) data will be based on the objectives of sublethal toxicity testing outlined in Section 7 of the Metal Mining EEM Technical Guidance Document (TGD). The two potential uses of sublethal toxicity testing as given in the TGD are:

1. To estimate the potential for exposure area impacts and to estimate the potential geographic extent of the sublethal response using the results from toxicity tests conducted on mining effluent.
2. To measure changes in effluent quality over time as a result of changes in water treatment, site runoff management, acid rock drainage (ARD) mitigation measures or mine process changes.

3. Subgroup Discussions

Following is a brief summary of several key discussions held by the Subgroup, with resulting recommendations. In some cases, the summary and recommendations represent a compromise, due to differing opinions on some of the following issues and on the mandate of this Subgroup.

3a. Discussions on: *measuring changes in effluent quality over time as a result of changes in water treatment, site runoff management, acid rock drainage (ARD) mitigation measures or mine process changes.*

Initial discussions of the Subgroup quickly reached the conclusion that it would not be possible for the MMEEM Review Team to evaluate “changes in water treatment, site runoff management, acid rock drainage (ARD) mitigation measures or mine process changes” since the reporting of this information is not required under the MMER. Therefore, it would not be possible to determine, if sublethal toxicity data collected under the initial monitoring phase of the EEM program has utility in measuring changes in effluent quality over time as a result of these specifically mentioned factors. This is primarily because information is not collected on a national scale for the evaluation of changes in: water treatment, site runoff management, acid rock drainage mitigation measures, or mine process changes as part of the EEM program. In other words, although the TGD indicates this as a potential use for sublethal toxicity data within the MMEEM Program, data collection under the program does not facilitate such comparisons on a national level. However, this does not mean that sublethal toxicity testing cannot be used for assessing changes in effluent quality over time. As well, it is important to note that information on changes in water treatment, site runoff management, acid rock drainage (ARD) mitigation measures or mine process changes can be available and recorded, for example as part of Environment Canada regional EEM summaries or interpretative reports prepared for individual mines.

Recommendation 1:

1a. The MMEEM Review Team should not bother trying to assess, at a national level, whether sublethal toxicity test data can be used to assess effluent quality changes over time as a result of the specific factors listed in section 7 of the TGD (i.e. “changes in water treatment, site runoff management, acid rock drainage (ARD) mitigation measures or mine process changes”) since this information is not required reporting under the MMER.

1b. The MMEEM Review Team should discuss if there is a need to collect information on treatment, mitigation management and process changes over time at a national level. If they decide there is a need, then the EEM program should develop reporting criteria to facilitate the collection and reporting of this type of information in a nationally consistent manner. If not, then the section 7 of the TGD should be revised to clearly indicate that these are not the only factors that could be considered in measuring changes in effluent quality over time.

Although it is not possible to assess, at a national level, changes in effluent quality related to the specific factors listed in section 7 (i.e. as a result of process/ mitigation/ management/ treatment changes over time), there may be value in trying to correlated changes in effluent chemistry over time with changes in sublethal toxicity outcomes. Establishing linkages between sublethal toxicity testing outcomes and effluent chemistry has the potential to: 1) evaluate changes at a particular site over time, and 2) establish national level trends on linkages between

chemistry and sublethal toxicity data. As well there maybe value in assessing effluent quality in the context of the variability in sublethal test outcomes over time. Further discussion on effluent quality analysis methods is outlined below.

Recommendation 2:

The MMEEM Review Team should establish a project to develop data and interpretation on potential linkages between effluent chemistry and sublethal effluent toxicity. This activity has two potential avenues that might be worth exploring:

- 2a. The relationship between changes in toxicity test outcomes and associated effluent chemistry over time; and
- 2b. The development of data integration methods with the goal of establishing whether there are national level trends in effluent chemistry that can be correlated with sublethal toxicity.

In the context of evaluating changes at a particular site over time, examining national trends in the SLT dataset or trying to establish national level trends on linkages between chemistry and sublethal toxicity, there are a variety of potential approaches that could be applied. Regardless of the method applied it is important to establish the goals or questions that an analysis would address in order to ensure the appropriate methodologies and levels. Examples could include: trends over time for a toxicity endpoint, clarifying variability, improving estimates of bioavailability, integrating data from effluent characterization and others. Data would also have to support the analysis and the question and/or goals.

In general, any approaches designed to improve the relevance of the sublethal toxicity data are useful. The subgroup noted that there are already a number of procedural mechanisms recommended that can be used to improve the relevance of sublethal tests with effluent in relation to the potential for impacts in the downstream environment at a site. These include:

1. using site water as the test dilution water,
2. adjusting the test dilution water to match site specific chemistry,
3. acclimating organisms to the site water conditions prior to test initiation.

Research has shown that these methods can have a significant effect on test endpoints and will improve the interpretation of test results. However, in practice it appears that they are rarely used in spite of being provided as guidance within the EEM TDG. The question arises as to how to use and interpret the sublethal toxicity data if this guidance is not being followed. Additionally, there is also no way to know, within the national database, whether those methods have been used.

Recommendation 3

- 3a. Attempts should be made to encourage mines to improve the usefulness of sublethal test results by following recommended approaches provided in the EEM TGC.
- 3b. In the submission of data to the national database, Operators should be able to indicate the type of control/dilution water used in testing and provide data related to

whether or not upstream site water, lab-simulated site water to better match site water quality or test organisms pre-acclimated to site water quality have been used.

3b. Discussions on: *estimating the potential for exposure area impacts and to estimate the potential geographic extent of the sublethal response using the results from toxicity tests conducted on mining effluent.*

As part of the evaluation of the potential for sublethal toxicity tests to estimate exposure area impacts it was decided that an objective review of the current state of the science was needed. The review summarized discussions in the scientific community regarding the linkages between results of lab-based sublethal toxicity testing on effluent and downstream field biological data.

The review focused on literature related to Metal Mining but because of the limited availability of studies a broader review including other industrial sectors was required. The literature search focused on studies available after 1995 to build on the 1997 Review of Methods for Sublethal Aquatic Toxicity Tests Relevant to the Canadian Metal-Mining Industry (Sprague, 1997; CANMET Report AETE Project 1.2.1). It covered different techniques currently in use to compare laboratory toxicity data and field data on benthos and fish and a few relevant studies prior to 1995 were included. The review addressed whether methodologies exist to effectively assess the utility of sublethal toxicity data in addition to commenting on the strengths and weaknesses of these methodologies. The intention was to focus on studies relevant to the metal-mining sector but this was broadened to include information from several other industrial and municipal sectors.

The review was conducted by a team from Stantec Consultants Ltd under the leadership of Dr. Bruce Kilgour. The report (Kilgour 2006) was produced independently for Natural Resources Canada. A summary of the section of the report presenting recommendations for data analysis is presented below. Note that the text below has been copied verbatim from Kilgour (2006) but also has a few changes in working as agreed upon by the subgroup. It is important to highlight that the subgroup members were in full agreement with the sections of the Stantec report which dealt with discussing and summarizing relevant studies as well as recommendations for data analysis. However, the subgroup disagreed on the content and interpretations contained in the other section of the report.

In the studies that were reviewed, many either did not adequately describe ecological condition, did not collect synoptic samples across a gradient of conditions, or did not have enough synoptic observations with which to test for associations between sublethal toxicity test endpoints and receiving environment assessment endpoints. For those studies with enough data, appropriate statistical tools generally were not used. Where the data were adequate and statistical methods were appropriate studies generally reported mixed results. Most of the studies used observational data from areas that had been previously impacted by effluents or other wastes. Therefore, interpretation of correlations could have been confounded for all of the studies. In spite of this, there were three general conditions when sublethal toxicity tended to co-occur with receiving-environment biological effects.

Final Report of the Subgroup on Sublethal Toxicity Test Data

1. Sublethal toxicity and ecological effects co-occurred more frequently when effluent comprised a relatively large fraction of the total receiving-environment flow volume and less frequently when effluent volumes were a smaller fraction of the receiving-water volume.
2. Sublethal toxicity and ecological effects co-occurred more frequently when sublethal tests were conducted using ambient receiving water.
3. Sublethal toxicity co-occurred more frequently with receiving-environment effects when multiple effluent samples are tested and averaged. In other words a single sublethal test may not a good predictor of potential receiving-environment impairment since one test cannot characterize the degree of variability in effluent quality.

Overall there is not enough data to demonstrate that sublethal toxicity tests are predictive of receiving-environment effects. Sublethal toxicity at the end-of-pipe often co-occurs with receiving-environment biological effects, but mainly when effluents dominate the receiving environment. Where effects co-occur, there tend to be inconsistencies in the endpoints that correlate. Most studies examined in this report that evaluated the relationship between sublethal toxicity and receiving-environment effects were observational studies conducted on receiving environments that had historical impacts that could have confounded the interpretation of correlation. There are undoubtedly some examples of environments that are not confounded by historical contamination and these would simplify the process of determining a more valid assessment. The data analytical approaches described below could be used to analyze the existing data available through the metal mining EEM program, though with the understanding that any conclusions related to concordance between sublethal and receiving-environment effects may be limited because of the confounding effects of historical contamination.

The multivariate nature of sublethal toxicity testing data within the federal metal-mining EEM program, and the multivariate nature of the ecological responses would benefit from consideration of multivariate techniques for assessing correlations (other statistical approaches could also be appropriate after careful consideration). A canonical correlation would be an appropriate approach for evaluating both the strength and nature of association between field and sublethal toxicity tests. Alternatively, a multivariate discriminant analysis could be used if the benthic effect sizes could be used to categorize effects as unimpaired (e.g., effects < 2 SDs) and impaired (effects > 2SDs). The resulting discriminant model is the categorical analogue to canonical correlation used for continuous variables. Both canonical correlation and discriminant analysis are useful for illustrating which of the toxicological endpoints vary with the benthic community endpoints or impairment classification. One could take the further step of using the discriminant model to predict the impairment class of the benthic community for future cases. The use of parametric canonical correlation and discriminant analysis has assumptions including multivariate normality and homogeneity of variances and covariances. Appropriate transformations of response and predictor variables, and the use of somewhat independent (uncorrelated) predictor variables will ensure more valid multivariate analysis.

Canonical correlation may also be an appropriate approach to evaluate the strength and nature of the toxicology-EEM fish endpoint relationship, while a discriminant model would be useful for assessing classification frequencies (as above for the benthic endpoints).

Recommendation 4:

The data analysis for comparisons of sublethal toxicity data with receiving environment effects should be done using the principles described in detail in the Stantec Report (Kilgour 2006).

3c. Additional subgroup discussions.

The subgroup also spent time discussing the uses and application of sublethal toxicity testing within the EEM program. Opinions within the subgroup differed widely. For example, some felt that sublethal tests should be used as regulatory endpoints in-and-of-themselves within the EEM program with “trigger” criteria for increased/decreased field studies. Others questioned the validity of the assumption that sublethal toxicity tests can be predictive of environmental effects. Some felt that SLT results can be a first indication of deteriorating effluent quality from changes at the site and be used to mitigate effluent toxicity to prevent further or new impacts.

All members agreed that toxicity testing can be an important tool in Investigation of Cause (IOC) studies. In these cases the full suite of toxicity test methodologies could be considered and in the case of historical contamination of sediments this might include sediment based tests (see Technical Guidance Document) without limiting to other available tools. The uses and application of sublethal toxicity testing is a key issue for the EEM Program and these need to be clearly delineated.

Recommendation 5:

The EEM Program should facilitate a discussion on the utility and application of sublethal toxicity tests within EEM and ideally, should involve all stakeholder groups.

Consideration of this recommendation highlighted an important and related issue, the lack of well defined goals for sublethal toxicity testing within the EEM Program. Even section 7 of the TGD, which provided the subgroup’s mandate, only lists “potential” uses. This represents a key area for improvement within the EEM Program. An additional feature of discussion on the utility and application of sublethal toxicity testing within EEM (recommendation 5) must also include a discussion on the overall goals for sublethal toxicity in the program. In other words, from the multiple possible uses and applications of sublethal toxicity, the EEM program must clearly define which uses and applications will be targeted. Unambiguously delineating how sublethal toxicity data will be used within the program will help to focus efforts and to ensure that data is collected efficiently and properly so that results can be assessed against expected performance.

Recommendation 6:

The EEM Program must clearly identify the goals (clear questions and hypotheses) for sublethal toxicity testing and how results will be applied to meet those goals. The application of sublethal toxicity must be clearly described, based on sound science. The nature and degree of effort invested in Recommendations 7 and 8 below is dependent on

the EEM program goals that are decided upon.

More importantly, it was apparent that the understanding of how sublethal toxicity testing can be used and applied within metal mining EEM was incomplete.

The discussion within the subgroup clearly illustrated that Environment Canada needs to provide leadership to facilitate a broader multi-stakeholder discussion on the use and application of sublethal toxicity testing within EEM.

The subgroup recognized that any discussions and decisions that might arise from recommendations 5 and 6 will be challenging. It will take some time to develop the understanding and clarity required to reach the endpoint of having well defined uses and applications. In fact the issue of clear goals for sublethal toxicity test results has been unresolved for a while and in some respect things have been stalled in recent years. The fact that goals, uses and applications for sublethal toxicity within EEM are not well articulated is reflective of the state of the science and low level of confidence of some stakeholders. From this perspective, it is premature to expect to make decisions on uses and application without first developing an improved understanding and building confidence based on robust study designs.

One way to tackle this issue is to continue the multi-stakeholder approach used by the Aquatic Effects Technology Evaluation (AETE) Program and the Toxicological Investigations of Mine Effluents (TIME) Network. Both of these programs were considered by many as very successful and excellent consensus-building efforts. This would involve developing a methodical, pragmatic and reasoned approach that begins by breaking the issues down into key questions, then supports efforts to develop answers, provides validation and finally moves towards integrating those activities in support of providing sound solutions and options. The organization structure of AETE and TIME would seem like excellent models to follow in terms of multi-stakeholder involvement as well as being able to pool resources for maximum effectiveness.

Recommendation 7:

A group of experts should be brought together with the goal of providing the science required to address recommendations put forward in the subgroup report. This group should have multi-stakeholder involvement and continue the style and tradition of AETE and TIME for providing solutions.

The subgroup discussions also highlighted some other inadequacies related to the national reporting and storage of EEM sublethal toxicity test data within the Regulatory Information Submission System (RISS). These are related to the ease of data handling and difficulties in extracting and utilizing SLT data from the RISS database. Efforts to interpret SLT data by some of the subgroup members illustrated inconsistencies on how sublethal toxicity testing data is collected, reported and stored. In the end, the data required a very time consuming independent verification involving regional staff and others before analysis could begin. Currently, every line needs to be double checked and validated thus limiting analysis only to the brave and persistent

Final Report of the Subgroup on Sublethal Toxicity Test Data

people who have the flexibility and resources to tackle such a large job. This situation clearly needs to be resolved to allow future analysis to be completed with greater efficiency and far less effort.

Recommendation 8:

Data collection, reporting and storage in the RISS database needs improvements and this is crucial for future discussions on the uses and applications of EEM SLT data. The EEM Program needs to ensure high quality data is collected and organized in the national database and that it fulfills the requirements in terms of uses and application.

Problems encountered in working with a database of information extracted from RISS.

- Data cannot be used as downloaded and requires considerable reformatting;
- Date formats are inconsistent and needs to be standardized;
- Inconsistent reporting of sublethal toxicity endpoints from consulting laboratories;
- The two Lemna test endpoints are measured but there is only a single field for these endpoints so data has been entered inconsistently into RISS or not at all;
- In working with the database there is too much reliance on comment fields to make sense of reported data; and
- There are many situations where the text descriptors for the mine site or the Final Discharge Location differ. These inconsistencies occur within and across years.

References.

Kilgour, B. 2006. Using Sublethal Toxicity Tests to Predict Receiving-Environment Effects. A report produced by Stantec Consulting Ltd. for Natural Resources Canada under contract no. 23440070071001. 33 pgs.

Attachment A



**Using Sublethal Toxicity Tests to
Predict Receiving-Environment
Effects**

File No. 160940051

January 9, 2007

Prepared for:

Natural Resources Canada

Attention: James McGeer

Prepared by:

Stantec Consulting Ltd.

1505 Laperriere Road

Ottawa, Ontario

K1Z 7T1

613-725-5104

Using Sublethal Toxicity Tests to Predict Receiving Environment EffectsJanuary 9, 2007

Executive Summary

This report provides a review of literature documenting the relationship between end-of-pipe sublethal toxicity tests and receiving-environment biological effects. At least 14 relevant research efforts have been carried out in Canada and the United States since 1995 including some major works carried out in Canada in support of the development of the metal mining environmental effects monitoring (EEM) program, and as part of national summaries of the pulp and paper EEM program. There is presently inadequate data to demonstrate that sublethal toxicity tests are predictive of receiving-environment effects. Sublethal toxicity at the end-of-pipe often co-occurs with receiving-environment biological effects, but mainly when effluents dominate the receiving environment. Where effects co-occur, there tend to be inconsistencies in the endpoints that correlate. Most (one exception) studies that evaluated the relationship between sublethal toxicity and receiving-environment effects were observational studies conducted on receiving environments that had historical impacts that confounded the interpretation of correlation. A more valid assessment of the correspondence between sublethal toxicity tests and receiving-environment effects could be accomplished through observational studies if the data were collected locations where there is no historical contamination. The data analytical approaches described below would be suitable, assuming that the data are from sites with no historical contamination. The data analytical approaches described below could also be used to analyze the existing data available through the metal mining EEM program, though with the understanding that any conclusions related to concordance between sublethal and receiving-environment effects will be limited because of the confounding effects of historical contamination.

The concordance between sublethal and biological effects data should be assessed using synoptic samples of sublethal toxicity and receiving-environment biological responses, collected across a gradient of conditions. The conclusion that whole-effluent toxicity tests are predictive (enough to be useful) of receiving-environment ecological effects will then depend on (1) the nature of the association between the toxicity and ecological endpoints, and (2) management's requirement for correct conclusions that the ecological indicator is "impaired". The strength of the association between toxicological and ecological endpoints can be established through correlation analysis, while contingency tables indicate the likelihood of making correct and incorrect conclusions in future assessments. Some understanding of the strength of the correlation, as well as the probability of correct/incorrect classification rates is required in order to fully justify the use of one endpoint to predict another. Multivariate procedures should be used to evaluate the concordance of the two monitoring components because both involve multiple endpoints.

Using Sublethal Toxicity Tests to Predict Receiving Environment Effects

Table of Contents

January 9, 2007

Table of Contents

1.0 Introduction.....	1
2.0 Establishing Concordance Between Indicators.....	3
3.0 Literature Review	6
3.1 Review of Studies and Analyses Pre-1995	6
3.2 Review of Studies and Analyses Post-1995	9
4.0 Summary and Recommendations.....	17
4.1 Summary of findings	17
4.2 Recommendations for data analysis	18
5.0 Opinion on Using Sublethal Tests in EEM.....	21
5.1 Goals of EEM in relation to sublethal toxicity testing	21
5.2 Early warning of potential impacts	21
5.3 longer term perspective on the outcomes of EEM Monitoring	22
5.4 Broader considerations on sublethal toxicity testing and EEM	22
6.0 Literature Cited	24

List of Figures

Figure 1. Conceptual biplot illustrating one approach to estimating the reliability of sublethal toxicity test endpoints for predicting receiving-environment ecological effects. 5

Introduction

Under the Metal Mining Effluent Regulations (MMER), all metal mines in Canada are required to conduct Environmental Effects Monitoring (EEM). The objective of EEM is to evaluate the effects of mine effluents on fish, fish habitat and the use of fisheries resources. This information may ultimately be used to assess the adequacy of the MMER for protecting aquatic environment resources. To meet the objective, all mines in Canada are required to conduct biological monitoring which includes a fish survey (fish population and fish tissue analysis) and a benthic macroinvertebrate survey, as well as effluent and water quality monitoring which includes effluent characterization, receiving water quality monitoring, and effluent sublethal toxicity testing.

The development of the Metal Mining EEM (MMEEM) program was based on recommendations from multiple stakeholders. When the program was developed, Environment Canada committed to reviewing the MMEEM program and recently developed the multi-stakeholder Metal Mining EEM Review Team (MMEEMRT). The purpose of the Review Team is to review the experiences of the first round of MMEEM studies and to provide recommendations to Environment Canada to improve the program. As part of the review, a subgroup was convened to provide recommendations on the review of sublethal toxicity testing data. The subgroup's review will be based on the objectives of sublethal toxicity testing outlines in Section 7 of the MMEEM Technical Guidance Document (TGD). The data available to the review team arise from the requirements for all mines in Canada to conduct four freshwater sublethal toxicity tests (fish, invertebrate, algae, plant) or three marine sublethal toxicity tests (fish, invertebrate, algae), depending on the receiver type. Currently this is done twice per year.

The purpose of sublethal toxicity testing is to answer the question: "Is there evidence from sublethal toxicity testing that mine effluent may affect fish, invertebrates and aquatic plants?" (Environment Canada, 2002). To answer this question, all mines in Canada conduct four freshwater sublethal toxicity tests (fish, invertebrate, algae, plant) or three marine sublethal toxicity tests (fish, invertebrate, algae), depending on the receiver type. Sublethal toxicity tests have the following suggested uses in EEM programs:

1. To estimate the potential for exposure-area impacts and to estimate the potential geographic extent of the sublethal response using the results from toxicity tests conducted on mining effluent.
2. To measure changes in effluent quality over time as a result of changes in water treatment, site runoff management, acid rock drainage (ARD) mitigation measures or mine process changes.

The toxicity subgroup also recognized a third use for sublethal toxicity data:

3. To estimate the relative contributions from various sources (discharges) to observed effects in the receiving environment (Scroggins et al., 2002b; Environment Canada, 2002)

Conceptually, the use of sublethal toxicity testing within MMEEM is to “provide an estimate of the extent of potential effects on biological components of the exposure area, whether these components are being directly measured in the field or not.” (Environment Canada, 2002; Scroggins et al., 2002a). Effluent toxicity tests are informative because they integrate the interactions among a complex suite of substances (effluent) in a highly standardized, controlled environment (Chapman, 2000). The tests are quick, with a turn-around time of about 1 to 2 weeks, and can thus provide relatively inexpensive, “real-time” information on effluent quality. The ability to extrapolate the results of single-species laboratory toxicity tests, however, to effects in complex environments on higher levels of organization has been debated (Cairns, 1983; 1988; Clements and Kiffney, 1994; Amman et al., 1999; Waller et al., 1999; Lapoint and Waller, 2000; Chapman, 2002).

The intent of this report is to provide an objective review of the state of the science of the potential for sublethal toxicity tests to estimate receiving-environment effects. The review specifically focuses on data that explore the relationship between toxicity and effects on fish communities, fish populations and communities of benthic invertebrates. A major aspect of the review is the identification of data analytic tools that have been used to demonstrate relationships between laboratory toxicity tests and field biological studies in studies that have analyzed the results from multiple field programs. Though the intention was to focus on studies relevant to the metal-mining sector, the review was broadened and includes information from several industrial and municipal sectors. In addition, though the initial intention was to focus on studies post 1995 (i.e., post the CANMET review by Sprague, 1997), this review includes a discussion of some of the pivotal pre-1995 papers that were the basis of post 1995 work.

In the sections that follow, the review first discusses “accepted” approaches for assessing the concordance between sets of indicators (**Section 2.0**) to provide the background necessary to critically examine the key papers both pre (**Section 3.1**) and post (**Section 3.2**) 1995. The review was limited to peer-reviewed articles or government reports, and specifically did not include summaries available from the first round of Canadian MMEEM programs. **Section 4.0** summarizes the major findings of the review and provides recommendations for potential future work.. A discussion of the utility of sublethal toxicity tests within an EEM context is provided in **Section 5.0**.

Establishing Concordance Between Indicators

Figure 1 illustrates the conceptual relationship that a toxicity test endpoint might have with a single ecological endpoint. It is this relationship that we analyze if we wish to know whether the toxicological endpoint is predictive of the ecological endpoint. The conclusion that whole-effluent toxicity tests are predictive (enough to be useful) of receiving-environment ecological effects will then depend on (1) the nature of the association between the toxicity and ecological endpoints, and (2) management's requirement for correct conclusions that the ecological indicator is "impaired". We typically have a decision criterion for the toxicity test, which when exceeded is used to declare that the effluent is "impairing" and is often the criterion used to predict a receiving-environment impact. We also typically have a decision criterion for the receiving-environment ecological endpoint, which again when exceeded is used to confirm that an ecological effect has occurred. When comparing a toxicity test endpoint with a receiving environment endpoint, each with their criteria applied, there are four possible outcomes. A "true-positive" (TP) result occurs when both endpoints indicate impairment, whereas a "true-negative" (TN) result occurs when both endpoints indicate no impact. False-positives (FP, toxicity endpoint indicates impairment when there is none) and false-negatives (FN, toxicity indicates no impairment when there really is one) are also very important frequencies to understand. High rates of false-positive and false-negative results occur when the association between the two endpoints is weak, while lower frequencies of these two errors occur when the association between endpoints is strong. The strength of the association between toxicological and ecological endpoints can be established through correlation analysis, while contingency tables indicate the likelihood of making correct and incorrect conclusions in future assessments. Some understanding of the strength of the correlation, as well as the probability of correct/incorrect classification rates is required in order to fully justify the use of one endpoint to predict another. The comparisons that yield higher rates of FP and FN could be very interesting from the point of view of understanding why there is a lack of concordance. Also, in those cases, it may be possible to refine the endpoint interpretation (e.g. cutoff criterion) to improve the strength of the correlation.

Swets (1988) and Murtaugh (1996) provide guidance on assessing contingency tables, and in examining how the decision criterion can be manipulated to alter/adjust correct and incorrect classification frequencies. As an example, Yoder and Rankin (1995) analyzed contingency tables to determine the likelihood of correctly declaring fish communities to be impaired on the basis of samples of benthic communities. Murtaugh (1996) recommends calculating the percent true positives (%TP) as

$$\%TP = \frac{TP}{TP + FN}$$

and the percent false-positives (%FP) as

$$\%FP = \frac{FP}{FP + TN}$$

Ninety-five percent confidence limits can be placed on these frequencies based on equations for the binomial distribution as per Zar (p 378, 1984).

Demonstrating that two endpoints are correlated requires synoptic measures of both endpoints across a gradient of conditions (from acceptable to unacceptable). Further, there must be enough synoptic observations to confidently demonstrate a strong, weak, or no correlation. Any less than about 10 observations would be considered a small data set for this type of analysis, while more than about 20 would provide reasonable statistical power. Synoptic measures should be roughly evenly distributed along the gradient. These conditions would appear to favour national level assessments of MMEEM data.

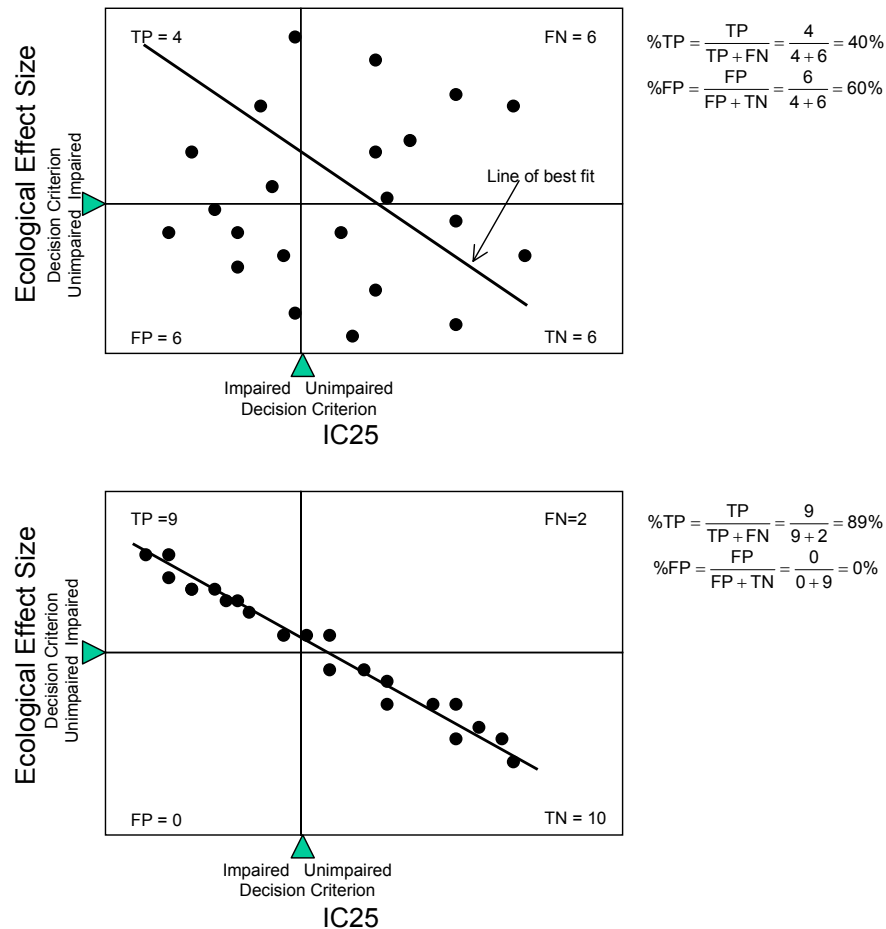


Figure 2. Conceptual biplot illustrating one approach to estimating the reliability of sublethal toxicity test endpoints for predicting receiving-environment ecological effects.

Figure Note: Two examples are given, one with a relatively low predictive ability (top panel) and one with a higher predictive ability (lower panel). Graphs are split into quadrants delineated by the decided criteria for effects (denoted by the green triangles) with TN indicating true negative, TP true positive, FN false negative and FP false positive. The calculation shown on each graph gives an estimate of the frequency (%) of TP and FP (i.e. good predictive ability). The upper panel shows the calculations for percent TP and FP decisions based on the observed scatter of the data, and selected endpoints. The lower panel illustrates the improvement in percent TP and FP outcomes based on a tighter data scatter.

Literature Review

Review of Studies and Analyses Pre-1995

Papers from the Sprague Review

Sprague's (1997) original review of sublethal toxicity tests, in part, attempted to quantify the likelihood of correctly declaring that receiving-environment biological indicators are impaired on the basis of results of an end-of-pipe sublethal test. Sprague concluded that sublethal toxicity tests were generally predictive of receiving-environment effects on the basis that 84% of studies (29 studies from 63 locations) reported an impaired receiving environment when effluent also caused sublethal toxicity. That review, and others since (e.g., Parkhurst, 1996), were based principally on the USEPA's CETTP (Complex Effluent Toxicity Testing Program) studies, Birge et al.'s (1989) South Elkhorn Creek Study (Kentucky), Eagleson et al.'s (1990) study in North Carolina, and Dickson et al.'s (1992) study of the Trinity River (Texas). The eight CETTP studies consisted of work from Scippo Creek, Ohio (Mount and Norberg-King, 1985), Ottawa River, Ohio (Mount et al., 1984), Skeleton Creek, Oklahoma (Norberg-King and Mount, 1986), Five Mile Creek, Alabama (Mount et al., 1985), Ohio River, West Virginia (Mount et al., 1986b), Kanawha River, West Virginia (Mount and Norberg-King, 1986), Naugatuck River, Connecticut (Mount et al., 1986a), and Back River, Baltimore Harbor, Maryland (Mount et al., 1986c).

The CETTP, South Elkhorn and Trinity River studies evaluated the association between sublethal toxicity testing of the receiving water and the condition of receiving-environment indicators (principally fish, benthos), whereas the Eagleson study in North Carolina evaluated the association between end-of-pipe test results and benthos. The original CETTP papers demonstrated concordance between sublethal test results and ecological effects by tabulating the number of times that the two components agreed (i.e., % true-positives, see Figure 1). Agreement was based on arbitrary levels of ecological impact and toxicity. Ecological impacts were classified according to the percent reduction (20-40, 40-60, 60-80, 80-100) in the number of taxa (zooplankton, benthos, or fish), while toxicity to *Ceriodaphnia* reproduction or fathead minnow growth was classified according to the percent increase in toxicity, in this case reductions, (again, 20-40, 40-60, 60-80, 80-100). The level of agreement between toxicity tests and ecological indicators was generally high (80 to 100% true positives) because the study sites were chosen to be in generally in poor condition producing both an impaired status for the receiving environment and effluent sublethal toxicity. Several reviews since have criticized these studies because the study locations were those that were generally known to be in an impaired state and with effluents known to be toxic (e.g., Parkhurst, 1996; Chapman, 2000). The study authors did not evaluate the frequency of false-positive or false-negative results. The CETTP studies also quantified the correlations among all possible pairs of toxicological and receiving-environment indicators. Correlations were generally weak or not significant, in part because of low sample sizes (n~7 or less), and because the sites were generally impacted and did not provide a necessary gradient of conditions. Some of the CETTP reports argued that the

response of the most sensitive toxicity test species should be compared to the most sensitive trophic level in the receiving environment. That is, they might have chosen *Ceriodaphnia* reproduction as the most sensitive toxicity test endpoint and number of benthic taxa as the most sensitive ecological indicator at Station A, while choosing fathead minnow growth and number of fish taxa at Station B, etc. Significant correlations based on the most sensitive endpoints were obtained from Skeleton Creek and Naugatuck River.

Data from the CETTP were re-evaluated by Marcus and McDonald (1992). They first demonstrated, in a simulation experiment, demonstrated that obtaining a stronger correlation for the most sensitive indicators than for the pairs of the individual components could easily have occurred by chance, and they thus somewhat dismissed the conclusions obtained for Skeleton Creek and Naugatuck River. They then demonstrated that the approach used to construct tables of percent correct classifications was misguided and biased towards producing overly optimistic results. They pointed out that qualify as a correct prediction of a “non-impaired” condition, all biological and all toxicological measures would have to have classified as unimpaired. To qualify as a correct prediction of impairment, only a single toxicological endpoint and a single ecological endpoint would have to classify as impaired. Marcus and McDonald demonstrated that those decisions produced a mathematical certainty of the conclusion; i.e., that toxicological endpoints were predictive of ecological endpoints. Unfortunately, Marcus and McDonald did not re-calculate the percent correct predictions. The final analysis conducted by Marcus and McDonald involved canonical correlation to test for associations between the sublethal toxicological endpoints and measured instream ecological endpoints. Canonical correlation is a statistical tool for defining linear relationships (associations) between two sets of data (e.g., toxicity data, ecological data) that both consist of multiple variables. They found statistically significant ($p < 0.05$) correlations for five of 11 (or 45%) of the CETTP data sets. Canonical correlation coefficients were relatively high (> 0.7) for each data set indicating that more than half (50%) of the variation in ecological effects could be “explained” by variation in sublethal toxicity. Small sample sizes, was the major reason that more the canonical correlations were not significant.

Dickson et al. (1992) combined the CETTP study data sets with data from South Eklhorn Creek (Birge et al., 1989) and the Trinity River (Dickson et al., 1989). They used canonical correlation to determine how toxicity endpoints and ecological endpoints covaried within data sets. Canonical correlations were generally high ($R > 0.7$) and statistically significant. Some of the correlations were highly influenced by having clustered data (i.e., many highly impaired and a few non-impaired locations), but not all. The strength of the canonical correlations, and the variables demonstrated to be important to the overall correlation varied over time in the Trinity River. The complicated results led the authors to conclude that one or more of the following caused the correlations: (1) the toxic components varied over time; (2) there were multiple significant pair-wise correlations between toxicological and ecological endpoints; and (3) correlations were due to chance alone. Dickson et al. then classified impairment status based on whether the observed toxicity or ecological condition fell outside a specified range of the data. Toxicity variables were summed, then “normalized” assuming a normal distribution. Counts of fish species and benthic taxa were normalized using a Poisson distribution. Based on the normalized data for both toxicological and ecological endpoints, observed values beyond an upper percentile (e.g., the 90th) were used to classify a site as impaired, while observed values

less than a lower percentile (e.g., the 10th) were used to classify a site as unimpaired. These percentiles were based on all the data from a given study, without *a priori* knowledge as to whether a site was impaired or not. The classification procedure resulted in a two-way contingency table for predicted and observed impacts. A Fisher's exact test was then used to evaluate the significance of the relationship between ambient toxicity and in-stream ecological impact. For their large, overall comparison, the authors determined that ambient toxicity tests correctly predicted a biological effect in almost 90% of cases where a biological effect was measured, but also incorrectly predicted an impact in nearly 60% of cases when there was no ecological impact.

Eagleson et al. (1990) presented an analysis of 43 comparisons of whole-effluent toxicity tests and measured in-stream biological communities. Most of the effluents tested were municipal wastewaters. The authors compared the "chronic value", a percent of effluent above which detrimental effects (reduced survival and reproduction of *Ceriodaphnia dubia*) occur, to the condition of benthic macroinvertebrates. The authors predicted biological impacts would occur only if the chronic toxicity value (EC25) was greater than the permitted in-stream waste concentration. The impairment status of the benthic community was subjectively determined from interpretation of the assemblage, with some consideration for number of taxa and the kinds of taxa present. The authors concluded that the whole-effluent toxicity tests correctly predicted the impairment status of the receiving stream in 88% of the cases, but did not report the strength of the correlation between toxicological and ecological measures, did not demonstrate a gradient in conditions across the study sites, and did not calculate the probability of incorrect classifications.

Papers not in the Sprague Review

Nimmo et al. (1990) used 48-h acute tests with *Ceriodaphnia dubia* to test water from the Chalk River, and acute and chronic tests to test water from the Clark Fork River and Whitewood Creek. In mid May of 1986, 500 fingerling trout were released into the Chalk River. Mortality of those fingerlings reached 80% within 48 h of release. Ambient waters produced significant mortality in *Ceriodaphnia dubia* over a 48-h period. There was no sublethal endpoint measured. The authors also found that ambient waters in the Clark Fork River (and its tributaries) produced variable toxicity included acute lethality and depressed reproduction depending on the location samples were collected. The authors concluded that variations in sublethal toxicity were "consistent with fishery information about depressed populations of catchable trout", presumably higher toxicity occurring in areas with depressed brown trout populations. In their third case study, the authors carried out an acute test with *Ceriodaphnia* above and below a mine effluent discharge on the Whitewood River, South Dakota. Stream water was considered acutely toxic to *Ceriodaphnia*, and that corresponded to an absence of fish and presence of "some" aquatic life. Downstream water was considered acutely toxic to *Ceriodaphnia*, and corresponded to an absence of fish and presence of "some" aquatic life. These three case studies did not present robust analyses of instream biological condition, and therefore it is difficult to assess the "predictiveness" of the sublethal toxicity tests in this case.

Review of Studies and Analyses Post-1995

Metal Contaminated Streams in New Zealand

Hickey and Clements (1998) evaluated the association between ambient toxicity and measures of benthic community composition in three streams in Coromandel Peninsula of New Zealand. Toxicity tests included the chronic reproduction assay with *Daphnia magna*, and a 96-h acute bioassay with an endemic mayfly, *Deleatidium* (Leptophlebiidae). Benthic communities were collected from riffle environments with 250-µm mesh, and identified to genus and species. No formal statistical tests were used to determine if the variations in toxicity tests and measures of benthic community composition from nine locations covaried. The authors concluded, however, that there was generally good agreement between the chronic toxicity test results (i.e., reproduction of *D. magna*) and benthic community composition. Sites with high metals concentrations in water tended to have reduced reproduction, coinciding with lower benthic community taxa richness, a lower number of mayfly taxa, and lower total numbers of mayflies. The acute toxicity test results with the mayfly *Deleatidium* did not vary among stations, and so did not covary with indices of benthic community composition. Based on data from three locations, this study did not have enough data to quantify the strength of the association between toxicity and ecological condition, or the frequency of correctly predicting a biological impairment.

Analysis by Hartwell and others, Chesapeake Bay

Hartwell et al. (1997) examined the associations between measured ambient water-column toxicity and sediment toxicity, to metrics of the fish community in Chesapeake Bay tidal tributaries. Water column bioassays included 7-d sheepshead minnow survival and growth, 7-d grass shrimp survival and growth, a copepod life-cycle survival and reproduction test, a bacterial luminosity bioassay, and bioassays with sago pondweed. Sediment bioassays included the 10-d sheepshead minnow embryo-larval survival and teratogenicity test, two 10-d amphipod survival and growth tests (i.e., two species), the 10-d polychaete survival and growth test and the *Spartina alterniflora* seed germination test. The authors calculated a “risk rank” based on the results of the water-column and sediment bioassays. Risk ranking has five components: (1) severity of effect; (2) degree of response; (3) test variability; (4) site consistency; (5) number of measured endpoints. Severity refers to the endpoint measured and were used as multipliers where mortality was 3, reproduction was given a 2 and growth as equal to 1. The risk score was calculated as:

$$\text{risk score} = \left[\sum \left[(\text{severity})(\% \text{response})(\text{CV}) \right] + (\text{consistency}) \right] / \sqrt{N}$$

where

- consistency = $[(N/2) - X]^3$,
 - where X is the number of non-significant endpoints

- N is the total number of endpoints,
- CV is the coefficient of variation, and

Fish were sampled with beach seines and bottom trawls deployed near mid channel. An Index of Biotic Integrity (IBI) as well as a measure of Diversity were calculated using the fish collection data. The IBI is a summary metric of the fish community that attempts to summarize the tolerance of the community to stress. High scores tend to occur in “healthier” fish communities.

Simple parametric Pearson correlations with a total of eight observations, demonstrated that risks determined from the sediment bioassays were correlated with observed fish community diversity and IBI scores, while risks estimated from water-column bioassays were not.

Analysis by Sarakinos and Rasmussen

Sarakinos and Rasmussen (1998) compared variations in benthic invertebrate community composition and total abundance with response “thresholds” calculated from chronic whole-effluent toxicity tests. The 7-d survival and reproduction bioassay with *Ceriodaphnia dubia*, the 96-h growth inhibition test with *Selenastrum capricornutum*, and the 7-d early-life-stage survival and growth tests with fathead minnows were tested on a single sample of effluent from a paper mill in Kingsley, QC. Results of toxicity tests were reported as the Maximum Allowable Toxicant Concentration (MATC; i.e., the geometric mean of the No Observed Effect Concentration, NOEC, and the Lowest Observed Effect Concentration, LOEC). The average toxicity of the sample was expressed as the geometric mean of the MATC for all assays. The whole effluent toxicity tests produced an average MATC of 3.6% effluent. The authors then collected 210 samples of benthic macroinvertebrates from within the 1% mixing zone of the mill’s effluent in the Nicolet River. The authors first used a Partial Canonical Correlation to relate descriptors of benthic community composition and % mill effluent after controlling for the underlying influences of natural habitat variables (i.e., current velocity). The analyses demonstrated a significant decrease in taxonomic richness at an effluent concentration of 16%, and a reduction in total abundances in locations where effluent concentrations were between 0 and 2%. The authors concluded that the standard whole-effluent test overestimated the effluent concentration that would elicit an ecological response in the receiving environment. This study is based on prediction from a single effluent, and does not represent multiple synoptic samples of toxicity and biology collected across a gradient. The results, therefore, demonstrate co-occurrence of sublethal toxicity and biological effects, but do not demonstrate a correlation. The level of agreement between the effluent concentration predicted to cause effects and the concentration in the receiving environment corresponding to effects on the benthic community were remarkably similar considering problems associated with estimating effluent dilutions in receiving environments (Girling et al., 2004).

USEPA Follow-Up to CETTP Studies

The USEPA (1999) summarized the various analyses provided in the original CETTP studies, and the follow-up analyses by Marcus and McDonald (1992) and Dickson et al. (1992). USEPA (1999) also included a set of comments from an independent biostatistician. Smith (1994, as

cited in USEPA, 1999) commented that the Canonical Correlation analyses may not have been the most appropriate for two reasons. First, Canonical Correlation assumes a linear association among variables, which may not have occurred, and which was not determined. Second, Canonical Correlation would fit “linear combinations of toxicity and instream response variables that correlate maximally”. Smith (1994, as cited in USEPA, 1999) concluded that the most meaningful associations between toxicity and instream responses might not have been represented by the linear combinations of variables that produced the maximum correlation. He further recommended the use of ordination techniques that objectively summarize multivariate sets of data. Smith proposed to ordinate the toxicological data, and the ecological data, then test for associations between the derived variables based on the ordination methods. It is not clear if the work was ever done.

River Goyt, UK

Maltby et al. (2000) conducted a variety of studies on the River Goyt, Derbyshire (UK) in the vicinity of a bleaching plant. A 48-h *Daphnia magna* immobilization test was conducted using whole effluent from the plant, while the same test was conducted in-situ with caged specimens upstream and downstream of the plant discharge. Cages of *Gammarus pulex* were deployed for six days upstream and downstream of the plant discharge to evaluate effects on feeding rate (leaf litter decomposition). Twenty-five samples of benthos were collected using a Hess sampler from all microhabitats in one location upstream and one location downstream of the plant discharge. Receiving-water toxicity and ecological studies were consistent with the whole-effluent tests. Downstream of the discharge there was a reduction in *D. magna* survival, in *G. pulex* survival and feeding rate, in detritus processing, and in biotic indices based on macroinvertebrate community structure. There were no specific tests used (or required) to demonstrate the agreement between whole-effluent tests and receiving-environment ecological condition, because the study evaluated a single point-source discharge, and had only two receiving-environment locations (upstream and downstream).

Analysis by Diamond and Daley

Diamond and Daley (2000) compiled a database of 250 dischargers across the United States and examined the relationships between whole-effluent toxicity test endpoints (*Ceriodaphnia dubia*, *Pimephales promelas*) and instream biological conditions as measured by inventories of benthic macroinvertebrates. Whole effluent toxicity tests were integrated for each facility in two general ways. First, the authors computed the average value for each WET endpoint reported and then compared the average value to pass/fail criteria used by state agencies. Second, the authors calculated the fraction of WET tests failed at each facility. Benthic community data were standardized because of differences, among States, in taxonomy and metrics used to summarize the community. They calculated the average relative difference in community metrics as:

$$\frac{\sum_{i=1}^N [(Upstream_i - Downstream_i)]}{N}$$

where N is the total number of metrics measured at each site. All benthic community assessments used in the study measured several metrics such as number of taxa, number of taxa belonging to the mayfly, stonefly and caddisfly groups, tolerance indices like the Hilsenhoff Biotic Index, and measures of dominance (e.g., percent dominance the most dominant taxon). Using the above formula, relative differences between upstream and downstream ranged from – 1.0 to +1.0 where a positive value indicated a higher degree of downstream impairment, and a value near zero indicated little or no difference. Sites with relative differences of < 0.15 were considered unimpaired.

Linear regression was used to determine relationships between “continuous” standard WET endpoints (LC50 and NOEC) and the derived benthic metric (average relative difference). Chi-square log-linear analysis of variance, and multiple discriminant analysis were used to test hypotheses related to categorical variables (i.e., pass/fail frequencies).

Dischargers that failed, 25% of their tests had $\leq 15\%$ chance of exhibiting an in-stream biological effect. Effluent dilution was the strongest predictor of the concordance between a WET test and the in-stream biological condition. Effluents that comprised > 80% of the receiving-environment volume under low-flow condition exhibited better relationships between WET and in-stream condition than effluents with greater dilution. Effluents that comprised < 20% of the low-flow receiving-environment volume had a low likelihood of correspondence to in-stream ecological effects, even if several toxicity tests were failures in a given year. The authors felt that WET tests were more predictive of in-stream ecological condition if several tests were conducted, more than one type of test was conducted, and when endpoints within a test were relatively consistent over time.

Problems with the study include decisions related to the benthic community metric (average relative difference). First, any of the selected metrics that were correlated in their responsiveness to effluents would over-weight and thus over-estimate the degree of impairment. Such a metric would be more valuable if it incorporated multiple orthogonal (or unrelated) metrics. Second, the cutoff applied to the benthic community metric (i.e., unimpaired was any site with an average relative difference < 0.15) was derived through what appears to have been a subjective process. There was no indication that effects were typically or generally statistically significant when they exceeded that value. Third, the formula for the overall benthic metric, though intended to provide a measure of relative (or percent) difference, did not estimate a percent difference at all. Percent differences are normally expressed relative to the upstream condition, such that the difference between upstream and downstream is divided by the upstream value. Here the difference between upstream and downstream was divided by the sum of the upstream and downstream values, somewhat underestimating the “relative” difference.

Post-Exposure Feeding Depression in *Daphnia*

McWilliam and Baird (2002) developed a post-exposure feeding test with *Daphnia magna* that they considered to be a sensitive indicator of water quality. Four-day old *Daphnia* were held in cages in each of 13-study area receiving environments, including reference areas, for 24 h. Feeding rates of *Daphnia* were determined with exposed specimens. Benthic invertebrates

were collected using kick-nets. The authors calculated the Biological Monitoring Working Party (BMWP) and Average Score per Taxon (ASPT) scores, two relatively common metrics used to summarize the pollution tolerance of invertebrate community samples in the UK and Europe. Feeding depression was observed at each of the “contaminated” sites, while invertebrates were considered impaired at only one location. The authors concluded that the feeding-depression test was a more sensitive, ecologically relevant diagnostic endpoint. Simple Pearson correlations were used to demonstrate a lack of relationship between the sublethal test result and metrics of the invertebrate community.

Upper Powell River, Virginia

Soucek et al. (2001) incubated clams in cages in 19 locations in the Pucket’s Creek watershed (Virginia) for a 30-d period. Benthic communities were sampled in the same reaches using EPA’s rapid bioassessment protocols. Clam survival over 30 days was positively correlated with the relative abundance of benthos belonging to the order Ephemeroptera, while clam growth was correlated with the abundance of benthos belonging to the collector-filterer functional feeding groups. Though this study demonstrated co-occurrence of effects in caged animals and benthic invertebrates, this was an observational study from a location with considerable historical impacts, thus confounding the correlation.

AETE Studies

The Aquatic Effects Technology Evaluation (AETE) program conducted a number of studies, including the literature review by Sprague (1997), as well as field studies evaluating different monitoring components. Field studies were carried out in 1997 (Beak and Golder, 1998). One of the questions tested during the field program was whether there was a link between effluent sublethal toxicity test results and instream biological (fish or benthos) responses. Beak International Incorporated (Beak, 1998) analyzed the data from Heath Steele. Based on % inhibition vs effluent concentration curve for *Ceriodaphnia* reproduction, fathead minnow growth, algae (*Selenastrum*) growth, duckweed (*Lemna minor*) growth, Beak estimated the toxicity of the receiving environment water in each of the five reaches for which there were biological response data, for three dates (i.e., the dates that the sublethal tests were conducted). They then correlated the estimated % inhibition with field measurements of biological responses. With that analysis, Beak observed that fish catch per unit effort (CPUE) and number of taxa decreased with predicted water toxicity to algae, *Ceriodaphnia*, duckweed and fathead minnow. They also observed that dominance of pollution-tolerant chironomids increased with increases in predicted water toxicity. Other indices of benthic community composition were unrelated to sublethal toxicity endpoints. The correlations observed simply demonstrated a gradient of responses in biological condition downstream of the effluent release. The results did not demonstrate a correlation between toxicological and biological responses because the observations were not synoptic. The data did provide another example of the co-occurrence of sublethal toxicity and receiving-environment effects, but receiving environment effects were as easily explained by historical inputs as by effluent quality at that time except for Heath Steele where considerable effort went into selecting the study site so as to specifically avoid historical contamination.

As summarized in Scroggins et al. (2002a), additional comparisons between sublethal toxicity endpoints and receiving-environment biological responses were made as part of studies at Dome Mine (Timmins) and Myra Falls (Vancouver Island), though more qualitatively. At Dome Mine the lowest IC25 values for *Ceriodaphnia*, fathead minnow, algae and duckweed representing effluent that was discharged to the river were < 6.25, 47, 27 and 3.7% effluent respectively. The estimated effluent concentrations in the river exceeded all of these values except for the fathead minnow IC25 (47%). The results of the sublethal toxicity tests, with the exception of the fathead minnow results, were considered evidence of potential receiving-environment effects. Benthic communities in the nearfield environment had decreased density and diversity, and were missing clams of the genus *Pisidium*. Adult pearl dace and yellow perch had higher condition in the nearfield environment, a finding somewhat contradictory to a mine-related effect. The lack of fathead minnow sublethal effects therefore co-occurred with a lack of effects in the two sentinel species in the receiving environment. The observed effects in the benthic community could as easily have been caused by historical contamination of the sediments, and not effluent quality.

Effects on the benthic community were observed at Myra Falls at metal concentrations in water that produced sublethal toxicity in *Ceriodaphnia*, duckweed and algae. Fathead minnows did not show any deleterious effects of mine effluent, and thus did not co-occur with the observed benthic community impairment. This study provides an example of the co-occurrence of sublethal and receiving-environment biological responses, but does not demonstrate a correlation between responses.

The three AETE studies combined provided three synoptic observations of sublethal toxicity and receiving-environment effects. The AETE effort thus produced a relatively small data set.

Salinas River, California

Anderson et al. (2003) examined the toxicity of agricultural runoff (tile drains) in the Salinas River in California. Samples of river water were collected and used in 96-h survival tests with *Ceriodaphnia dubia*. Benthic macroinvertebrates were collected using a traveling kick and sweep technique following typical USEPA guidance (Barbour et al., 1999). With a total of 16 stations, correlations between indices of benthic community composition and survival of *C. dubia* were assessed using non-parametric Spearman Rank tests. Number of mayfly taxa and percent of the benthic fauna as chironomids were generally positively correlated with survival of *C. dubia*.

Leading Creek, Ohio

Kennedy et al. (2003) evaluated growth of the clam *Corbicula* in situ, conducted acute 48-h tests with *Ceriodaphnia dubia*, chronic growth tests with *C. dubia*, and collected benthos within the vicinity of a discharge for a coal-processing effluent. Fecundity of *C. dubia*, and growth of *Corbicula* were impaired by the effluent. Mayflies were significantly reduced (99%) downstream of the effluent discharge point. This study presented the results from a single location only, and demonstrates the co-occurrence of effects, not a correlation between synoptic measures.

Canadian Pulp and Paper EEM

A variety of papers have been written summarizing the results from the sublethal studies conducted within the federal pulp and paper EEM program, and the studies summarized here have reported on the associations between sublethal endpoints and receiving-environment effects. Scroggins et al. (2002b) reported that sublethal tests conducted on an effluent from a pulp mill correctly predicted the “zone of potential effects” (ZPE) in a receiving water, agreeing with effects observed in biological surveys. The calculation of the ZPE is described in Environment Canada’s Technical Guidance Documents for the EEM programs for both the pulp and paper and metal mining sectors. Scroggins et al. (2002b, 2005) suggested that overlapping zones from multiple effluents could also be demonstrated with the ZPE calculation.

Moody (2002), Borgman et al. (2004) evaluated the relationship between sublethal toxicity data and field effects using data from 16 pulp mills in Ontario. Moody evaluated the association between sublethal toxicity tests and in-stream ecological effects. In her study, the ZPE was calculated for 16 mills in Ontario, on the basis of the most sensitive IC25 of mill effluents: one for each sublethal test species (inhibition of *Ceriodaphnia dubia* reproduction, *Selenastrum capricornutum* growth, and fathead minnow growth). The relationship between the zones of potential effect and actual extent of effects observed in the field was rated strong, moderate or weak based on the proximity of the ZPE estimate at the location of the affected biota. The relationship between the results of fathead minnow testing and the fish survey was described as strong or moderately strong at 40% of mills on rivers, and 80% of mills on lakes. Fathead minnow tests typically underestimated effects in the field, including six effluents that produced no toxic response at the end-of-pipe, but where there were apparent effects on natural field populations of fish, likely reflecting the effects of historical contamination in the field.

Both Moody (2002), Borgmann et al. (2004) and Scroggins et al. (2005) reported on the “Lab-to-Field” rating scheme that related whole-effluent toxicity to impacts observed in the benthic community. In this scheme, the number of potential mill-related effects for field measurements was compared to the severity of the effluent toxicity as determined by the lab tests. A rating scheme was applied to both toxicity test results and ecological endpoints. Each of the sublethal tests was rated based on the most sensitive IC25. Percent effects and thus an LTF rating were calculated for the sentinel fish survey endpoints, using all possible comparisons for each gender and species of fish collected in reference and exposure areas. Percent effects and an LTF rating were also calculated for invertebrate communities based differences in abundance, richness, Bray-Curtis distance measure, Simpson’s Diversity, Simpson’s Evenness, percent EPT (Ephemeroptera, Plecoptera, Trichoptera), as well as key major groups (e.g., Oligochaeta, Arthropoda, Chironomidae, Amphipoda, Lumbriculidae). Regression analysis of LTF scores revealed that the relationship between *Ceriodaphnia* reproduction test and benthic invertebrate field survey measurements was significant ($r=0.79$). The authors concluded that there was insufficient data to determine if this approach could be used as a predictive tool, and also concluded that the LTF rating scheme would benefit from a more sensitive species or life stage of fish to strengthen the sublethal test-to-fish survey relationship. Having ranked effects on a scale from 1 to 5 minimized the scale across which correlations between monitoring components might have occurred. The relationships might have been stronger had the authors

simply used the IC25 as the predictor of ecological effects, and if the ecological effects had more simply been expressed as the percent of endpoints indicating an effect.

Walker et al. (2005) conducted a national-scale analysis incorporating available EEM data for all mills across the country, including those used in the Moody etc. Ontario study. They calculated effect sizes for the invertebrate community survey data, as well as for the sentinel fish survey data. Benthic effect sizes were based on observed differences between reference and exposure areas in both abundance and richness. Fish effect sizes were based on differences in condition, liver weight, and gonad weight between reference and exposure areas expressed as a percentage of the reference area average value. Preliminary analysis, involving a series of bivariate correlations, revealed mostly a lack of correlation between laboratory toxicity (IC25s) and field responses. The few statistically significant (though weak) correlations that were found were in the opposite direction than that predicted by the laboratory toxicity tests. For example, one of the significant correlations was due to higher levels of fish toxicity measured in the laboratory being associated with improved fish condition in the field. Walker et al. (2005) ascribed the general lack of correlation (or counterintuitive correlations) to a variety of factors, including the fact that laboratory toxicity tests typically do not account for a multitude of direct and indirect interacting effects that occur in the field, including eutrophication (the most common field effect measured in the EEM studies), variable effluent dilution, the modifying effects of the receiving environment on effluent toxicity, and a large number of other abiotic and biotic modifying factors (see Marcus and McDonald 1992, Clements and Kiffney 1996, Chapman 2000 for further discussion).

Colorado REMAP Studies

In 1993 and 1994, U.S. Environmental Protection Agency (U.S. EPA) conducted a Regional Environmental Monitoring and Assessment Program (REMAP) survey on wadeable streams in Colorado's (USA) Southern Rockies Ecoregion (Griffith et al., 2004). The surveys collected data on fish and macroinvertebrate assemblages, physical habitat, and sediment and water chemistry and toxicity. The authors measured toxicity of surface waters at 73 locations using 48-h *Pimephales promelas* and *Ceriodaphnia dubia* lethality tests. Stations were classified as affected or not on the basis of water quality (exceeding dissolved metals criteria or not), and toxicity test results (i.e., more toxic than controls or not). Many metrics of macroinvertebrate and fish community composition were statistically significantly different (based on analysis of variance) between sites classified as having toxic or non-toxic surface waters. The study was conducted in a region that had been subject to pressures from mining operations for several years. Historical contamination, therefore, confounded the interpretation the association between toxicity and biological effects.

Summary and Recommendations

Summary of findings

Section 2.0 described some recommended approaches to developing the burden of evidence that one indicator can be reliably used to predict the condition of another. Many of the papers reviewed in Sections 3.0 above either did not adequately describe ecological condition, did not collect synoptic samples across a gradient of conditions, or did not have enough synoptic observations with which to test for associations between sublethal toxicity endpoints and receiving-environment biological endpoints. For those studies with enough data, appropriate statistical tools generally were not used. Where the data were adequate and statistical methods were appropriate studies generally reported mixed results. All but one of the studies used observational data from areas that had been previously impacted by effluents or other wastes. The presence of historical contamination confounds the conclusions related to the nature of the association between sublethal toxicity and receiving-environment biological effects. Therefore, interpretation of correlations was confounded for most of the studies. In spite of this, there were three general conditions when sublethal toxicity tended to co-occur with receiving-environment biological effects.

Condition 1: Sublethal toxicity and ecological effects co-occurred more frequently when effluent comprised a relatively large fraction of the total receiving-environment flow volume (i.e., > 20% or so, Diamond and Daley, 2000), and less frequently when effluent volumes were a smaller fraction of the receiving-water volume (see also Eagleson et al., 1990).

Condition 2: Sublethal toxicity and ecological effects co-occurred more frequently when sublethal tests were conducted using ambient receiving water. Ambient-environment tests take into account mixing and any changes in toxicity associated with alterations in chemistry of the final effluent (i.e., pH, suspended solids).

Environment Canada has published numerous multi-application methods for testing acute and sublethal toxicity in aquatic test samples (e.g., EPS 1/RM/9 (acute lethality to rainbow trout); EPS 1/RM/11 (acute lethality to *Daphnia* spp.); EPS 1/RM/21 (sublethal toxicity to *Ceriodaphnia*); EPS 1/RM/22 (sublethal toxicity to fathead minnow); EPS 1/RM/25 (sublethal toxicity to algae); and EPS 1/RM/28 (sublethal toxicity to early life stages of rainbow trout). These more generic test methods allow the use of receiving water as control/dilution water in testing and have specific chapters that provide direction on the direct testing of effluent-impacted receiving waters. There are two other Environment Canada acute lethality method documents cited in the MMER (and PPER) but these methods are used specifically for effluent compliance testing (EPS 1/RM/13 – rainbow trout Reference Method and EPS 1/RM/14 – *Daphnia magna* Reference Method) and are not used in the EEM programs. These more specific Reference methods are used for pass or fail compliance monitoring and require the use of laboratory dilution water (i.e., reconstituted water, groundwater, or dechlorinated tap water).

As the EEM sublethal toxicity tests are not used as compliance tests in MMER or PPER regulations, these more multi-purpose methods can be used to monitor effluents directly (as currently in the EEM program) or to monitor receiving waters that have been impacted by the effluent discharged from an industrial site (i.e., within the effluent mixing zone, typically the near field area).

Condition 3: Sublethal toxicity co-occurred more frequently with receiving-environment effects when multiple effluent samples are tested and averaged. Diamond and Daley (2000) found that a single sublethal test is not a good predictor of potential receiving-environment impairment since one test cannot characterize the degree of variability in effluent quality.

Recommendations for data analysis

There is presently inadequate data to demonstrate that sublethal toxicity tests are predictive of receiving-environment effects. Sublethal toxicity at the end-of-pipe often co-occurs with receiving-environment biological effects, but mainly when effluents dominate the receiving environment. Where effects co-occur, there tend to be inconsistencies in the endpoints that correlate. All but one of the studies that evaluated the relationship between sublethal toxicity and receiving-environment effects were observational studies conducted on receiving environments that had historical impacts that confounded the interpretation of correlation. A more valid assessment of the correspondence between sublethal toxicity tests and receiving-environment effects could be accomplished through observational studies if the data were collected at locations where there is no historical contamination. The data analytical approaches described below would be suitable, assuming that the data are from sites with no historical contamination. The data analytical approaches described below could also be used to analyze the existing data available through the metal mining EEM program, though with the understanding that any conclusions related to concordance between sublethal and receiving-environment effects could be limited because of the confounding effects of historical contamination.

As from Section 2.0, an evaluation of the nature and strength of associations between toxicological and biological data is required, as is an assessment of toxicological data to reliably declare biological measures to be impaired or not. The multivariate nature of sublethal toxicity testing data within the federal metal-mining EEM program, and the multivariate nature of the ecological responses requires a consideration of multivariate techniques for assessing correlations. In the Canadian EEM programs, the sublethal endpoints are represented by growth and reproduction of various taxa. The responses are the concentration of effluents causing a 25% reduction in growth and reproduction relative to a control group. The ecological endpoints are represented by the fish population survey and a survey of benthic macroinvertebrates.

The invertebrate community survey produces counts of various species in both reference and exposure areas. Those data are used to estimate key response variables (total abundance, and number of taxa – number of families). The technical guidance documents for the metal-mining and pulp and paper EEM programs recommend calculating “effect sizes” for measures of

benthic community composition (i.e., abundance, richness) that are expressed relative to the background variability (i.e., the standard deviation, either estimated from the reference area, or estimated from a pooled estimate from reference and exposure areas). That re-expression of effect sizes is just like a signal-to-noise ratio, and is a commonly used approach to quantify the magnitude of effects. Effect sizes calculated in this way from any site can be compared to the effect sizes compared from any other site, despite differences in receiving-environment type, sampling methods, taxonomy, etc. Effect sizes for abundance and richness can be directly compared with observed effects in the sublethal toxicity tests. A canonical correlation would be the appropriate approach for evaluating both the strength and nature of the association. Alternatively, a multivariate discriminant analysis could be used if the benthic effect sizes could be used to categorize effects as unimpaired (e.g., effects < 2 SDs) and impaired (effects > 2SDs). The resulting discriminant model is the categorical analogue to canonical correlation used for regression questions. Both canonical correlation and discriminant analysis are useful for illustrating which of the toxicological endpoints vary with the benthic community endpoints or impairment classification. One could take the further step of using the discriminant model to predict the impairment class of the benthic community for future cases. The use of parametric canonical correlation and discriminant analysis has assumptions including multivariate normality and homogeneity of variances and covariances. Appropriate transformations of response and predictor variables, and the use of somewhat independent (uncorrelated) predictor variables will ensure more valid multivariate analysis. Violation of assumptions (which generally always occurs) is less worrisome with larger datasets (Green, 1979) typically associated with EEM analyses (Lowell et al., 2003; Walker et al., 2005).

The fish survey produces weights (liver, gonad, body), lengths (total, standard, fork), and age data for generally two species of fish, two genders, from each of two areas (reference and exposure). With normally two species, two genders and five response endpoints, there are up to 20 total endpoints that comprise the results of an EEM fish survey. Assuming a single species and gender, there are only five responses that are measured synoptically with sublethal tests at a mine. As per Walker et al. (2005), the percent difference in each of the fish population endpoints relative to fish from a reference area would be a rational transformation of the data in order to examine correlations with the toxicological data. Canonical correlation would be the appropriate approach to evaluate the strength and nature of the toxicology-biology relationship, while a discriminant model would be useful for assessing classification frequencies (as above for the benthic endpoints).

A lack of correlation between sublethal and receiving-environment biological endpoints, as observed by Walker et al. (2005), might reflect differences in effluent concentrations in the receiving environment. Oddly, few of the studies in Section 3.0 considered the influence of effluent concentration on the sublethal-biology correlation. The influence of the effluent concentration should be factored out, either through a tool like stepwise-multiple regression (see Stanfield and Kilgour, 2006, for an example with receiving-environment-type data).

As described by Swets (1988) and Murtaugh (1996), and as described in Section 2.0 above (see also Figure 1), some analysis of the frequency of making correct and incorrect predictions about the condition of biological endpoints must be made. This analysis requires that we define, *a priori*, the magnitude of changes in receiving-environment biological endpoints that we deem

important. The federal EEM programs have previously defined critical effect sizes (Environment Canada, 2005; Kilgour et al., 2005, Kilgour et al., 2006) that should be adopted for this exercise. Criteria for the sublethal endpoints also need to be specified, preferably *a priori*, potentially something like an EC25. Alternatively, inspection of biplots of sublethal vs receiving-water biological endpoints may suggest sublethal cutoffs that provide reasonable classification frequencies. Kilgour (1997), for example, used fitted regressions between predictor (benthos) and dependant (fish) variables to determine the likelihood of making correct statements about the impairment status of fish communities using benthic communities, and chose as the cutoff for the predictor variable the value that coincided (along the regression line) with the critical value for the fish community. An understanding of the true-positive (TP), true-negative (TN), false-negative (FN), and false-positive (FP) frequencies is necessary in order to fully understand the contribution that sublethal tests could potentially make to effluent management programs. These classification frequencies could be constructed using a canonical correlation/canonical variates analysis, but the mathematics would be difficult to communicate to lay practitioners. It would be more sensible to calculate the classification frequencies for bivariate combinations of sublethal and biological endpoints (as per Walker et al., 2005).

Opinion on Using Sublethal Tests in EEM

Goals of EEM in relation to sublethal toxicity testing

Sublethal toxicity testing can be justified as a component of EEM programs a variety of ways, depending on the objectives of the monitoring program. Environment Canada developed the EEM program to characterize the condition of biological resources within aquatic environments receiving effluents from mines. The stated objective of EEM is to “evaluate the effects of liquid mine effluents on the condition of fish, fish habitat and the use of fisheries resources” (Walker et al., 2003; Environment Canada, 2005). Ultimately, Environment Canada is attempting to ensure healthy aquatic ecosystems with unimpaired human uses, as well as diverse and functioning aquatic ecosystems (see Kilgour et al., 2005). Unimpaired human uses are determined directly through analysis of fish tissues for contaminants of anticipated concern. The diversity and functioning of aquatic ecosystems is ensured when the dominant trophic levels (fish) are healthy and are characteristic of what would normally be found in unimpaired reference conditions (Kilgour et al., 2005). It is not always possible to directly or quickly quantify fish communities, or other higher trophic levels. In some cases there is a recognized desire to have an early warning of potential long-term problems, and not monitor the trophic level we are protecting (i.e., fish, see Kilgour et al., 2005), because impacts could be irreversible by the time we detect them. Therefore, other components of the aquatic environment such as sentinel fish populations and benthos are usually monitored because they are believed to be predictive of effects on fish. Within this approach there is also an implied willingness to accept some amount of change in early-warning indicators such as sentinel fish population or the benthic community (Kilgour et al., 2005, 2006).

Early warning of potential impacts

One of the potential uses of sublethal toxicity testing is to provide an early marker of developing effects in the receiving environment. Within this context, the Canadian EEM programs for pulp and paper, mining, and now the municipal wastewater sectors have endpoints that could be considered as providing this early-warning indication. For example, sentinel fish population endpoints and indices of benthic community composition are considered early-warning indicators (as per Cairns et al., 1993; Kilgour et al., 2005). Effects on indices of benthic community composition that deviate from the mean reference response by more than ± 2 standard deviations are considered a “warning” of the potential for other effects requiring further monitoring. Effects smaller than that are acceptable. Benthic community studies do not elicit a real concern unless effects are large (> 2 SDs) and get larger over time (either spatially or in terms of effect size), or when the community is dominated by one or a few species that are known to be very tolerant to extremely degraded conditions (e.g., *Tubifex* worms in freshwater, or the polychaete *Capitella* in saltwater; Kilgour et al., 2005). These very large effects are considered important because they tend to co-occur with impaired fish communities (Yoder and Rankin, 1995; Kilgour, 1997). Similarly, differences in gonad size, liver size or growth between

reference fish and fish exposed to an effluent of < 25% are considered acceptable, while effects larger are considered “warning” of potential degradation requiring continued monitoring. Effects that get larger over time, either in terms of effect size or geographic extent are considered unacceptable because they indicate the receiving environment cannot assimilate the inputs. Growth, energy storage and energy utilization tend to be affected in populations where the fish community has been altered through overfishing, recruitment failure, stocking programs, or the invasion of habitat by non-native species (e.g., Munkittrick and Dixon, 1989a,b). Indicators of growth, energy storage and energy use can, therefore, be used to predict when the fish community has been altered from its normal composition (Munkittrick and Dixon, 1989a,b; Gibbons et al., 2004; Kilgour et al., 2005).

longer term perspective on the outcomes of EEM Monitoring

Data from the federal EEM program will be used to determine if the existing Metal-Mining Effluent Regulations (MMER) are protective of the environment. Data will be collected over several years through many “cycles”, after which a burden of evidence of the protectiveness of the effluent regulations will be determined. The principal data used to determine if there were effects, and to determine if the receiving environments were protected, will be sentinel fish population parameters and indices of benthic community composition. Conclusions of the condition of the receiving environments can be made in the absence of sublethal toxicity tests. Sites that continue to show a lack of receiving-environment effects year-after-year would be used as evidence that the effluent regulation has been protective, and sublethal toxicity studies would be unnecessary. Sites that show effects on fish and benthos, would provide evidence that the environment had not been protected and that the environmental damage had occurred. In this case sublethal toxicity testing would be unnecessary for making that conclusion however, it is conceptually possible that sublethal test could have provided an early indication that damage was occurring. Sites that produced effects would trigger the investigation of cause phase of EEM, which would include a variety of studies. The sublethal toxicity information can be used as part of the burden of evidence that effects are effluent related. Sublethal toxicity tests could be used to demonstrate, for example, that effluent is non-toxic, supporting a conclusion that observed receiving-environment effects are potentially due to other inputs and/or historical contamination. Sublethal toxicity tests can also be used to identify specific effluent processes causing effects (Hewitt et al., 2005).

Broader considerations on sublethal toxicity testing and EEM

The federal EEM program could be considered an inefficient effluent management tool because the overall feedback time, through revision of the MMER's, takes a long time (upwards of a 20-y process). Where there is a desire to manage effluents more proactively, tests are required that provide rapid response. Sublethal tests can provide very rapid response, whereas field studies tend to require more time. As a monitoring tool, end-of-pipe effluent toxicity tests (lethal, sublethal) provide a biological measure of variations in effluent quality. Small and large process changes, differences in ore-body chemistry, start-up, shut-down, etc., can all influence end-of-pipe toxicity. End-of-pipe toxicity tests give an immediate measure, and provide an opportunity to prevent continued discharge if the risks of a receiving-environment impact are considered

high. Most international approaches to the management of risks associated with discharge of liquid effluents require dischargers to conduct toxicity testing at the end of pipe (normally acute lethality testing), while only a few require dischargers to conduct in-stream biological monitoring (Kilgour et al., 2003). The EEM program in Canada is unique in requiring acute and sublethal toxicity testing, as well as receiving-environment biological studies (Power and Boumphrey, 2004). NPDES permits in the USA generally specify limits on water quality and for acute and chronic whole-effluent tests (Hudiburgh, 1995; Power and Boumphrey, 2004) with receiving-environment biological studies generally carried out by state government agencies. Sublethal toxicity tests are currently used as part of Water-Effects-Ratio (WER) studies to determine the effects of local surface water quality on the toxicity of effluent-related substances and the concentrations of substances that pose risks to aquatic biota (CCME, 2003).

Sublethal toxicity tests could be used as a measure of effluent quality that was predictive of effects in the receiving environment **if the sublethal endpoint reliably predicts a receiving-environment effect**. If that assumption is valid, then the sublethal toxicity test would make a useful contribution to EEM **if a reliable prediction of receiving-environment effects can be made in less time and with less effort and resources than is required for a valid field study**.

In consideration of the above, the frequencies of testing and the turn-around time for results should also be part of the considerations. If frequent enough, sublethal testing permits real-time assessment of process changes, interruptions or perturbations in treatment systems. Ecological studies do not lend themselves to this type of study since the outcome remains unknown until an effect is observed (or not). On the other hand, short-term interruptions or perturbations, etc may not produce long-term impacts downstream.

Despite meeting the above assumptions, sublethal toxicity tests represent “snapshots” in time taken in a laboratory context. Sublethal toxicity studies are also unable to detect/address impacts associated with sediment quality, genotoxicity, bioaccumulation, dietary toxicity, trophic transfer of contaminants, and enrichment (Black et al., 1996). Therefore, the relevance of sublethal toxicity tests as a routine effluent monitoring tool will be questioned. Attempts to mimic the natural environment will only be seen as exactly that. Unless (perhaps even with) validated linkages between sublethal toxicity test endpoints and receiving environment impacts, a lack of confidence will likely generally lead proponents to conducting field studies in order to confirm effects on field-based early-warning indicators and prior to effluent management (Black et al., 1996; Lowell et al., 2000; Suter et al., 2002; Chapman and Anderson, 2005).

Literature Cited

Amman, L.P., W.J. Birge, K.L. Dickson, P.B. Dorn, M.E. LeBlanc, D.I. Mount, B.R. Parkhurst, H.R. Preston, S.C. Schimmel, A. Spacie and G.B. Thursby. 1999. Predicting instream effects from WET tests, Discussion Synopsis. In, D.R. Grothe, K.L. Dickson and D.K. Reed-Judkins (eds), Whole Effluent Testing: An Evaluation of Methods and Prediction of Receiving System Impacts. SETAC Press, Pensacola.

Anderson, B.S., J.W. Hunt, B.M. Phillips, P.A. Nicely, K.D. Gibert, V. De Vlaming, V. Connor, N. Richard and R.S. Tjeerdema. 2003. Ecotoxicological impacts of agricultural drain water in the Salinas River, California, USA. *Environmental Toxicology and Chemistry*, 22:2375-2384.

Barbour, M.T., J. Gerritsen, B.D. Snyder and J.B. Stribling. 1999. Rapid bioassessment protocols for use in streams and Wadeable rivers: periphyton, benthic macroinvertebrates, and fish. Second Edition. EPA 841-B-99-002.

Beak and Golder (Beak International and Golder Associates). 1998. Summary and cost effectiveness evaluation of aquatic effects monitoring technologies applied in the 1997 AETE field evaluation program. Beak Reference 20776.1.

Beak International Incorporated. 1998. 1997 Field Program Final Report, Heath Steele Site, New Brunswick. Report prepared for Aquatic Effects Technology Evaluation (AETE) Program, Canada Centre for Mineral and Energy Technology (CANMET), Natural Resources Canada. AETE Project 4.1.3.

Birge, M.T., H.A. Black and T.M. Short. 1989. A comparative ecological and toxicological investigation of a secondary waste-water treatment plant effluent and its receiving stream. *Environmental Toxicology and Chemistry*, 8:437-450.

Black, J.A., D.T. Burton, G.M. DeGraeve, M.A. Heber, N.E. LeBlanc and M.C. Lewis. 1996. Session 7, Workshop Summary and Conclusions, Chapter 11 in D.R. Grothe, K.L. Dickson and D.K. Reed-Judkins (eds), Whole Effluent Toxicity Testing: An Evaluation of Methods and Prediction of Receiving System Impacts. SETAC Press, Pensacola.

Borgman, A.I., M.J. Moody and R.P. Scroggins. 2004. The lab-to-field (LTF) rating scheme: a new method of investigating the relationships between laboratory sublethal toxicity tests and field measurements in environmental effects monitoring studies. *Human and Ecological Risk Assessment*, 10:683-707.

Cairns, J. Jr. 1983. Are single species toxicity tests alone adequate for estimating environmental hazard? *Hydrobiologia*, 100:47-57

Cairns, J. Jr. 1988. Should regulatory criteria and standards be based on multispecies evidence? *Environmental Professional*, 10:157-165.

Cairns, J. Jr., P.V. McCormick and B.R. Niederlehner. 1993. A proposed framework for developing indicators of ecosystem health. *Hydrobiologia*, 263:1-44.

CCME (Canadian Council of Ministers of the Environment). 2003. Canadian Water Quality Guidelines for the Protection of Aquatic Life, pages 83 to 87, 142 to 144.

Chapman PM. 2000. Whole effluent toxicity testing – usefulness, level of protection and risk assessment. *Environmental Toxicology and Chemistry*, 19:3–13.

Chapman, P.M. 2002. Whole effluent toxicity testing -usefulness, level of protection, and risk assessment. *Environmental Toxicology and Chemistry*, 19:3-13.

Chapman, P.M., and J. Anderson. 2005. A decision-making framework for sediment contamination. *Integrated Environmental Assessment and Management*, 1:163-173.

Clements, W.H. and P.M. Kiffney. 1994. Integrated laboratory and field approach for assessing impacts of heavy metals at the Arkansas River, Colorado. *Environmental Toxicology and Chemistry*, 13:397-404.

Diamond, J. and C. Daley. 2000. What is the relationship between whole effluent toxicity and instream biological condition? *Environmental Toxicology and Chemistry*, 19:158-168.

Dickson, K.L., W.T. Waller, J.H. Kennedy and L.P. Ammann. 1992. Assessing the relationship between ambient toxicity and instream biological response. *Environmental Toxicology and Chemistry* 11:1307-1322.

Dickson, K.L., W.T. Waller, J.H. Kennedy, W.R. Arnold, et al. 1989. A water quality and ecological survey of the Trinity River, Volumes I and II. Final Report. City of Dallas Water Utilities, Dallas, TX.

Eagleson, K.W., D.L. Lenat, L.W. Ausley and F.B. Winborne. 1990. Comparison of measured instream biological responses with responses predicted using the *Ceriodaphnia dubia* chronic toxicity test. *Environmental Toxicology and Chemistry*, 9:1019-1028.

Environment Canada. 2002. Metal mining guidance document for aquatic environmental effects monitoring. Environment Canada, Gatineau, QC.

Gibbons, W.N. and K.R. Munkittrick. 1994. A sentinel monitoring framework for identifying fish population responses to industrial discharges. *Journal of Aquatic Ecosystem Health*, 3:227-237.

Girling, A.E., A.M. Riddle, G.M. Mitchell, P.K. Chown, D. Tinsley, C. Buckler, I. Johnson and R. Benstead. 2004. Estimating spatial patterns of effluent exposure concentrations in direct toxicity assessment studies. *Ecotoxicology*, 13:449-461.

Green, R.H. 1979. Sampling Design and Statistical Methods for Environmental Biologists. John Wiley, New York.

Griffith, M.B., J.M. Lazorchak and A.T. Herlihy. 2004. Relationships among exceedances of metals criteria, the results of ambient bioassays, and community metrics in mining-impacted streams. *Environmental Toxicology and Chemistry*, 23:1786-1795.

Hartwell, S.I., C.E. Dawson, E.Q. Durell, R.W. Alden, P.C. Adolphson, D.A. Wright, G.M. Coelho, J.A. Magee, S. Ailstock And M. Norman. 1997. Correlation of measures of ambient toxicity and fish community diversity in Chesapeake Bay, USA tributaries - urbanizing watersheds. *Environmental Toxicology and Chemistry*, 16:2556–2567.

Hewitt, L.M., M.G. Dubé, S.C. Ribey, J.M. Culp, R. Lowell, K. Hedley, B. Kilgour, C. Portt, D.L. MacLachy and K.R. Munkittrick. 2005. Investigation of cause in pulp and paper environmental effects monitoring. *Water Quality Research Journal of Canada*, 40:261-274.

Hickey, C.W. and W.H. Clements. 1998. Effects of heavy metals on benthic macroinvertebrate communities in New Zealand streams. *Environmental Toxicology and Chemistry*, 17:2338-2346.

Hudiburgh, G.W. Jr. 1995. The Clean Water Act. In, G.M. Rand (ed), *Fundamentals of Aquatic Toxicology, Effects, Environmental Fate, and Risk Assessment*. 2nd edition.

Kennedy, A.J., D.S. Cherry and R.J. Currie. 2003. Field and laboratory assessment of a coal processing effluent in the Leading Creek Watershed, Meigs County, Ohio. *Archives of Environmental Contamination and Toxicology*, 44:324-331.

Kilgour, B.W. 1997. Fish-benthos correlations and effects on benthos that reflect significant effects on fish communities in southern Ontario streams. PhD thesis, University of Waterloo, Waterloo, Ontario. 190 pp.

Kilgour, B.W., D.G. Dixon and R.P. Lanno. 2003. Recommended framework to Integrate environmental quality objectives for aquatic environments. Jacques Whitford Environment Limited Project No. ONO50483 prepared for Environment Canada, Gatineau, QC.

Kilgour, B.W., M.G. Dubé, K. Hedley, C.B. Portt and K.R. Munkittrick. 2006. Aquatic environmental effects monitoring guidance for environmental assessment practitioners. *Environmental Monitoring and Assessment*.

Kilgour, B.W., K.R. Munkittrick, C.B. Portt, K. Hedley, J. Culp, S. Dixit, G. Pastershank. 2005. Biological criteria for municipal wastewater effluent monitoring programs. *Water Quality Research Journal of Canada*, 40:374-387.

Stanfield, L.W. and B.W. Kilgour. 2006. Effects of Percent Impervious Cover on Fish and Benthos Assemblages and In-Stream Habitats in Lake Ontario Tributaries. In, R. M. Hughes, L.

Wang, and P. W. Seelbach (eds). Influences of landscape on stream habitats and biological assemblages. American Fisheries Society, Symposium 48: 577-599, Bethesda, Maryland.

Lapoint, T.W. and W.T. Waller. 2000. Field assessment in conjunction with whole effluent toxicity testing. *Environmental Toxicology and Chemistry*, 19:14-24.

Lowell, R.B. J.M. Culp and M.G. Dubé. 2000. A weight-of-evidence approach for northern river risk assessment: integrating the effects of multiple stressors. *Environmental Toxicology and Chemistry*, 19:1182-1190.

Lowell, R.B., S.C. Ribey, I.K. Ellis, E.L. Porter, J.M. Culp, L.C. Grapentine, M.E. McMaster, K.R. Munkittrick and R.P. Scroggins. 2003. National assessment of the pulp and paper environmental effects monitoring data. National Water Research Institute, NWRI Contribution No. 03-521.

Maltby, L., S.A. Clayton, H. Yu, N. McLoughlin, R.M. Wood and D. Yin. 2000. Using single-species toxicity tests, community-level responses, and toxicity identification evaluations to investigate effluent impacts. *Environmental Toxicology and Chemistry*, 19:151-157.

Maltby, L. and G.A. Burton. 2006. Field-based effects measures. *Environmental Toxicology and Chemistry*, 25:2261-2262.

Marcus, M.D. and L.L. McDonald. 1992. Evaluating the statistical bases for relating receiving water impacts to effluent and ambient toxicities. *Environmental Toxicology and Chemistry*, 11:1389-1402.

McWilliam, P.A. and D.J. Baird. 2002. Application of postexposure feeding depression bioassays with *Daphnia magna* for assessment of toxic effluents in rivers. *Environmental Toxicology and Chemistry*, 21:1462-1468.

Moody, M. 2002. Assessment of relationship between laboratory sublethal toxicity and field measurements through the review of the Ontario Region Environmental Effects Monitoring (EEM) studies. Saskatchewan Research Council, Environment Branch, Saskatoon. SRC Publication No. 11415-1E01.

Mount, D.I. and T.J. Norberg-King (eds). 1985. Validity of effluent and ambient toxicity tests for predicting biological impact, Scippo Creek, Circleville, Ohio. Duluth MN. US Environmental Protection Agency, Office of Research and Development, EPA600/3-85/044.

Mount, D.I. and T.J. Norberg-King (eds). 1986. Validity of effluent and ambient toxicity tests for predicting biological impact, Kanawha River, Charleston, West Virginia. Duluth MN. US Environmental Protection Agency, Office of Research and Development, EPA600/3-86/006.

Mount, D.I., A.E. Steen, T.J. Norberg-King (eds). 1985. Validity of effluent and ambient toxicity testing for predicting biological impact on Five Mile Creek, Birmingham, Alabama. Duluth MN: USEPA Office of Research and Development. EPA/600/8-85/015.

Mount, D.I., N.A. Thomas, T.J. Norberg, M.T. Barbour, T.H. Roush and W.F. Brandes. 1984. Effluent and ambient toxicity testing and instream community response on the Ottawa River, Lima, Ohio. Duluth MN: US Environmental protection Agency, Office of Research and Development. EPA-600/3-84-080.

Mount, D.I., T.J. Norberg-King, and A.E. Steen (eds). 1986a. Validity of effluent and ambient toxicity tests for predicting biological impact, Naugatuck River, Waterbury, Connecticut. Duluth MN. US Environmental Protection Agency, Office of Research and Development, EPA600/8-86/005.

Mount, D.I., A.E. Steen, T.J. and Norberg-King, (eds). 1986b. Validity of effluent and ambient toxicity tests for predicting biological impact, Ohio River, near Wheeling, West Virginia. Duluth MN. US Environmental Protection Agency, Office of Research and Development, EPA600/3-85/071.

Mount, D.I., A.E. Steen, T.J. and Norberg-King, (eds). 1986c. Validity of effluent and ambient toxicity tests for predicting biological impact, Back River, Baltimore harbor, Maryland. Duluth MN. US Environmental Protection Agency, Office of Research and Development, EPA600/8-86/001.

Munkittrick, K.R. and D.G. Dixon. 1989a. An holistic approach to ecosystem health assessment using fish population characteristics. *Hydrobiologia* 188/189: 122-135.

Munkittrick, K.R. and D.G. Dixon. 1989b. Use of white sucker (*Catostomus commersoni*) populations to assess the health of aquatic ecosystems exposed to low-level contaminant stress. *Canadian Journal of Fisheries and Aquatic Sciences*, 46: 1455-1462.

Murtaugh, P.A. 1996. The statistical evaluation of ecological indicators. *Ecological Applications*, 6:132-139.

Nimmo, D.R., M.H. Dodson, P.H. Davies, J.C. Green, M.A. Kerr. 1990. Three studies using *Ceriodaphnia* to detect nonpoint sources of metals from mine drainage. *Research Journal of the Water Pollution Control Federation*, 62:7-15.

Norberg-King, T.J, and D.I. Mount. 1986. Validity of effluent and ambient toxicity tests for predicting biological impact, Skeleton Creek, Enid, Oklahoma. Duluth MN: US Environmental Protection Agency, Office of Research and Development. EPA/600/8-86/002.

Parkhurst, B.R. 1996. Discussion-Initiation Paper 10.2, Predicting Receiving System Impacts from Effluent Toxicity. In, Grothe, D.R., K.L. Dickson and D.K. Reed-Judkins (eds), *Whole Effluent Toxicity Testing: An Evaluation of Methods and Prediction of Receiving System Impacts*. SETAC Press, Pensacola.

Power, E.A. and R.S. Bousphrey. 2004. International trends in bioassay use for effluent management. *Ecotoxicology*, 13:377-398.

Sarakinos, H.C. and J.B. Rasmussen. 1998. Use of bioassay-based whole effluent toxicity (WET) tests to predict benthic community response to a complex industrial effluent. *Journal of Aquatic Ecosystem Stress and Recovery*, 6: 141–157

Sarazin-Delay, C.L. Benthic Invertebrate Community Survey using the Reference Condition Approach for INCO Ltd. Cooperative Freshwater Ecology Unit, Laurentian University, Sudbury. Report.

Scroggins, .R., G. van Aggelen and J. Schroeder. 2002a. Monitoring sublethal toxicity in effluent under the metal mining EEM program. *Water Quality Research Journal of Canada*, 37:279-294.

Scroggins, R.P., J.A. Miller and A.I. Borgmann and J.B. Sprague. 2002b. Sublethal toxicity findings by the pulp and paper industry for cycles 1 and 2 of the Environmental Effects Monitoring program. *Water Quality Research Journal of Canada*, 37:21-48.

Scroggins, R.P., A.I. Borgmann, J.A. Miller and M.J. Moody. 2005. Strategies for monitoring environmental effects of industrial effluents, In, C. Blaise and J.-F. Ferard (eds), *Small-Scale Freshwater Toxicity Investigations*, Vol. 2, 139-167.

Soucek, D.J., T.S. Schmidt and D.S. Cherry. 2001. In situ studies with Asian clams (*Corbicula fluminea*) detect acid mine drainage and nutrient inputs in low-order streams. *Canadian Journal of Fisheries and Aquatic Sciences*, 58:602-608.

Sprague, J.B. 1997. Review of methods for sublethal aquatic toxicity tests relevant to the Canadian metal-mining industry. Sponsored by: Canada Centre for Mineral and Energy Technology and the Mining Association of Canada on behalf of the Aquatic Effects Technology Evaluation (AETE) Program.

Suter, G.W. II, S.B. Norton and S.M. Cormier. 2002. A methodology for inferring the causes of observed impairments in aquatic ecosystems. *Environmental Toxicology and Chemistry*, 21:1101-1111.

Swets, J.A. 1988. Measuring the accuracy of diagnostic systems. *Science*, 240:1285-1293.

United States Environmental Protection Agency (USEPA). 1999. A review of single species toxicity tests: are the tests reliable predictors of aquatic ecosystem community responses? Office of Research and Development, Washington. EPA/600/R-97/114.

Walker, S.L., R.B. Lowell and J.P. Sherry. 2005. Preliminary analysis of pulp and paper environmental effects monitoring data to assess possible relationships between the sublethal toxicity of effluent and effects on biota in the field. *Water Quality Research Journal of Canada*, 40:251-260.

Walker, S.L., S.C. Ribey, L. Trudel and E. Porter. 2003. Canadian Environmental Effects Monitoring: experiences with pulp and paper and metal mining regulatory programs. *Environmental Monitoring and Assessment*, 88:311-326.

Waller, W.T., L.P. Ammann, W.J. Birge, K.L. Dickson, P.B. Dorn, N.E. LeBlanc, D.I. Mount, B.R. Parkhurst, H.R. Preston, S.C. Schimmel, A. Spacie and G.B. Thursby. 1996. Predicting instream effects from WET tests. In, D.R. Grothe, K.L. Dickson and D.K. Reed-Judkins (eds), *Whole Effluent Toxicity Testing: an evaluation of methods and prediction of receiving system impacts*. SETAC Press, Pensacola.

Yoder, C.O. and E.T. Rankin. 1995. Biological criteria program development and implementation in Ohio. In, W.S. Davis and T.P. Simon (eds), *Biological Assessment and Criteria, Tools for Water Resource Planning and Decision Making*. Lewis Publishers, Boca Raton.

Zar, J.H. 1984. *Biostatistical Analysis*, Second Edition. Prentice Hall, Englewood Cliffs, New Jersey.