

Environment
Canada

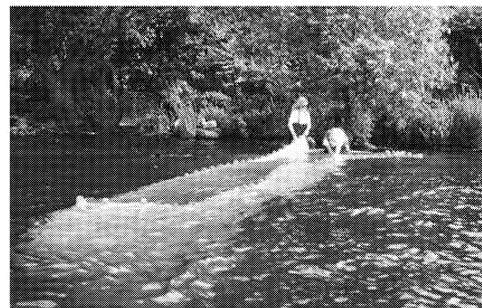
Environnement
Canada

Canada

National Assessment of Pulp and Paper Environmental Effects Monitoring Data: Findings from Cycles 1 through 3



R.B. Lowell, B. Ring, G. Pastershank, S. Walker, L. Trudel and K. Hedley



NWRI Scientific Assessment Report Series No. 5

To obtain additional copies:

National EEM Office
351 Saint. Joseph Blvd, 8th Floor
Place Vincent Massey
Gatineau, Québec
K1A 0H3
Phone: (819) 997-1535
EEM-ESEE@ec.gc.ca

or
Science Liaison Branch
National Water Research Institute
Environment Canada
867 Lakeshore Road, P.O. Box 5050
Burlington, Ontario L7R 4A6
nwriscience.liaison@ec.gc.ca

To view an html version or download a copy in pdf format visit the NWRI web site: <http://www.nwri.ca>

Library and Archives Canada Cataloguing in Publication

Main entry under title:

National assessment of pulp and paper environmental effects monitoring data: findings from cycles 1 through 3

(NWRI Scientific Assessment Report Series, ISSN 1499-5905; no. 5)

Issued also in French under the title: Évaluation nationale des données des études de suivi des effets sur l'environnement des fabriques de pâtes et papiers : Résultats des cycles 1 à 3

Includes bibliographical references.

ISBN 0-662-39350-3

Cat. no. En40-237/5-2005E

1. Wood-pulp industry – Waste disposal – Environmental aspects – Canada.
2. Effluent quality – Canada.
3. Fishes – Effect of water pollution on – Canada.
4. Benthic animals – Effect of water pollution on – Canada.
5. Water quality bioassay – Canada.
- I. Lowell, Richard Bruce, 1956- .
- II. National Water Research Institute (Canada)
- III. Series.

TD195.P37N37 2005 676'.042'0971 C2005-980047-X

Internet:

En40-237/5-2005E-PDF
0-662-39407-0

En40-237/5-2005E-HTML
0-662-39408-9

This report may be cited as follows:

Lowell, R.B., B. Ring, G. Pastershank, S. Walker, L. Trudel and K. Hedley. 2005. National Assessment of Pulp and Paper Environmental Effects Monitoring Data: Findings from Cycles 1 through 3. National Water Research Institute, Burlington, Ontario. NWRI Scientific Assessment Report Series No. 5. 40 p.

NWRI Scientific Assessment Report Series

The National Water Research Institute

Environment Canada's National Water Research Institute (NWRI) is Canada's largest freshwater research facility, with five locations across the country: at the Canada Centre for Inland Waters, Burlington, Ontario and the National Hydrology Research Centre, Saskatoon, Saskatchewan; and in Gatineau, Québec; Fredericton, New Brunswick; and Victoria, British Columbia. From these locations, NWRI leads initiatives in all regions of Canada and internationally to protect and sustain the quality and quantity of Canadian waters.

NWRI Scientific Assessment Report Series

On behalf of the Canadian government, NWRI leads and publishes scientific assessments of priority freshwater issues in Canada. Typically, these assessment reports are written by experts from NWRI, government, universities, and industry, and they encapsulate current scientific knowledge, trends, and information and program needs. As authoritative documents, they are intended to assist water science decision-makers, resource managers, and the research community in setting research priorities and in developing sound management policies and practices.

This report and previous ones in the series are available at **www.nwri.ca**

No. 1 – Threats to Sources of Drinking Water and Aquatic Ecosystem Health in Canada (2001)

No. 2 – National Assessment of Pulp and Paper Environmental Effects Monitoring Data (2003)

No. 3 – Threats to Water Availability in Canada (2004)

No. 4 – A Decade of Research on the Environmental Impacts of Pulp and Paper Mill Effluents in Canada (1992–2002) (2004)

Acknowledgement

We would like to particularly thank the following persons for reviewing various drafts of this report and for their suggestions for future enhancements to EEM data assessment: Peter Chapman (EVS Environment Consultants, Vancouver), Susan Jobling (Beyond the Basics Ltd., U.K.), Kelly Munkittrick (University of New Brunswick, Saint John), Keith Somers (Dorset Environmental Science Centre, Ontario), Nardia Ali and Wesley Plant (Environment Canada, Ontario Region), and Raymond Chabot and Isabelle Matteau (Environment Canada, Quebec Region). We are also grateful to the members of the EEM National Team and Science Committee for providing comments, background information, and insights throughout this study. The extensive analyses summarized in this report benefited from the able assistance of Gilles Champagne (Environment Canada, Gatineau) and Isaac Wong (Environment Canada, Burlington). We also thank Sylvie Desjardins (Environment Canada, Gatineau) and Marla Sheffer (Ottawa) for their help with editing the report.

Abstract

Under the *Fisheries Act*, Canadian pulp and paper mills are required to conduct Environmental Effects Monitoring (EEM) to assess effects potentially caused by their effluents. This report presents the results of the national assessment of EEM data collected from the receiving environments of pulp and paper mills across Canada for Cycles 1 through 3 (covering the years 1992-2004), with particular emphasis on the last four years (Cycle 3). The field survey response patterns observed for effluent-exposed fish and benthic invertebrates have, for the most part, shown a high degree of consistency through time over the last two cycles of data collection. The national average response pattern measured for fish in both Cycles 2 and 3 was one typically associated with nutrient enrichment overlaid by metabolic disruption. That is, exposed fish have consistently shown evidence of increased food availability or increased food absorption (fatter, faster growing, with larger livers) together with disruption of allocation of resources to reproduction (smaller gonads). The national average response for benthic invertebrate communities in both Cycles 2 and 3 was indicative of eutrophication, ranging from mild to more pronounced, partly depending on habitat type. More specifically, benthic invertebrate communities exposed to pulp mill effluent have commonly exhibited increases in abundance, together with some combination of increases, decreases or no change in taxon richness, depending on the degree of eutrophication. For both the fish and benthic invertebrate field surveys, other response patterns were also observed at some mills; furthermore, some possible shifts in patterns of effects were observed between Cycles 2 and 3, but these shifts were mostly relatively small. The results obtained in Cycle 4 and beyond, together with more focused studies at some mills, will help to address the biological significance (if any) of these shifts. The sublethal toxicity data showed clear improvements in effluent quality from Cycle 1 to Cycle 2, with, for the most part, no further changes in effluent quality in Cycle 3. Based on critical effect sizes developed for the fish and benthic invertebrate field survey endpoints, initial estimates are that approximately 20% of the monitored mills are expected to conduct only sublethal toxicity testing in Cycle 4, while roughly another 20% are expected to progress to determinations of the extent and magnitude of effects or, for some mills, investigations of the causes of the effects. Previous to the EEM program, a broad view of pulp and paper mill effluent effects on aquatic biota was not available at a national level in Canada (or, at this geographic scale, in other jurisdictions). The extensive database generated by the EEM program, and the subsequent analyses, have provided a fairly robust picture of the effects of these effluents on receiving waters across the country.

Résumé analytique

Aux termes de la *Loi sur les pêches*, les usines canadiennes de pâtes et papiers sont tenues de réaliser des études de suivi des effets sur l'environnement (ESEE) afin d'évaluer les effets potentiels de leurs effluents. Le présent rapport présente les résultats de l'évaluation nationale des données des ESEE recueillies dans les milieux récepteurs des usines canadiennes de pâtes et papiers pendant les cycles 1 à 3 (entre 1992 et 2004) et plus particulièrement pendant les quatre dernières années (cycle 3). Les études de terrain sur les poissons et les invertébrés benthiques exposés aux effluents montrent, pour la plupart, un haut niveau de cohérence pendant les deux derniers cycles de collecte des données. La répartition nationale des réponses moyennes des poissons pendant les cycles 2 et 3 révèle une réponse typiquement associée à un enrichissement en éléments nutritifs doublée d'une perturbation métabolique. En effet, les poissons exposés montrent des signes très clairs d'une augmentation de la disponibilité de la nourriture ou d'un accroissement de l'absorption des aliments (poissons plus gras, croissance plus rapide, foie plus gros) ainsi que d'une perturbation de l'allocation des ressources à la reproduction (gonades plus petites). La réponse nationale moyenne des communautés d'invertébrés benthiques pendant les cycles 2 et 3 indique une eutrophisation plus ou moins prononcée selon le type d'habitat. Plus précisément, les communautés d'invertébrés benthiques exposées aux effluents des usines de pâte sont généralement plus abondantes et montrent une richesse taxonomique croissante, décroissante ou stable selon le niveau d'eutrophisation. Les études sur les poissons et les invertébrés benthiques ont permis d'observer d'autres types de réponses à certaines usines. De plus, on a relevé de possibles changements de réponse entre les cycles 2 et 3, mais ces changements sont pour la plupart relativement faibles. Les résultats des cycles 4 et suivants, ainsi que des études plus spécifiques entreprises à certaines usines, permettront d'évaluer l'importance biologique de ces changements. Les données sur la toxicité sublétales indiquent une nette amélioration de la qualité des effluents entre les cycles 1 et 2, suivie, dans la plupart des cas, d'une stabilisation de la qualité des effluents au cycle 3. En nous basant sur les seuils critiques d'effet établis pour les mesures terminales des études sur les poissons et les invertébrés benthiques, nous estimons qu'au cycle 4, environ 20 % des usines surveillées n'auront à effectuer que des essais de toxicité sublétales alors que 20 % devront déterminer l'ampleur et la portée des effets ou, dans certains cas, enquêter sur les causes de ces derniers. Avant le programme d'ESEE, nous ne disposions pas d'une vue d'ensemble, à l'échelle du Canada (ou d'un territoire de taille comparable), des effets sur le biote aquatique des effluents des usines de pâtes et papiers. La riche base de données générée par le programme d'ESEE et par les analyses subséquentes nous a permis d'obtenir une image solide des effets de ces effluents sur les eaux réceptrices du pays.

Executive Summary

Under the *Fisheries Act*, the *Regulations Amending the Pulp and Paper Effluent Regulations* require Canadian pulp and paper mills to conduct Environmental Effects Monitoring (EEM) to assess effects potentially caused by their effluent. These studies generally include some or all of the following key components: a fish population field survey to assess effects on fish, a benthic invertebrate community field survey to assess effects on fish habitat, and studies to assess effects on the usability of fisheries resources, including a study of dioxins and furans in fish tissue and a tainting study. In addition, sublethal toxicity testing is conducted to aid in assessing effluent quality.

The EEM program is structured into “cycles,” whereby a mill conducts an EEM study every three to six years. The purpose of this report is to present and discuss the results of the national assessment of EEM data collected from the receiving environments of pulp and paper mills across Canada for Cycle 3, which was conducted from April 1, 2000 to April 1, 2004. In addition, these findings are compared to those from earlier cycles.

Two complementary quantitative approaches were used to provide an assessment of the overall effects of pulp and paper mill effluents on aquatic biota: 1) tabulation of the results of individual mill comparisons and 2) meta-analyses to investigate national patterns of effects. The response patterns observed for effluent-exposed fish and benthic invertebrates during Cycle 3 were, for the most part, quite similar to those observed during Cycle 2.

The national average response pattern measured for fish in both Cycles 2 and 3 was one typically associated with nutrient enrichment overlaid by metabolic disruption. That is, exposed fish have consistently shown evidence of increased food availability or increased food absorption (fatter, faster growing, with larger livers) together with disruption of allocation of resources to reproduction (smaller gonads), in comparison to reference area fish. Further, at a national level, the reduction in fish gonad size has remained virtually unchanged over two cycles of data collection. This metabolic disruption may include some aspect of endocrine disruption associated with problems in producing sufficient sex steroid hormones. Other observed response patterns for fish have included nutrient enrichment without measurable metabolic disruption, nutrient limitation, and chemical toxicity. Tainting was confirmed at one mill and dioxins and furans exceeded fish tissue guideline levels at three mills.

The national average response for benthic invertebrate communities in both Cycles 2 and 3 was indicative of eutrophication, ranging from mild to more pronounced, partly depending on habitat type. More specifically, benthic invertebrate communities exposed to pulp mill effluent have commonly exhibited increases in abundance, together with some combination of increases, decreases or no change in taxon richness, depending on the degree of eutrophication. Other observed benthic invertebrate response patterns have included toxicity or smothering effects.

The sublethal toxicity data showed clear improvements in effluent quality from Cycle 1 to Cycle 2, with, for the most part, no further changes in effluent quality in Cycle 3. It should be noted, however, that the sublethal toxicity reporting methodology and selection of tests do not track all aspects of effluent quality (and some tests showed low sensitivity to effluent exposure). In particular, they do not currently measure nutrient enrichment effects of the effluent, which was the most commonly observed kind of effect measured in the fish and benthic invertebrate field surveys, although the tests could be modified to provide some information on nutrient enrichment effects. Also of note, and consistent with the reductions in gonad size measured in the fish field survey, the most sensitive sublethal toxicity tests were those that measured a reproductive endpoint (although these did not include fish tests).

Although the fish and benthic invertebrate responses measured in the field surveys were highly consistent between Cycles 2 and 3 at a national level, some possible shifts in patterns of effects were also observed. The Cycle 3 fish data showed an overall lessening of the increased growth and condition observed in Cycle 2. On a habitat-specific basis, the benthic invertebrate data showed evidence of a possible lessening of toxicity/smothering effects in marine habitats, as well as more pronounced eutrophication in freshwater depositional habitats. The causes of these possible shifts are unknown at present, and it is possible that they are related to other factors such as background temporal variability in the receiving environment or changes in study designs between cycles. The results obtained in Cycle 4 and beyond, together with more focused studies at some mills, will help to answer these questions.

Because not all statistically significant effects are necessarily considered serious, Environment Canada has developed critical effect sizes (CES) for the fish and benthic invertebrate endpoints. These CES are used to help identify those differences that could be important and where more information (e.g., extent and magnitude of effects) is desirable in order to better understand the ecological significance of these differences. The use of these CES, together with the magnitudes and statistical significance of effects measured during Cycles 2 and 3, has identified a subset of mills where monitoring efforts can be reduced or where more extensive monitoring should be conducted. Finer-resolution evaluations will be made at the regional level, but initial estimates are that approximately 20% of the monitored mills are expected to conduct only sublethal toxicity testing in Cycle 4, while roughly another 20% are expected to progress to determinations of the extent and magnitude of effects or, for some mills, investigations of the causes of the effects.

Previous to the EEM program, a broad view of pulp and paper mill effluent effects on aquatic biota was not available at a national level in Canada (or, at this geographic scale, in other jurisdictions). The extensive database generated by the EEM program, and the subsequent analyses, have provided a fairly robust picture of the effects of these effluents on receiving waters across the country.

Table of Contents

Executive Summary	v
1. Introduction.....	1
1.1 <i>The National EEM Program.....</i>	<i>1</i>
1.2 <i>Objectives of the Report.....</i>	<i>2</i>
1.3 <i>Effect Indicators and Critical Effect Size.....</i>	<i>2</i>
2. Overview of Studies Conducted in Cycle 3.....	4
3. General Methods.....	5
3.1 <i>Data Preparation and Analysis</i>	<i>5</i>
3.2 <i>Procedure for Determining National Response Patterns</i>	<i>6</i>
4. Fish Survey	7
4.1 <i>Data Processing and Study Designs.....</i>	<i>8</i>
4.2 <i>Summary of Effect Sizes</i>	<i>11</i>
4.3 <i>Response Patterns and Meta-analyses.....</i>	<i>15</i>
5. Fisheries Resources and Usability	19
6. Benthic Invertebrate Community Survey	19
6.1 <i>Data Processing and Study Designs.....</i>	<i>19</i>
6.2 <i>Summary of Effect Sizes</i>	<i>20</i>
6.3 <i>Response Patterns and Meta-analyses.....</i>	<i>23</i>
7. Sublethal Toxicity Testing.....	27
7.1 <i>Introduction.....</i>	<i>27</i>
7.2 <i>Monitoring Changes in Effluent Quality Among Cycles.....</i>	<i>28</i>
7.3 <i>Summary and Future Considerations</i>	<i>32</i>
8. Summary and Conclusions.....	32
9. Glossary	35

10.	Acronyms / Abbreviations.....	37
11.	References.....	38

List of Tables

Table 1:	Fish population and benthic invertebrate community survey endpoints and their respective critical effect sizes (CES) for EEM for pulp and paper mills.....	3
Table 2:	General summary of Cycle 3 pulp and paper EEM studies.....	4
Table 3:	Number of mills that encountered problems with fish field surveys in Cycles 2 and 3.....	8
Table 4:	Distribution list of freshwater and marine/estuarine species used over three cycles ^a	9

List of Figures

Figure 1:	Example of a meta-analysis summary figure. The effect size was measured as either Hedges' d or Fisher's z-transform(see text).	7
Figure 2:	Distribution of measured percent differences between exposure and reference area fish in Cycles 2 and 3	12
Figure 3:	Number of exposure versus reference fish comparisons showing no significant difference	14
Figure 4:	Grand means for five key fish endpoints for Cycles 2 (C2) and 3 (C3).	16
Figure 5:	Five key fish endpoints, by gender, for Cycles 2 (C2) and 3 (C3). M = male, F = female. Error bars represent 95% confidence intervals.....	17
Figure 6:	Five key endpoints, by species, for Cycle 3. Error bars represent 95% confidence intervals.	18
Figure 7:	Distribution of measured percent differences between exposure and reference areas for the benthic invertebrate survey (control/impact designs only) for a) abundance, b) taxon richness, c) Bray-Curtis and d) Simpson's evenness.....	21
Figure 8:	Number of mills showing no significant difference, a significant difference less than the CES of $\pm 2SD$ and a significant difference greater than the CES of $\pm 2SD$	22
Figure 9:	Predicted response patterns for benthic invertebrate communities (Lowell <i>et al.</i> 2003).	24
Figure 10:	Control/impact grand means for Cycles 2 (n = 62) and 3 (n = 55). Error bars represent 95% confidence intervals.....	25
Figure 11:	Abundance and taxon richness, by habitat, for Cycles 2 (C2) and 3 (C3).26	
Figure 12:	Comparison of sublethal toxicity to freshwater species in Cycles 1, 2 and 3. Note that for the rainbow trout test, the EC ₂₅ is the effective concentration for 25% of the embryos	29
Figure 13:	Comparison of sublethal toxicity to marine species in Cycles 1, 2 and 3.....	30
Figure 14:	Percentages of sublethal tests showing no effect at highest concentration31	

1. Introduction

1.1 *The National EEM Program*

Under the *Fisheries Act*, the 1992 *Pulp and Paper Effluent Regulations* (PPER) prescribe discharge limits for total suspended solids and biochemical oxygen demand, and require effluent to be non-acutely lethal to fish. These limits reflect what secondary treatment of effluents can achieve and provide a national baseline standard that is intended to protect fish, fish habitat and the use of fisheries resources. At the time the regulations were developed, it was acknowledged that there was uncertainty about the effectiveness of the new limits for protecting the diverse variety of aquatic environments receiving pulp and paper mill effluents in Canada.

In order to assess the adequacy of the effluent regulations for protecting the aquatic environment, Environmental Effects Monitoring (EEM) was included as a requirement in the 1992 PPER. In May 2004, the *Regulations Amending the Pulp and Paper Effluent Regulations* (RAPPER) came into force and further clarified the requirements of the 1992 PPER. Canadian pulp and paper mills that are subject to the RAPPER are required to conduct studies on their receiving environments in order to assess and monitor effects potentially caused by their effluent. These studies may include the following components:

- a fish population survey to assess fish health;
- a benthic invertebrate community survey to assess effects on fish habitat; and
- studies to assess effects on the usability of fisheries resources, including a study of dioxins and furans in fish tissue and a tainting study.

In addition, sublethal toxicity testing and measurements of supporting water and sediment quality variables are conducted to contribute to the program in areas such as interpretation of biological data and assessing effluent quality.

The EEM program is structured into “cycles,” whereby a mill conducts an EEM study every three to six years with both monitoring and interpretation phases. At the beginning of each cycle, each mill is required to develop a site-specific study design in collaboration with Environment Canada regional staff. At the end of each cycle, each mill must submit an interpretative report that summarizes its monitoring results. The EEM program uses a tiered approach to monitoring, with initial studies carried out to characterize and assess the condition of the receiving environment followed by targeted or focused studies to determine the extent and magnitude of effects, where effects are detected and confirmed, or a reduced level of monitoring, where effects are not found. Technical guidance is developed by Environment Canada on all aspects of EEM studies, including design, analyses and interpretation of data. Additional information on the EEM program is available at <http://www.ec.gc.ca/eem/>.

Currently, most mills have completed their third cycle of monitoring and reporting. Cycle 1 (completed in 1996) was used primarily as a baseline to gain better

understanding of the variability of the field measurements, to identify problems and to provide recommendations for future cycles. The second cycle results showed that mills were successful in reducing the toxicity of their effluent and that their effluent quality had vastly improved since the promulgation of the 1992 PPER (many Canadian mills upgraded their effluent treatment systems just prior to Cycle 2). The national assessment of biological monitoring data for Cycle 2 EEM (Lowell *et al.* 2003, 2004) showed that pulp and paper mill effluents continued to affect fish health and habitat. The predominant response patterns seen for fish were a decrease in gonad weight and increases in liver weight, condition and weight at age. These response patterns are believed to be indicative of some form of metabolic disruption or impairment of endocrine functioning in combination with a nutrient enrichment effect. The national average response pattern for the benthic invertebrate community surveys in Cycle 2 was mild to moderate eutrophication.

1.2 *Objectives of the Report*

The purpose of this report is to present and discuss the results of the national assessment of environmental effects monitoring (EEM) data collected from the receiving environments of pulp and paper mills across Canada for Cycle 3, which was conducted from April 1, 2000, to April 1, 2004. The data analyses focused on the following questions:

- 1) What are the types and magnitude of effects of pulp and paper mill effluents on adult fish and benthic invertebrate communities?
- 2) How are these effects influenced by habitat types (for benthic invertebrates), species and gender (for fish)?
- 3) How does Cycle 3 compare with Cycle 2 in terms of magnitudes of effects and types of effects?
- 4) How do measured effects compare with predefined critical effect sizes established to help determine which mills are having the most serious impacts?

In addition to the above analyses, sublethal toxicity data were analyzed to assess changes in effluent quality over three cycles of data. The effects of pulp and paper mill effluents on the use of fisheries resources (levels of dioxins and furans in fish tissue and tainting) were also examined.

1.3 *Effect Indicators and Critical Effect Size*

In EEM, an effect is defined as a statistically significant difference in one of the measurements taken in fish (fish survey) or benthic invertebrates (benthic invertebrate community survey) from the exposure area relative to those from a reference area or along a gradient of effluent exposure (Lowell *et al.* 2002). Cycle 2 data for pulp and paper mills showed that the vast majority of mills reported a statistically significant effect in at least one of the indicator measurements. Because not all statistically significant effects are considered serious, Environment Canada has developed critical effect sizes (CES) for fish and benthos (Environment Canada 2004b). These CES are used to help identify those differences that could be important and where more information (e.g.,

extent and magnitude of effects) is desirable in order to better understand the ecological significance of these differences. For fish, the CES are based on 1) the magnitude of observed pulp mill effluent effects that were previously demonstrated in Canada and Sweden, 2) natural variations typically observed and 3) the magnitude of effects observed at the worst mills in Cycle 2 (Environment Canada 2004b, Lowell *et al.* 2003). For benthic communities, the CES are also based on the magnitude of effects measured during Cycle 2, as well as the concept that effects exceeding the “normal range” of variability in reference areas are important (Lowell 1997, Kilgour *et al.* 1998, Lowell *et al.* 2003). The indicator endpoints and CES used to determine effects on fish and benthic invertebrates are described in Table 1; background on these endpoints including calculation methods, is available in Environment Canada (2004a).

Table 1: Fish population and benthic invertebrate community survey endpoints and their respective critical effect sizes (CES) for EEM for pulp and paper mills.

Endpoints	Endpoint provides information on:	Critical effect size ^a
<i>Fish population survey</i>		
Weight of gonads relative to body weight	Reproduction	± 25%
Liver weight relative to body weight	Condition	± 25%
Body weight relative to body length	Condition	± 10%
Age	Survival	b
Body weight relative to age	Growth	b
<i>Benthic invertebrate survey</i>		
Total abundance	Number of animals	± 2SD
Taxon richness	Number of taxa or kinds of animals	± 2SD
Simpson's evenness	Measure of how evenly the animals are distributed among the taxa	± 2SD
Bray-Curtis index	Measure of dissimilarity in community composition among sites	+ 2SD

^a CES for fish endpoints are expressed as percent (%) difference of exposure relative to reference data. CES for the benthic invertebrate surveys apply only to control/impact designs and are expressed as within ±2 standard deviations (SD) of the reference data. For gradient designs, a significant correlation in two consecutive cycles is used in place of the CES. More information on CES can be obtained from Lowell (1997) and Lowell *et al.* (2002, 2003), Environment Canada 2004 a and b.

^b CES for these age-related endpoints have not been developed due to uncertainties in techniques used to age some species of fish.

In this national assessment report, the distribution of measured effect sizes at mills across Canada are presented and the measured effects are compared with the CES in Table 1. At the individual mill level, the EEM program uses exceedance of CES (when confirmed in two consecutive cycles) to help identify those mills expected to conduct more focused monitoring to assess the extent and magnitude of effects or to investigate the cause of the observed effects at a site. At the level of this national assessment, however, the use of CES is not to address the need for any particular mill to conduct more extensive monitoring, but rather to help to estimate the total number of mills that

will be conducting more extensive monitoring in Cycle 4. As more information becomes available, the use of the CES in the EEM program will evolve.

2. Overview of Studies Conducted in Cycle 3

In Cycle 3, 112 mills conducted EEM studies, including 29 marine/estuarine and 66 freshwater field surveys. Table 2 shows a regional summary of the number and type of field surveys as well as the sublethal toxicity studies conducted by mills for Cycle 3. There were 62 standard fish surveys, which included 4 studies using molluscs, and 87 standard benthic invertebrate surveys; all mills conducted sublethal toxicity studies. Very few mills used alternatives for fish and benthic surveys; these included caged bivalves, mesocosms and other research. In total, 24 mills were exempted from conducting a fish survey because the effluent concentration in the exposure area was less than 1% within 250 m of the point of discharge. In addition, conditions at numerous sites made conducting the surveys unfeasible for either the fish survey or the benthic invertebrate survey. At these mills, site-specific exemptions were given for various reasons, such as variability of habitat, extreme tides, unsafe conditions for sampling and lack of suitable reference areas. Due to these exemptions, there were 11 mills operating in Cycle 3 that did not conduct field surveys and submitted only sublethal toxicity data.

Table 2: General summary of Cycle 3 pulp and paper EEM studies.

Region	No. of mills that conducted EEM studies ^a	Fish surveys			Benthic invertebrate surveys		No. of mills that used alternatives	No. of mills that conducted only sublethal toxicity tests
		No. of standard surveys	No. of dilution exemptions	No. of site-specific exemptions	No. of standard surveys	No. of site-specific exemptions		
Atlantic	19	8	5	3	16	2	3 (fish research & benthic mesocosm; fish research; caged bivalve)	0
Quebec	42 ^b	29	10	0	31	8	0	8
Ontario	20	14	1	4	17	2	1 (1 fish mesocosm)	1
Prairie and Northern	10	5	4	0	9	0	1 (fish and benthic mesocosms)	0
Pacific and Yukon	21	6	4	7 ^c	14	2	0	2
Total	112	62^d	24	14	87^e	14	5	11

^a There were an additional 10 mills that were non-operational for a long period of time and did not conduct an EEM study in Cycle 3.

^b There are three mills in the Quebec Region that are on a different monitoring timeline; therefore, their data for Cycle 3 are not available at this time.

^c Four mills from the Pacific and Yukon Region conducted extended sublethal toxicity tests on topsmelt (growth and survival).

^d Includes 1 joint study in the Pacific and Yukon Region. A joint study refers to a study that was completed concurrently by more than one mill that discharge in the same water body and share sampling areas.

^e Includes a total of 4 joint studies, three in Pacific and Yukon Region and one in the Ontario Region.

3. General Methods

3.1 Data Preparation and Analysis

This section briefly describes the general methodology used to carry out the national assessment of data from the benthic invertebrate and fish surveys conducted in Cycle 3, which, for the most part, is the same as that used in the Cycle 2 national assessment (see Lowell *et al.* 2003 for further details).

Two complementary quantitative approaches were used to synthesize the large number of independent studies to provide a national overview of the effects of pulp mill effluents: 1) tabulation of the results of individual mill comparisons and 2) meta-analyses. The tabulations, which use an easy-to-understand format (e.g., number of studies with significant effect at the level of the individual mill), are useful to illustrate the main findings of the assessment in histogram figures in the report. One limitation of this approach is that the significance level is dependent not only on the magnitude of difference between an exposure and reference area, but also on the sample size, which can limit the ability to detect some effects that may be occurring. Therefore, we also used meta-analyses, which do not have the same limitations as individual study tabulations. Meta-analysis is a technique used to statistically examine the magnitude of effects in a way that loses less information due to the constraints of individual study sample sizes and scale of measurement (Hedges and Olkin 1985, Rosenberg *et al.* 2000, Gurevitch and Hedges, 2001). This latter approach treats the separate studies essentially as replicates; as such, it is possible to look at questions that are difficult to examine at the individual mill level (e.g., the influence of fish gender or species or habitat type on effluent effects in the field). A full description of how meta-analysis was used for the Cycle 2 national assessment can be found in Lowell *et al.* (2003).

There are two main categories of sampling designs for the benthic invertebrate community survey: control/impact (sampling stations located in reference areas and exposure areas; this was the most commonly used design for the invertebrate survey in Cycles 1 through 3) and gradient (sampling stations located along an exposure gradient). For the fish survey, only the control/impact sampling design is used. Further information on EEM study designs and respective analyses for the fish and benthic invertebrate surveys is provided in Glozier *et al.* (2002), Lowell *et al.* (2002, 2003) and the Technical Guidance Document (Environment Canada 2004a).

The national assessment focused on near-field effects in order to investigate the more pronounced effects that are occurring nationally for the control/impact benthic invertebrate and fish surveys. Some mills collected data from multiple areas (e.g., multiple near-field areas). Data from more than one area were pooled only if warranted based on inspection of the interpretative reports. A statistical assessment tool (SAT) developed by the National Water Research Institute of Environment Canada was used to calculate the magnitude and statistical significance of effects for the five key fish and four key benthic invertebrate endpoints (control/impact study design only). For gradient-

design benthic invertebrate studies, regression analysis was used to analyze changes in the benthic invertebrate endpoints with increasing distance from the effluent outfall.

The primary steps common to both the invertebrate and fish analyses were as follows. First, electronic data were screened for obvious errors (missing data fields, obvious data entry errors, etc.). SAT aided with the selection of the appropriate data for analysis, including removal of outliers. Finally, SAT or a spreadsheet were used to analyze the data via ANOVA (Analysis of Variance), ANCOVA (Analysis of Covariance), or regression analysis, depending on the study design, to determine the relationship between exposure and reference areas (or correlation coefficient for gradient designs) for each of the endpoint values for each mill. The ANOVA and ANCOVA analyses provided area means, standard deviations and the magnitudes of effects, which were required for subsequent tabulations and meta-analyses of measured effects for the control/impact-type designs. The regression analyses provided correlation coefficients for tabulations and meta-analyses of the remaining gradient benthic invertebrate designs. The significance level (α) used for the ANOVA, ANCOVA, and regression analyses was set at 0.05 for the purposes of the tabulations presented here.

The fish results presented here were derived from log-transformed data, and the invertebrate results were derived from non-transformed data. The one exception was the fish age results, which were derived from log age+1 data, for backwards compatibility. For comparative purposes, both the fish age data and the invertebrate data were also analyzed using the log transformation (not adding 1), and the fish age data were further analyzed using no transformation (not presented here). The meta-analyses revealed that the chosen transformation (or lack of) had virtually no effect on the results at a national scale. It should be noted, however, that choice of transformation can sometimes have important consequences at the level of the individual mill, and the transformation should be chosen accordingly for mill-specific evaluations. In particular, it is recommended that the log endpoint+1 transformation not be used in future individual mill analyses; see the Technical Guidance Document (Environment Canada, 2004a) for further information on choice of data transformation.

3.2 *Procedure for Determining National Response Patterns*

Meta-analysis is a set of statistical procedures used to quantitatively synthesize the results of a large number of independent studies. Further, it permits overall response patterns to be determined. The meta-analyses required determination of a standardized magnitude of effect. For the more commonly employed control/impact designs, the Hedges'd effect size was calculated as the difference between the exposure and reference means, divided by the pooled standard deviation (this value is multiplied by a correction factor that accounts for the effects of small sample sizes) (Rosenberg *et al.* 2000). When pooling control/impact and gradient design data, it was necessary to use a second, related effect size measure, the Fisher's z-transform (Rosenthal 1991, Rosenberg *et al.* 2000, Lowell *et al.* 2003).

The main meta-analytical results are presented in the following summary format (Fig. 1). The standardized effect size is on the x-axis, with the vertical line representing a zero effect. The result for each mill grouping is presented as a horizontal 95% confidence interval about a vertical tick mark indicating the average effect size for that grouping of mills. Mill distributions to the right of the zero effect line indicate that the average effect associated with effluent exposure was an increase in the measured endpoint. Similarly, mill distributions to the left of the zero effect line indicate an effluent-associated decrease in the measured endpoint. The increase or decrease is statistically significant for the group as a whole if the 95% confidence interval does not overlap the zero effect line. Larger mill groupings (that are non-significant as a whole) can be composed of smaller subgroups, some or all of which may be significantly different from zero. All of the meta-analysis results in the following sections will use this graphical representation of the data.

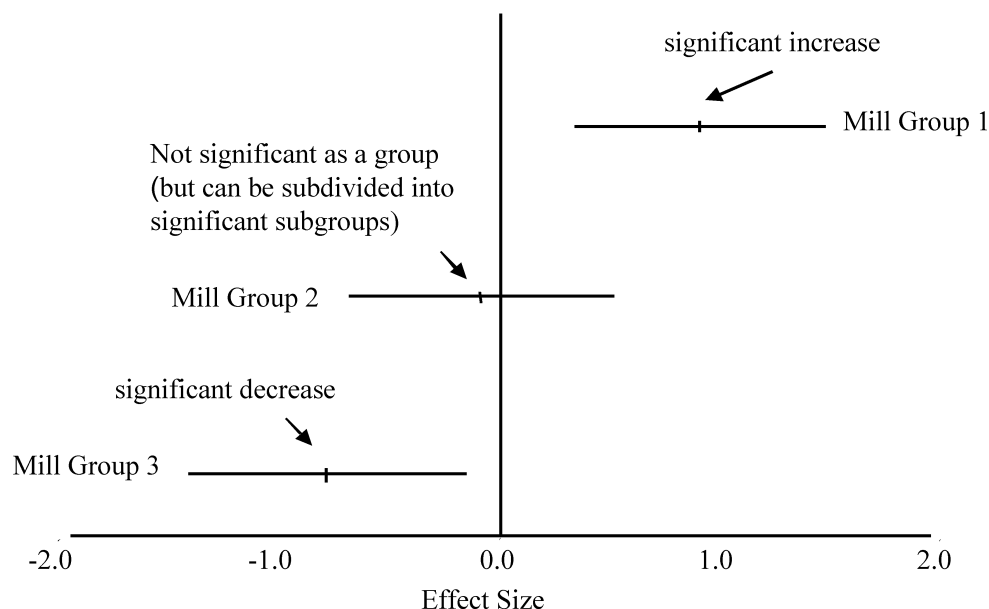


Figure 1: Example of a meta-analysis summary figure. The effect size was measured as either Hedges' d or Fisher's z -transform (see text).

4. Fish Survey

The adult fish survey is used to determine if the effluent is affecting fish populations by comparing effluent-exposed fish with those from reference areas. The survey uses fish growth, reproduction, condition and age structure to assess the overall health of exposed fish. These are assessed via measurements of five core fish endpoints: weight at age, relative gonad and liver weights, condition (body weight relative to length), and age (Table 1).

4.1 Data Processing and Study Designs

In Cycle 3, data for 62 fish surveys were submitted electronically. A total of 54 of these surveys contained adult fish data that had sufficient replication to conduct statistical analysis. A small number of alternative studies (e.g., caged bivalves, mesocosms) were also conducted (Table 2), which, due to their unique nature, are not included in the following summary of results. Prior to analysis, the electronically submitted fish data were screened for errors and incomplete data. In both Cycles 2 and 3, the majority of mills encountered one or more problems with portions of the fish field survey, and these are summarized in Table 3. The types and numbers of problems reported were similar between these two cycles.

Table 3: Number of mills that encountered problems with fish field surveys in Cycles 2 and 3.

Problems	Cycle 2	Cycle 3
Too few fish (<12 fish at exposure and/or reference sites)	28	24
Exposure site (i.e., fish not caught)	1	7
Reference site (i.e., fish not caught)	5	4
Only one sentinel species	33	27
Incomplete/poor reporting / <12 fish after removal of outliers and/or immatures	18	14
Fish immature	17	15
Aging problems / not aged	6	6

Overall, the majority of the submitted data were of good quality for both cycles, and it was possible to correct a number of the data submission problems manually. After accounting for problems with electronically submitted data, only six submitted fish studies in Cycle 2 and five submitted fish studies in Cycle 3 were omitted entirely due to the kinds of problems summarized in Table 3 (i.e., no valid fish comparisons available for the study). Note that a “comparison,” as used for the fish survey here and later in this report, refers to an exposure versus reference comparison for one fish species and gender. Thus, a mill that conducts a fish survey on two genders for two species would have data for four comparisons for each endpoint, assuming that all comparisons provided sufficient data.

Table 4 lists the fish species collected at all mills over three cycles. The two most commonly utilized species collected in Cycles 1 and 2 were the white and longnose suckers. The white sucker was also used frequently during Cycle 3. While this comparatively large species was the mostly commonly used, the proportion of small-bodied fish species increased substantially in Cycle 2, with this higher proportion carrying over to Cycle 3 (Table 4; Lowell *et al.* 2003). Increased use of small-bodied species has been encouraged due to concerns regarding large-bodied fish mobility and subsequent questions about movement of sampled fish in and out of exposure areas. It

should be noted, however, that recent research has shown that white sucker mobility is often minimal, except during spawning periods (Doherty *et al.* 2004).

Table 4: Distribution list of freshwater and marine/estuarine species used over three cycles^a.

Species	Scientific name	Cycle 1	Cycle 2	Cycle 3
Freshwater species				
<i>Large-bodied fish^b</i>				
White sucker	<i>Catostomus commersoni</i>	46	32	25
Longnose sucker	<i>Catostomus catostomus</i>	24	20	11
Walleye	<i>Stizostedion vitreum</i>	10	1	4
Yellow perch	<i>Perca flavescens</i>	7	9	6
Shorthead redhorse	<i>Moxostoma macrolepidotum</i>	8	1	2
Mountain whitefish	<i>Prosopium williamsoni</i>	5	1	1
Brown bullhead	<i>Ameiurus nebulosus</i>	4	3	4
Largescale sucker	<i>Catostomus macrocheilus</i>	4	3	1
Prickly sculpin	<i>Cottus asper</i>	2		2
Silver redhorse	<i>Moxostoma anisurum</i>	4	3	2
Common carp	<i>Cyprinus carpio</i>	2	1	1
Peamouth	<i>Mylocheilus caurinus</i>	2		
Rock bass	<i>Ambloplites rupestris</i>	2	2	7
Burbot	<i>Lota lota</i>	1	1	
Channel catfish	<i>Ictalurus punctatus</i>	1	1	
Goldeye	<i>Hiodon alosoides</i>	1		1
Northern pike	<i>Esox lucius</i>	1		
Smallmouth bass	<i>Micropterus dolomieu</i>	1	1	
Lake whitefish	<i>Coregonus clupeaformis</i>	1		
Fallfish	<i>Semotilus corporalis</i>		2	2
Bridgelip sucker	<i>Catostomus columbianus</i>			3
Total large-bodied fish		126	81	72
<i>Small-bodied fish^b</i>				
Spottail shiner	<i>Notropis hudsonius</i>	4	2	2
Pumpkinseed	<i>Lepomis gibbosus</i>	1	1	1
Lake chub	<i>Couesius plumbeus</i>	2	1	
Fathead minnow	<i>Pimephales promelas</i>	1		
Threespine stickleback	<i>Gasterosteus aculeatus</i>	1		
Johnny darter	<i>Etheostoma nigrum</i>		6	6
Longnose dace	<i>Rhinichthys cataractae</i>		5	5
Trout-perch	<i>Percopsis omiscomaycus</i>		4	1
Iowa darter	<i>Etheostoma exile</i>		2	

Species	Scientific name	Cycle 1	Cycle 2	Cycle 3
Logperch	<i>Percina caprodes</i>		1	
Bluntnose minnow	<i>Pimephales notatus</i>		1	
Blacknose dace	<i>Rhinichthys atratulus</i>		1	2
Common shiner	<i>Luxilus cornutus</i>		1	
Redside shiner	<i>Richardsonius balteatus</i>		1	
Mimic shiner	<i>Notropis volucellus</i>		1	1
Mottled sculpin	<i>Cottus bairdi</i>		1	2
Slimy sculpin	<i>Cottus cognatus</i>		1	
Spoonhead sculpin	<i>Cottus ricei</i>		1	1
Torrent sculpin	<i>Cottus rhotheus</i>			1
Pearl dace	<i>Margariscuss margarita</i>			1
Golden shiner	<i>Notemigonus crysoleucas</i>			1
Emerald shiner	<i>Notropis atherinoides</i>			1
Total small-bodied fish		9	30	25
Marine/estuarine species				
<i>Large-bodied fish^b</i>				
Winter flounder	<i>Pleuronectes americanus</i>	7	2	1
English sole	<i>Pleuronectes vetulus</i>	5		
Atlantic tomcod	<i>Microgadus tomcod</i>	5		
Starry flounder	<i>Platichthys stellatus</i>	2		
Longhorn sculpin	<i>Myoxocephalus octodecemspinosus</i>	2	2	1
Slender sole	<i>Eopsetta exilis</i>	1		
Cunner	<i>Tautoglabrus adspersus</i>	1		1
Widow rockfish	<i>Sebastes entomelas</i>	1		
Quillback rockfish	<i>Sebastes maliger</i>	1		
Rock sole	<i>Pleuronectes bilineatus</i>	1		
Shorthorn sculpin	<i>Myoxocephalus scorpius</i>		2	2
Total large-bodied fish		26	6	5
<i>Small-bodied fish^b</i>				
Atlantic silverside	<i>Menidia menidia</i>	1		
Threespine stickleback	<i>Gasterosteus aculeatus</i>	2	3	3
Mummichog	<i>Fundulus heteroclitus</i>	1	3	4
Rock gunnel	<i>Pholis gunnellus</i>			1
Total small-bodied fish		4	6	8
Crabs and molluscs				
Dungeness crab	<i>Cancer magister</i>	6		
Rock crab	<i>Cancer irroratus</i>	3		
Oyster	<i>Crassostrea gigas</i>	2	2	1
Blue mussel	<i>Mytilus edulis</i>	2	3	2
Green crab	<i>Carcinus maemus</i>	1		

Species	Scientific name	Cycle 1	Cycle 2	Cycle 3
Macoma	<i>Macoma balthica</i>		1	1
Waved whelk	<i>Buccinum undatum</i>		1	2
Total crabs and molluscs		14	7	6

^a The data in this table include information from both successful and unsuccessful studies (modified from Environment Canada 1997; Courtenay *et al.* 2002 ; Munkittrick *et al.* 2002 ; Lowell *et al.* 2003).

^b Small-bodied fish are defined here as a fish species that has a median size (as sampled during the EEM study) of 150 mm or less (see also Environment Canada 2004a).

4.2 Summary of Effect Sizes

The key fish comparisons summarized in this section focus on the endpoints for which Environment Canada has developed CES — gonad weight, liver weight and condition factor. The range and distribution of measured exposure versus reference area percent differences in Cycle 3 were quite similar to those in Cycle 2 (Fig. 2). For Figures 2 and 3, measured differences were calculated as exposure area minus reference area adjusted means, expressed as a percentage of the reference area adjusted mean. Figure 2 focuses on comparisons where exposure versus reference ANCOVA slopes were parallel (the majority of comparisons; see Fig. 3 and Environment Canada 2004a). Note also that Figure 2 provides a complete summary of the measured differences (i.e., includes both statistically significant and non-significant differences). Similar to Cycle 2, Cycle 3 condition factor showed the narrowest range in exposure versus reference area percent differences (-15% to 25%); this comparatively narrow range led to the development of a smaller CES for condition than for gonad and liver weights. Cycle 3 gonad weight percent differences ranged from approximately -70% to 70%, while liver weight differences ranged from -55% to 120%, again with a highly similar data distribution to that observed in Cycle 2. Thus, at a national level, we observed a high reproducibility between Cycles 2 and 3 in magnitude and distribution of measured differences. This is likely related to the consistency of response patterns from one cycle to the next that is discussed in section 4.3.

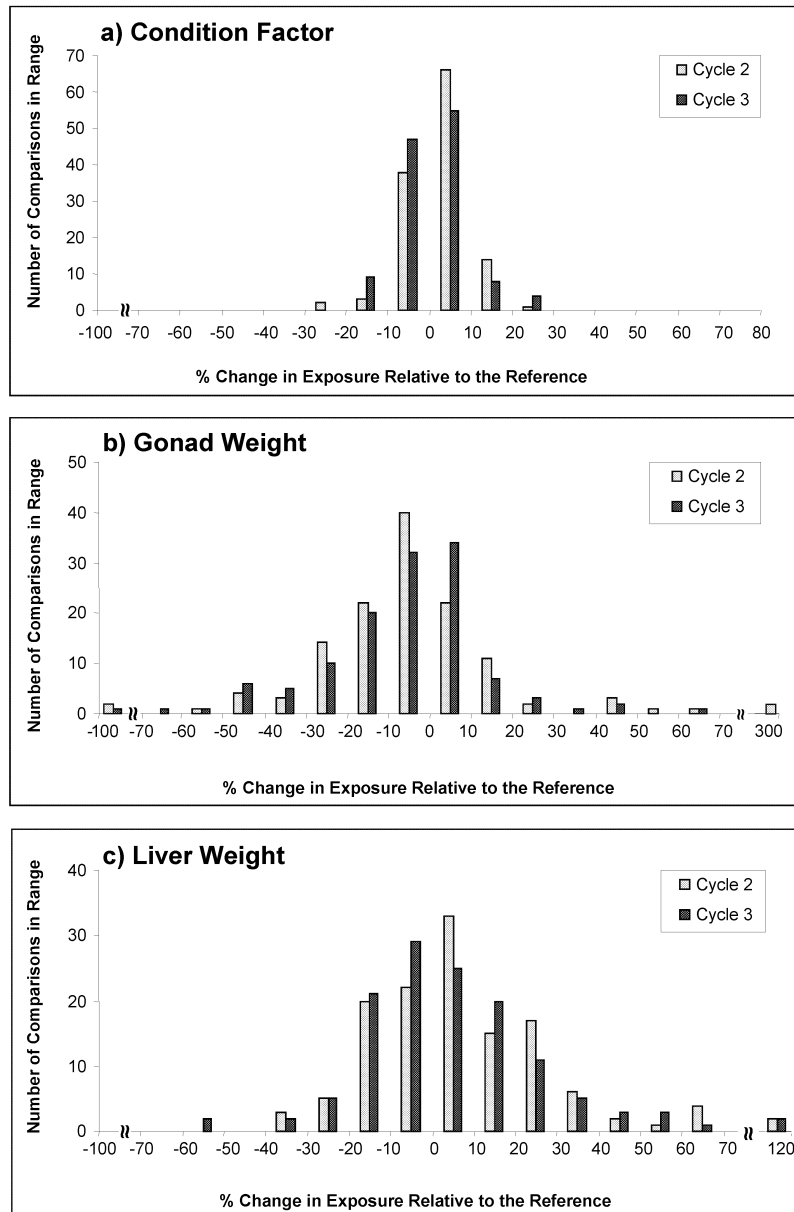


Figure 2: Distribution of measured percent differences between exposure and reference area fish in Cycles 2 and 3 for: a) condition factor, b) gonad weight and c) liver weight.

Figure 3 shows the number of comparisons that showed no statistically significant effect versus a significant effect for condition, gonad weight and liver weight. The significant effects are further divided into three categories. The first of these is a significant interaction. This occurs when the exposure versus reference area slopes are statistically different in the ANCOVA analysis; that is, when the slopes can be considered to be non-parallel. In this case, the effect of effluent exposure is to, for example, affect allocation to gonad weight differently for fish of different sizes, relative to fish in reference areas, resulting in non-parallel exposure versus reference slopes for ANCOVA regressions of gonad weight against body weight. The last two significant effect categories in Figure 3 occur only when the exposure and reference area slopes are not significantly different (no significant interaction, so effect of effluent exposure is approximately the same for fish of different sizes). The first of these last two categories consists of comparisons where the exposure versus reference area ANCOVA adjusted means were significantly different, but with an effect magnitude smaller than the CES. The second of these last two categories consists of comparisons where the exposure versus reference area adjusted means were significantly different, with an effect magnitude that exceeded the CES. Note that CES have not yet been developed for the case where ANCOVA slopes are not parallel. See Environment Canada (2004a) for further information on ANCOVA procedures and interpretation.

Summing the comparisons illustrated in Figure 3 reveals that between 40% and 60% of the comparisons for each endpoint were significant. Furthermore, 20–25 of the significant comparisons for each endpoint exceeded the CES. Thus, the number of effects that exceed the CES is relatively consistent among endpoints.

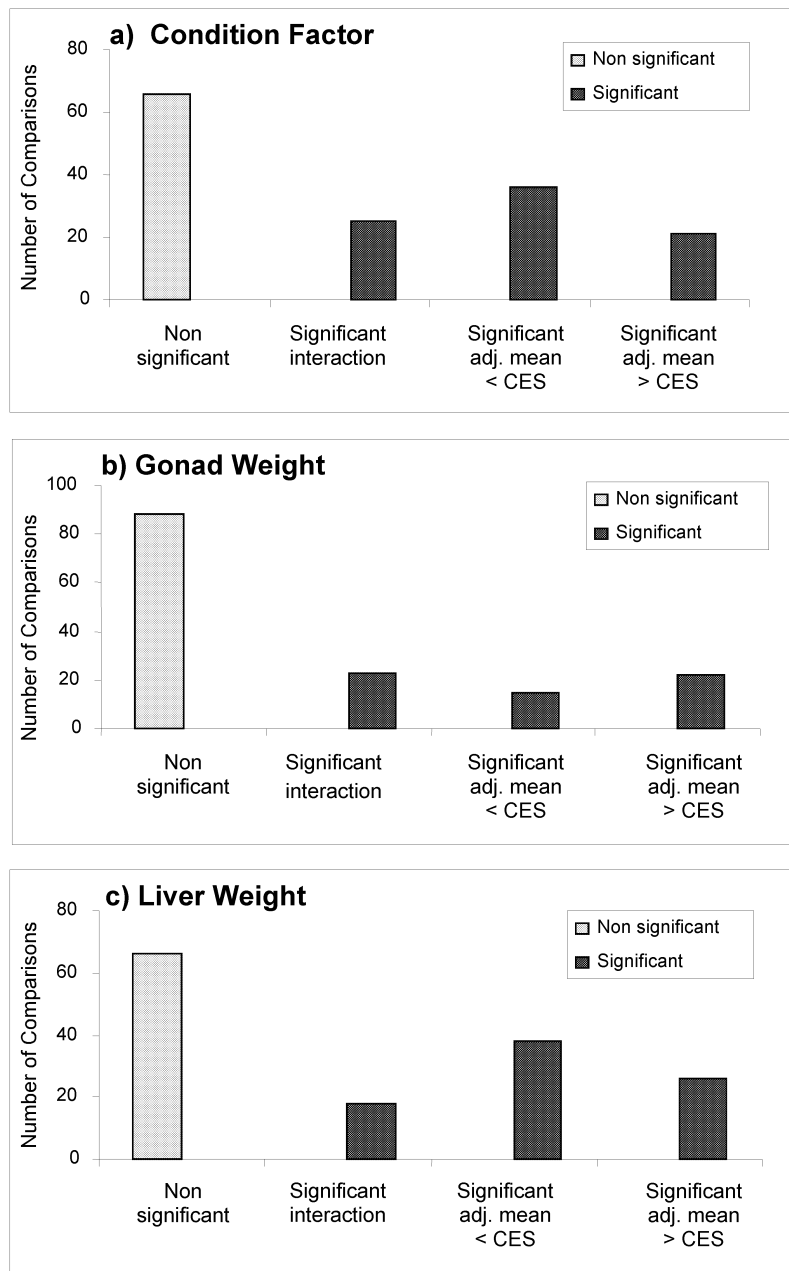


Figure 3: Number of exposure versus reference fish comparisons showing no significant difference a significant interaction, a significant difference in adjusted means less than CES and a significant difference in adjusted means greater than CES for a) condition factor, b) gonad weight and c) liver weight.

For fish studies, a mill will conduct focused monitoring (including magnitude and geographic extent) if any CES is exceeded, with statistical significance, in the same direction for the same sex and species for two consecutive cycles. There was this kind of exceedance of CES in Cycles 2 and 3 at approximately 15% of the mills conducting a fish survey. Of these latter mills showing consistently large effects in both cycles, approximately three-quarters exceeded the CES for condition factor, one-half for gonad weight and one-quarter for liver weight. Half of these mills exceeded the CES for two endpoints, while the other half exceeded the CES for only one endpoint. For mills expected to conduct focused monitoring, condition and liver size were usually greater in the exposure areas (a sign of nutrient enrichment), while gonad size was usually smaller in the exposure areas (a sign of disruption of resource allocation to the gonads; see section 4.3).

4.3 *Response Patterns and Meta-analyses*

Three main response patterns were observed in the Cycle 2 EEM fish survey (Lowell *et al.* 2003). These responses, and the changes in the EEM endpoints associated with them, have been widely described in the literature (see Munkittrick *et al.* 1991, 1994, 2000 for reviews). It should be noted, however, that other response patterns may also occur at some mills. The first of the three main patterns, nutrient enrichment, was generally associated with increases in gonad and liver weight, as well as increases in condition and often growth rate (weight at age). The second main pattern, nutrient limitation together with chemical toxicity or other inhibitory effects, was associated with decreases in these endpoints. Nutrient limitation is defined broadly here to include some combination of limited availability of food, appetite suppression and/or internal alteration of food absorption (leading to decreases in several endpoints). Factors other than nutrient availability (such as chemical toxicity) may also contribute to decreases in several of the endpoints. Chemical toxicity may occasionally lead to increased liver size (as part of the detoxification mechanism), together with decreases in the other endpoints.

The third, and most prominent, response pattern seen in Cycle 2 was associated with increases in condition and liver weight, and decreases in gonad weight. This was the national average response pattern (Fig. 4) and is generally indicative of nutrient enrichment coupled with metabolic disruption (Munkittrick *et al.* 2000). Thus, the national averages (grand means) in Figure 4 show that fish in effluent-exposed areas were significantly faster growing (increased weight at age), were significantly fatter (increased condition), and had significantly larger livers than fish in reference areas, but exposed fish also had significantly smaller gonads. This response pattern can occur when effluent exposure disrupts normal allocation of resources to gonadal development and may include some element of endocrine disruption associated with difficulties in producing sufficient sex steroid hormones (Munkittrick *et al.* 1991, Van Der Kraak *et al.* 1992, Damstra *et al.* 2002).

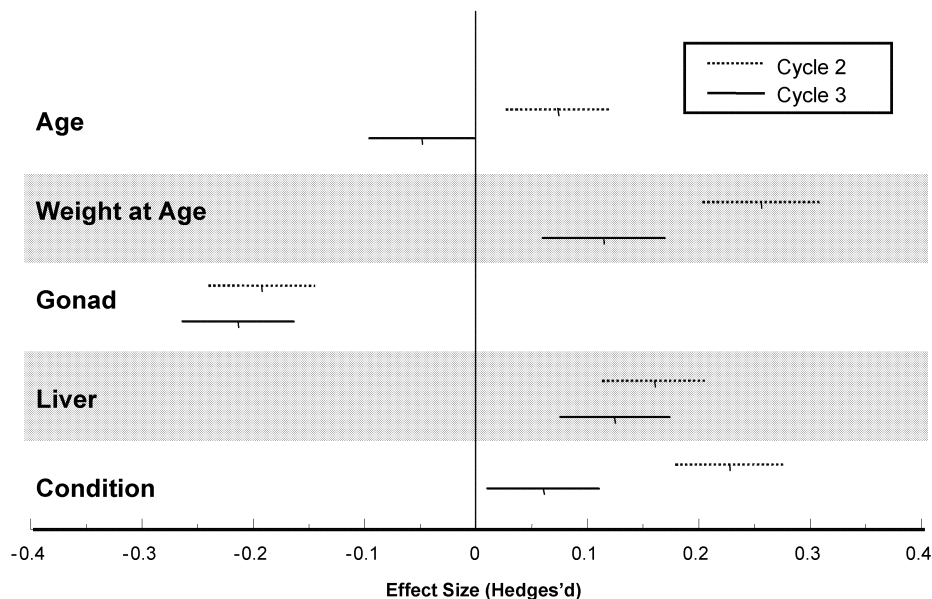


Figure 4: Grand means for five key fish endpoints for Cycles 2 (C2) and 3 (C3). Error bars represent 95% confidence intervals. Number of comparisons for: Age (C2 = 133; C3 = 138), Condition (C2 = 123; C3 = 123), Gonad (C2 = 126; C3 = 124), Liver (C2 = 128; C3 = 129), Weight at Age (C2 = 100; C3 = 105).

While Cycle 3 showed a similar diversity of response patterns, the national average response was quite similar to Cycle 2, with exposed fish exhibiting significant increases in growth rate, condition and liver size, but a significant decrease in gonad size (Fig. 4). Thus, in Cycle 3, the national average pattern was again consistent with one of nutrient enrichment coupled with metabolic disruption. Although the depression in gonad size remained unchanged since Cycle 2, the increase in condition and growth rate was lessened, perhaps signalling a reduction in nutrient enrichment effects in Cycle 3. An additional change that was observed between cycles was a switch from increased average age for exposed fish in Cycle 2 to decreased average age in Cycle 3. Note that the age and weight at age endpoints are included here for informational purposes, although they should be interpreted with care due to difficulties in aging fish at some mills (for this reason, these two endpoints are not being used to direct mills to more extensive monitoring in Cycle 4). Future cycles of data collection will provide more information on this issue.

Figure 5 shows the results of meta-analyses carried out to look at the influence of gender on fish responses. Greater differences were observed between genders in Cycle 3 compared with Cycle 2, where gender differences were observed only for the gonad endpoint. When breaking the results down by gender for Cycle 3, it is apparent that the lessening of nutrient enrichment effects was most pronounced for females (liver and condition endpoints). The change in the age response was also greater for females. In

contrast, the female versus male gonad response in Cycle 3 was almost identical to that in Cycle 2, with the gonad size reduction being significantly greater for females in both cycles.

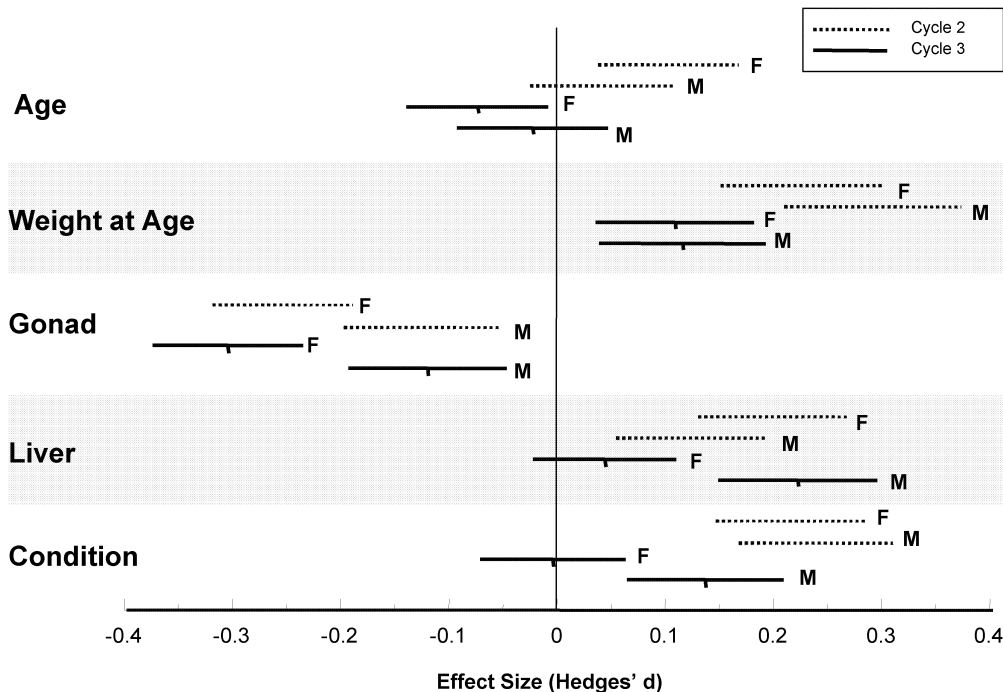


Figure 5: Five key fish endpoints, by gender, for Cycles 2 (C2) and 3 (C3). M = male, F = female. Error bars represent 95% confidence intervals. Number of comparisons for Age (C2: F = 67, M = 66); C3: F = 73, M = 65); Condition (C2: F = 61, M = 62; C3: F = 65, M = 58); Gonad (C2: F = 66, F = 60; C3: F = 65, M = 59); Liver (C2: F = 64, M = 64; C3: F = 70, M = 59); Weight at Age (C2: F = 50; M = 50; C3: F = 56, M = 49).

The national responses for Cycle 3 can also be subdivided by species (Fig. 6). The white sucker national average response pattern was the same as the overall national average pattern, and was consistent with nutrient enrichment in combination with metabolic disruption (characterized by increased liver size, condition and growth rate and decreased gonad size). This was likely a reflection of the large proportion of surveys that used white suckers as sentinel species (see Table 4). None of the other species was used at a large numbers of mills in Cycle 3. Nonetheless, some of the species that were less commonly used are included in Figure 6 for comparative purposes. Interpretations of the average responses shown by these other species should be done with care, given their low frequency of use. Of note, however, is the response pattern observed for mummichogs. This was the most commonly used marine species, and it exhibited a significant decrease in all the endpoints, generally indicative of a nutrient limitation or toxicity response pattern. This response is consistent with the overall inhibitory response pattern observed for benthic invertebrate communities in marine habitats, which will be discussed in

section 6. The fish data were not subdivided by habitat due to a lesser diversity of sampled habitats relative to the invertebrates (leading to low sample sizes for some habitat types) as well as their higher mobility relative to benthic invertebrates (so more difficult to assign to either erosional or depositional habitat types).

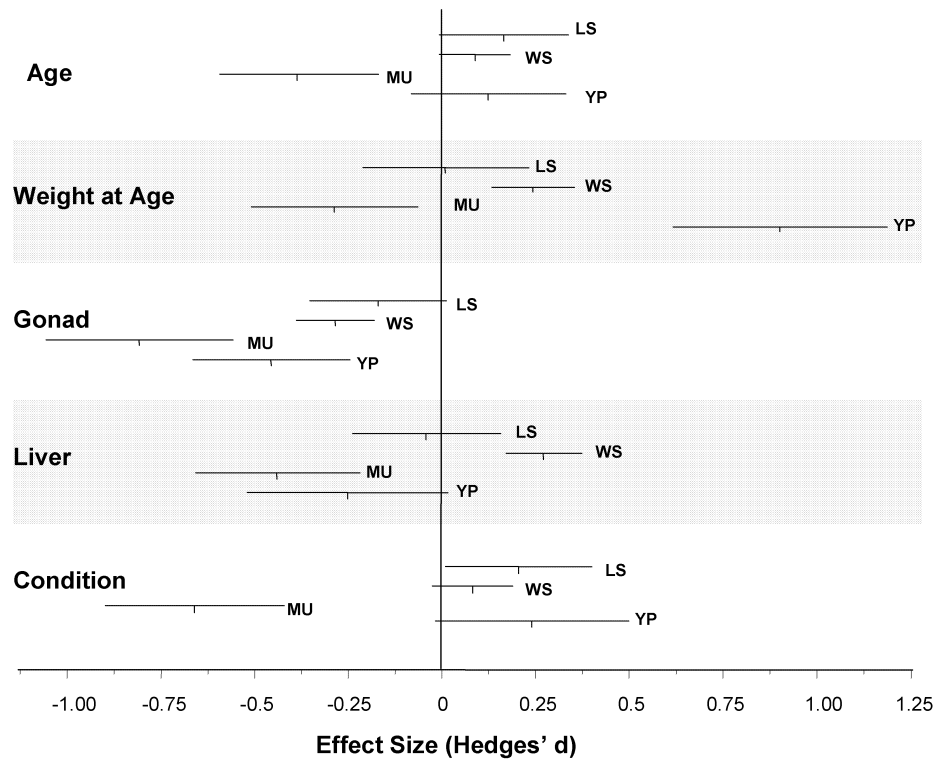


Figure 6: Five key endpoints, by species, for Cycle 3. Error bars represent 95% confidence intervals. LS = long-nose sucker (n = 10), WS = white sucker (n = 34), MU = mummichog (n = 7), YP = yellow perch (n = 6). Note that the number of comparisons (n) for each species is an average of all five endpoints. There were cases where certain measurements were not completed on some fish; therefore, the number of comparisons would be slightly lower for those endpoints.

Overall, the response patterns in Cycle 3 were quite similar to those observed in Cycle 2 (Fig. 4). The reduction in gonad size was unchanged between cycles. The Cycle 3 responses did, however, suggest a lessening in nutrient enrichment effects relative to Cycle 2. That is, while there were still significant increases in condition and growth rate in exposed areas in Cycle 3, they were not as great as observed in Cycle 2. This slight shift in fish response may have been due to a variety of possible causes, as discussed in section 8.

5. Fisheries Resources and Usability

Two components are included in the EEM program to evaluate the impact of effluent on the usability of fisheries resources by humans: tainting tests and analyses of levels of dioxins and furans in edible fish tissue. Tainting tests, which are designed to determine if the taste of the fish is altered, were undertaken at two mills in Cycle 3, the same mills where fish tainting occurred during Cycle 2 (see Lowell *et al.* 2003). The results of the tainting study at one mill confirmed that effluent was continuing to taint fish, although the difference from control method showed decreased intensity due to mill process and effluent treatment improvements. The results of the other tainting study showed no tainting in Cycle 3.

For Cycle 3, six mills were required to undertake fish tissue analyses for dioxins and furans. One freshwater mill conducted dioxins and furans tissue analyses, and five coastal mills continued to monitor fish tissue due to fisheries restrictions that had been put in place due to dioxins and furans found in the tissues of fish collected in areas near those mills. When concentrations of the dioxins and furans in fish tissue taken from fish exposed to pulp mill effluent were compared to the Health Canada fish consumption guidelines for dioxins and furans (Environment Canada 2004a), it was found that fish exposed to coastal mill effluent at three sites exceeded the guidelines.

6. Benthic Invertebrate Community Survey

The second primary component of the EEM program is the benthic invertebrate community survey, which assesses the impacts of mill effluent on fish habitat. The benthic invertebrate survey helps to supply information on the aquatic food resources available for fish and on the degree of habitat degradation due to organic enrichment or other forms of physical and chemical contamination. The four endpoints used to assess the effects of pulp and paper effluent on benthic invertebrate communities are abundance, taxon richness (number of taxa), Simpson's evenness and the Bray-Curtis index of dissimilarity. Taxa were analyzed at the family level (or above when data were reported only at a higher level); for further discussion of the rationale for the level of taxonomic resolution, see Bowman and Bailey (1997), Bailey *et al.* (2001), Lenat and Resh (2001) and Culp *et al.* (2003).

6.1 Data Processing and Study Designs

In Cycle 3, 67 mills used control/impact designs and 20 mills used gradient designs. Approximately the same number of surveys, by study designs, were conducted in Cycle 2. In addition, in Cycle 3, two mills used benthic invertebrate mesocosms as an alternative method, which were not included in the national analysis. Similar to the fish survey, the benthic invertebrate data were initially screened for errors; a few difficulties were encountered with the electronically submitted data, including the following: 1) raw numbers in electronic files not matching numbers in interpretative reports, 2) benthic invertebrate densities not reported as numbers per square metre, 3) number of stations in electronic files not matching number of stations in interpretative reports and 4) inclusion

of organisms and fragments of organisms in the electronic data that were not included in the interpretative report analysis. Overall, most of the data submitted were of good quality, and most mistakes were corrected manually when they were discovered. Only three benthic invertebrate surveys could not be included in the national analysis due to data quality errors in electronic submission.

6.2 *Summary of Effect Sizes*

The Cycle 3 distribution and range in measured exposure versus reference area percent differences for abundance, taxon richness, Simpson's evenness and the Bray-Curtis index of dissimilarity were similar to those in Cycle 2 (Fig. 7; Lowell *et al.* 2003). As was found for fish, this underscores the reproducibility of measured differences at a national scale in Cycle 3 relative to Cycle 2 and likely reflects the consistency of response patterns from one cycle to the next (see also section 6.4). The measured differences were calculated as the exposure area mean minus the reference area mean, expressed as a percentage of the reference area mean. Figure 7 provides a complete summary of the differences calculated for control/impact mills (i.e., includes both statistically significant and non-significant differences). This figure does not include mills that conducted a gradient design, since percent differences (as computed here) cannot be calculated for this type of study design; note, however, that standardized effect sizes can be calculated for gradient designs for the purposes of meta-analysis (see section 3 and section 6.4).

Similar to Cycle 2, the abundance endpoint showed the most extreme range, varying from a decrease of 80% to an increase of over 5500%. The percent differences distribution for taxon richness ranged from a decrease of 67% to an increase of 147%. Ranges for the Bray-Curtis index and the Simpson's evenness index were -33% to 471% and -58% to 227%, respectively. Note that most of the Bray-Curtis values were positive, due to the method of calculation. The few negative values were due to unusual data distributions.

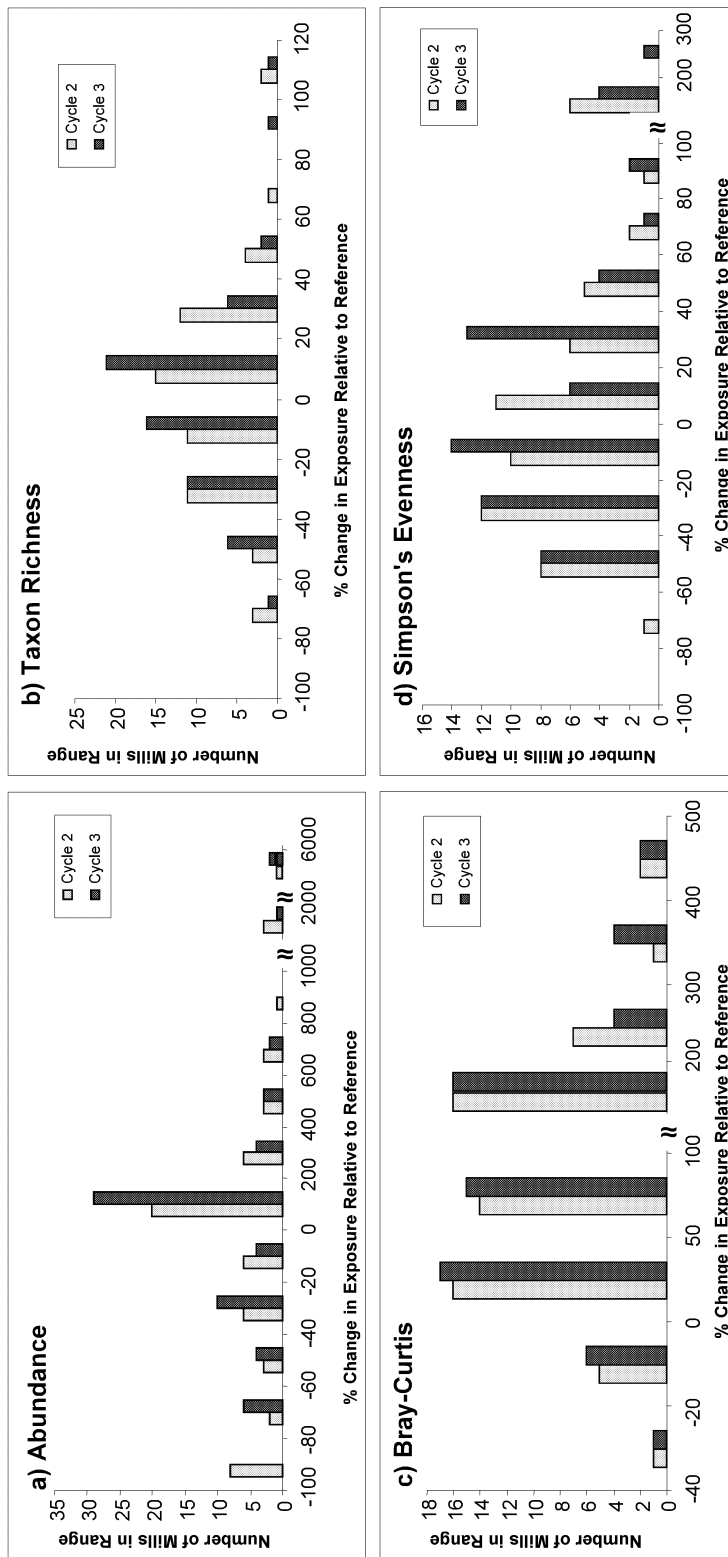


Figure 7: Distribution of measured percent differences between exposure and reference areas for the benthic invertebrate survey (control/impact designs only) for a) abundance, b) taxon richness, c) Bray-Curtis and d) Simpson's evenness.

Figure 8 was developed using all of the Cycle 3 results (control/impact and gradient designs) and shows the number of mills that had no significant effect versus a significant effect for the four endpoints. The figure also further divides the mills that had a significant effect into those where the effect was less than the CES of ± 2 standard deviations ($\pm 2SD$) and those where the effect exceeded the CES of $\pm 2SD$. For both Cycle 2 and Cycle 3, the highest percentages of the mills were non-significant for taxon richness, abundance and Simpson's evenness. The Bray-Curtis index was the most sensitive endpoint in both cycles, and this is shown by a high number of mills having a significant effect that exceeded $\pm 2SD$.

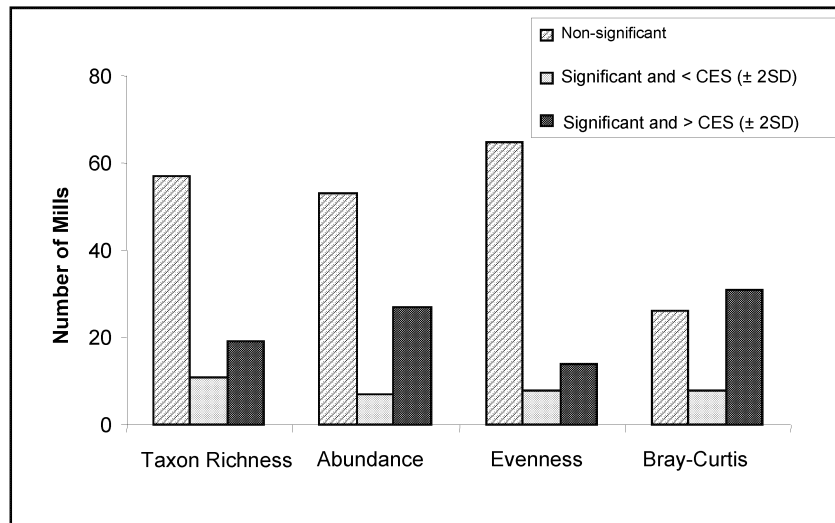


Figure 8: Number of mills showing no significant difference, a significant difference less than the CES of $\pm 2SD$ and a significant difference greater than the CES of $\pm 2SD$. Includes both control/impact designs ($n = 67$) and gradient designs ($n = 20$). Note that CES have not yet been developed for gradient designs; therefore, mills that used a gradient design and found a statistically significant effect are included under “significant and $> CES$.” See text for additional information. Note also that for the gradient designs, the Bray-Curtis index was not calculated because it requires highly site-specific information that was not available at a national scale (Lowell *et al.* 2003).

Although Figure 8 shows that many mills did not show a significant effect when considering each endpoint individually (except Bray-Curtis), approximately 54% of the mills that conducted benthic invertebrate studies found at least one significant effect in at least one of the benthic invertebrate endpoints in Cycle 3. Approximately half of these mills exceeded the CES of $\pm 2SD$ for taxon richness or abundance in Cycle 3.

For control/impact designs, a mill will conduct focused monitoring (including magnitude and geographic extent) for Cycle 4 if the CES for abundance or taxon richness was exceeded (with statistical significance) in the same direction for two consecutive cycles (Cycles 2 and 3). For gradient designs, if a mill measured a statistically significant effect (significant correlation coefficient) for abundance or taxon richness in the same direction for both Cycles 2 and 3, the mill would conduct more focused monitoring in Cycle 4. Given the inherent lower power for detecting effects when using the gradient design relative to the control/impact design (Lowell *et al.* 2003), it is likely that only gradient mills having particularly large effects would conduct more extensive monitoring. Similarly, only control/impact mills having particularly large effects would conduct more extensive monitoring, due to the dual requirement for control/impact designs of statistical significance and exceedance of the CES.

By examining the number of exceedances of CES in both Cycles 2 and 3 as an indicator of sites where more serious impacts are occurring, it is estimated that more extensive monitoring effort would be undertaken at approximately 20% of the mills conducting a benthic invertebrate community survey. Approximately half of these mills exceeded the CES for abundance, about one-third exceeded the CES for taxon richness, and the remaining mills exceeded the CES for both abundance and taxon richness. The majority of these mills exhibited significantly higher endpoint values in the exposure area; however, most mills where the CES was exceeded solely for taxon richness showed significantly lower endpoint values in the exposure area. This was also observed in Cycle 2. The significant increases observed in abundance were usually an indication of overall eutrophication. The decreases in taxon richness, in turn, were likely a reflection of more pronounced eutrophication or toxicity/smothering (see following section).

6.3 *Response Patterns and Meta-analyses*

Three primary benthic invertebrate community response patterns were observed in Cycle 2 (Lowell *et al.* 2003). These are illustrated in Figure 9, which shows expected changes in abundance, taxon richness and community structure with changing effluent quality (pulp mill effluent effects on invertebrates reviewed in Lowell *et al.* 1995, 2000; Chambers *et al.* 2000; Culp *et al.* 2000; Lowell and Culp 2002). In this context, “effluent quality” refers to that experienced by organisms exposed in the field, as distinguished from effluent effects measured under controlled conditions in the laboratory. The x-axis progresses from better quality effluent and less deleterious effects on the left to poorer quality effluent and more deleterious effects on the right. More deleterious effects may also be associated with past historical effects (such as the smothering effects of fibre mats generated over many years of mill operation).

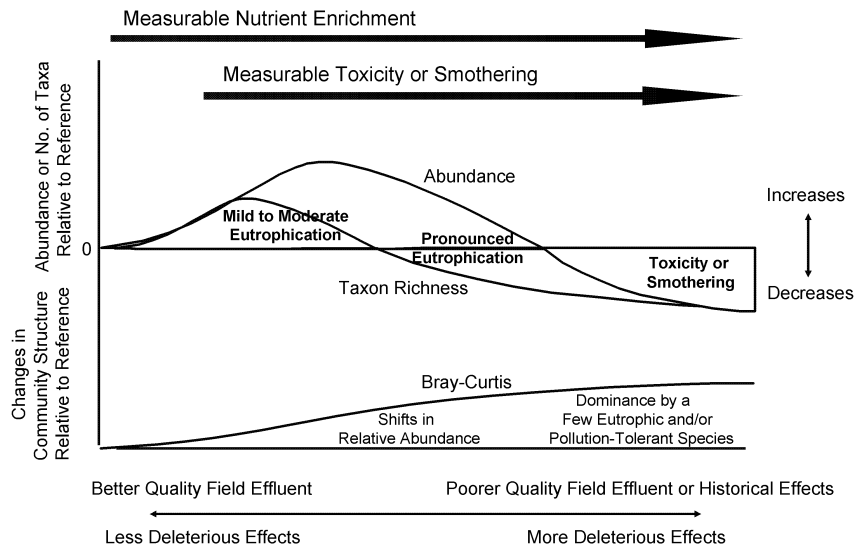


Figure 9: Predicted response patterns for benthic invertebrate communities (Lowell *et al.* 2003).

In general, nutrient enrichment (or eutrophication) increases from left to right, along with increasing toxicity or smothering effects (Fig. 9). Nutrient enrichment can often be measured at lower effluent concentrations than toxicity, and toxic effects are often masked by eutrophication at low to medium concentrations. Mild eutrophication is typified by increases in both total abundance and taxon richness. Progressing to the right, moderate eutrophication is typically associated with lessened increases in taxon richness, although further increases in abundance may still occur. More pronounced eutrophication is commonly associated with decreases in taxon richness, even while abundance is still greater than that found in reference areas. Finally, decreases in both taxon richness and abundance are typically a sign of overall inhibitory effects, such as toxicity or smothering.

In the EEM program, changes in invertebrate community composition are measured by changes in the Bray-Curtis index of dissimilarity. The value of this index usually increases with poorer quality effluent, reflecting changes in community structure (particularly community composition). This is illustrated at the bottom of Figure 9. Evenness similarly measures changes in community structure and typically decreases (or sometimes increases) with poorer quality effluent. However, it should be emphasized that changes in community composition are not always tied to changes in total abundance and taxon richness, as there are many different ways in which benthic communities may be affected in the field. Due to complex direct and indirect effects (e.g., substitution of more sensitive species by less sensitive species), effluent exposure may lead to pronounced effects on community composition without large effects on abundance or taxon richness, and vice versa.

The national average response pattern in Cycle 2 was consistent with one of mild to moderate eutrophication, as indicated by increases in abundance and no national average change in taxon richness (although subgroups of mills showed either significant increases or significant decreases in taxon richness) (Fig. 10; Lowell *et al.* 2003). Note that Figure 10 focuses on control/impact mill data, the most commonly used design in the EEM program. This allowed direct comparisons among all four endpoints (calculation of the Bray-Curtis endpoint for gradient mills requires highly site-specific information that was not readily available at a national scale; Lowell *et al.* 2003).

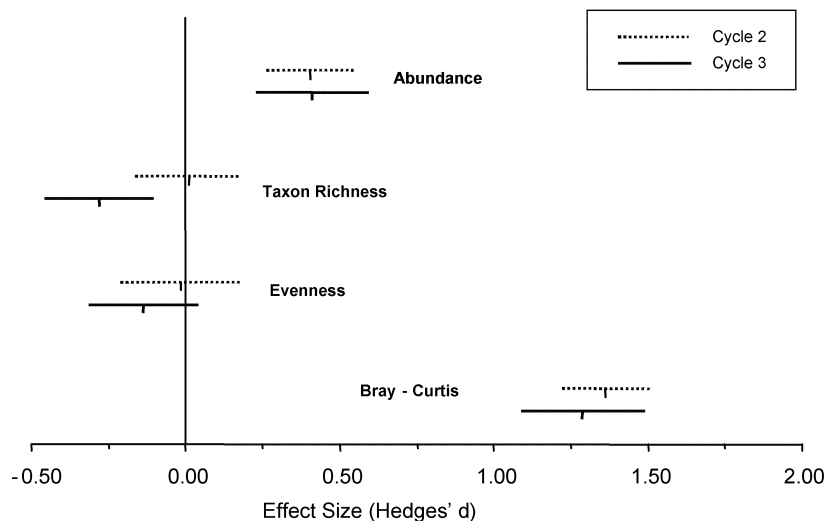


Figure 10: Control/impact grand means for Cycles 2 (n = 62) and 3 (n = 55). Error bars represent 95% confidence intervals.

In Cycle 3, the national average response pattern was similar to that observed in Cycle 2, although some shifts in the degree of response were also observed (Fig. 10; grand means averaging data across habitats on a national basis). Average abundance showed no change between cycles. In Cycle 3, however, there was a significant decrease in taxon richness in exposure areas relative to reference areas, indicative of a possible shift toward more pronounced eutrophication relative to Cycle 2 (mostly due to a response shift in freshwater depositional habitats; see below). As with abundance, there was also a high degree of overlap between cycles for the evenness and Bray-Curtis endpoints, particularly for the control/impact mills. Including data (not shown) from the lower number of mills using the gradient design had little effect on the overall response patterns shown in Figure 10, except for a shift in evenness slightly to the right in Cycle 2 and slightly to the left in Cycle 3. A decrease in evenness in exposure areas relative to reference areas (significant for gradient mills in Cycle 3, but not observed in Cycle 2) is a fairly commonly observed effect of stressors on benthic invertebrate communities.

In Cycle 2, distinct differences in response patterns were observed among habitat types, and similar differences were also seen in Cycle 3 (Fig. 11). The control/impact and gradient data are combined in Figure 11, given that the Bray-Curtis endpoint is not included and in order to provide a larger number of comparisons for the less commonly used habitat types. In both cycles, erosional river habitats exhibited increases in abundance and at least a tendency towards increases in number of taxa, indicative of mild to moderate eutrophication.

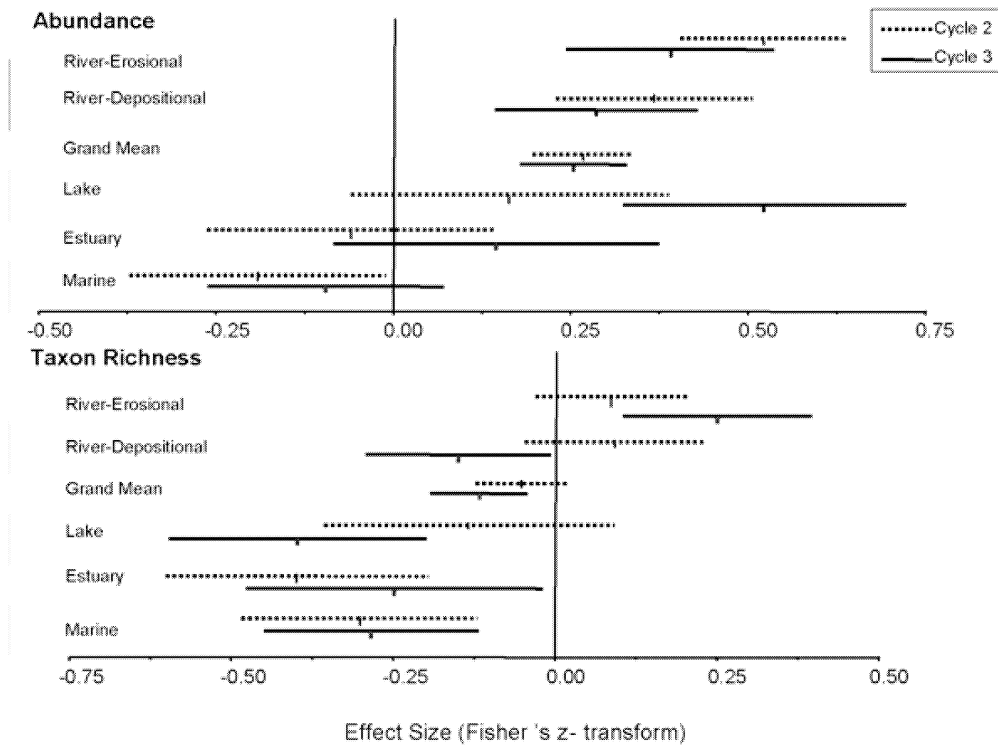


Figure 11: Abundance and taxon richness, by habitat, for Cycles 2 (C2) and 3 (C3). Error bars represent 95% confidence intervals. Number of studies for: river-erosional (C2 = 34; C3 = 21), river-depositional (C2 = 21; C3 = 23), lake (C2 = 7; C3 = 9), estuary (C2 = 8; C3 = 10), marine (C2 = 13; C3 = 12).

In marine and estuarine habitats, taxon richness exhibited significant decreases during both cycles (Fig. 11). In marine habitats, abundance showed a significant decrease in Cycle 2, but was not significant in Cycle 3. A similar slight shift to the right was seen in estuarine habitats, perhaps reflecting a lessening of effects measured in marine-type habitats in Cycle 3. The high degree of overlap between cycles makes any firm conclusions premature, however.

In freshwater depositional habitats (river-depositional and lakes), abundance showed similar increases during both cycles. In Cycle 3, however, taxon richness was significantly decreased in exposure areas relative to reference areas in these habitats (this was not the case in Cycle 2). This is suggestive of a shift toward more pronounced eutrophication in these habitats during Cycle 3. This also led to the overall shift in the taxon richness grand mean in Figure 10.

As was also found for the fish survey, the overall benthic response patterns were fairly similar between Cycles 2 and 3. The abundance, Bray-Curtis and (for control/impact mills) evenness grand means all showed a high degree of overlap. There were, however, some shifts in the degree of response in some habitat types that may have resulted from a number of possible causes, as discussed in section 8.

7. Sublethal Toxicity Testing

7.1 Introduction

Under the EEM program, each Canadian pulp and paper mill is required to measure the sublethal toxicity of its final effluent discharge. Mills conduct a battery of three sublethal toxicity tests twice a year (summer and winter) for each year of the cycle, including: 1) early life stage development of fish, 2) reproduction of an invertebrate and 3) algal growth inhibition or reproduction. The sublethal toxicity testing component of EEM follows prescribed acceptable test methods and includes species choices for different receiving environments (see Environment Canada 2004a for specific test methods).

For the purposes of this national assessment, the results of sublethal toxicity testing are used to measure changes in effluent quality over time. On a site-specific basis, sublethal toxicity data can also be used to understand and estimate the relative contribution, in multiple discharge situations, to observed effects in the receiving environment.

The endpoint that was used to measure effluent quality was the *inhibiting concentration for 25% effect (IC₂₅)* — i.e., that concentration causing performance (e.g., growth, reproduction) that was 25% inferior to that of the control organisms. In a freshwater test, if full-strength effluent did not cause 25% inhibition/effect, then the endpoint was reported as $\geq 100\%$ concentration. For estuarine/marine tests, the highest concentration range for effluents that were salinity-adjusted was summarized in a category designated $\geq 60\%$. Note that for estuarine/marine tests, if full-strength effluent did not cause 25 % inhibition, then the endpoint was reported as a category designated $\geq 60\%$ because it was the highest concentration that could be tested due to the standard procedure of adding hypersaline brine to the effluent for salinity adjustments. The method for adjusting salinity has more recently been changed to the addition of dry salts. This now makes it possible to test full strength effluents.

7.2 Monitoring Changes in Effluent Quality Among Cycles

Sublethal toxicity endpoints can be used to compare the quality of effluent at different times. Comparisons among results for Cycles 1, 2 and 3 were made for the distributions of endpoints from all Canadian mills for each freshwater and estuarine/marine test species (Figs. 12 and 13). For each test, all IC_{25} s were compiled, and the percentage of tests falling into each defined category range of effluent concentration (i.e., $\geq 100\%$ v/v, $\geq 60\%$ to $< 100\%$ v/v, $\geq 36\%$ to $< 60\%$ v/v, etc.) was calculated. For example, Figure 12A illustrates the results of sublethal toxicity tests conducted on *Ceriodaphnia dubia* during Cycles 1, 2 and 3 of the EEM program. The vertical bars indicate the percentage of tests conducted in each cycle in which *C. dubia* exhibited a 25% decrease in function at that threshold of concentration. A high proportion of tests carried out in Cycle 1 showed this effect at concentrations of effluent below 1%; however, during Cycles 2 and 3, between 25% and 30% of mills did not show this effect at concentrations equal to or greater than 100% effluent.

The Cycle 3 sublethal data sets were similar to the Cycle 2 results for the freshwater test species (*C. dubia*, fathead minnow and *Selenastrum capricornutum*), confirming the improvement in effluent quality from Cycle 1 (Fig. 12; note that *S. capricornutum* is now classified as *Raphidocelis subcapitata*). This improvement in effluent quality was attributed in the national assessment of Cycle 2 EEM data to the installation of advanced effluent treatment after Cycle 1. The results of the rainbow trout toxicity test, which is used only in British Columbia, however, indicated an increase in moderate toxicity of effluent (i.e., $> 36\%$ v/v to $> 60\%$ v/v) in comparison with Cycles 1 and 2 (Fig. 12C). The reason for this will remain unknown without an in-depth analysis of the individual mills. Problems with egg viability, however, were reported. To resolve this issue, the testing period was moved from January to May, which may have altered the degree of effluent toxicity that was measured.

For the marine toxicity test results, in particular the toxicity tests with *Champia parvula*, topmelt and inland silversides, Cycle 3 results were similar to the Cycle 2 results. The inland silverside growth assay results for Cycle 3 confirmed the Cycle 2 improvement in effluent quality in comparison with Cycle 1 (Fig. 13). The Cycle 3 results for the topmelt growth inhibition test, which showed no toxicity to effluent at full concentration, were similar to the Cycle 2 results, but no comparison could be made with Cycle 1, as this test was not used at that time (Fig. 13). Although there is a wide spread in the distribution of the toxicity data for the *Champia* reproduction assay, the data for both Cycles 2 and 3 generally indicate an improvement in effluent quality compared with the Cycle 1 test results (Fig. 13).

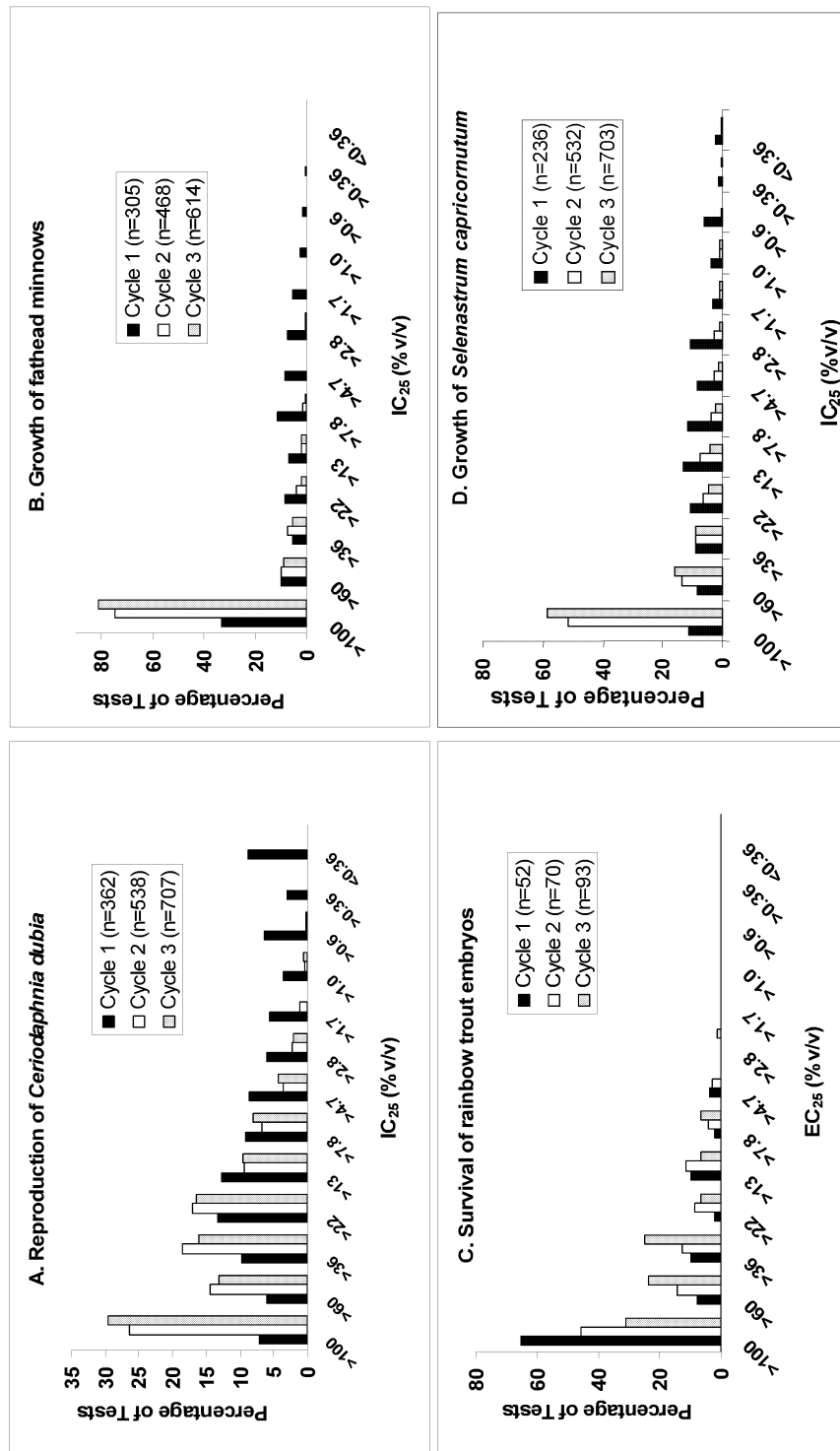


Figure 12: Comparison of sublethal toxicity to freshwater species in Cycles 1, 2 and 3. Note that for the rainbow trout test, the EC_{25} is the effective concentration for 25% of the embryos.

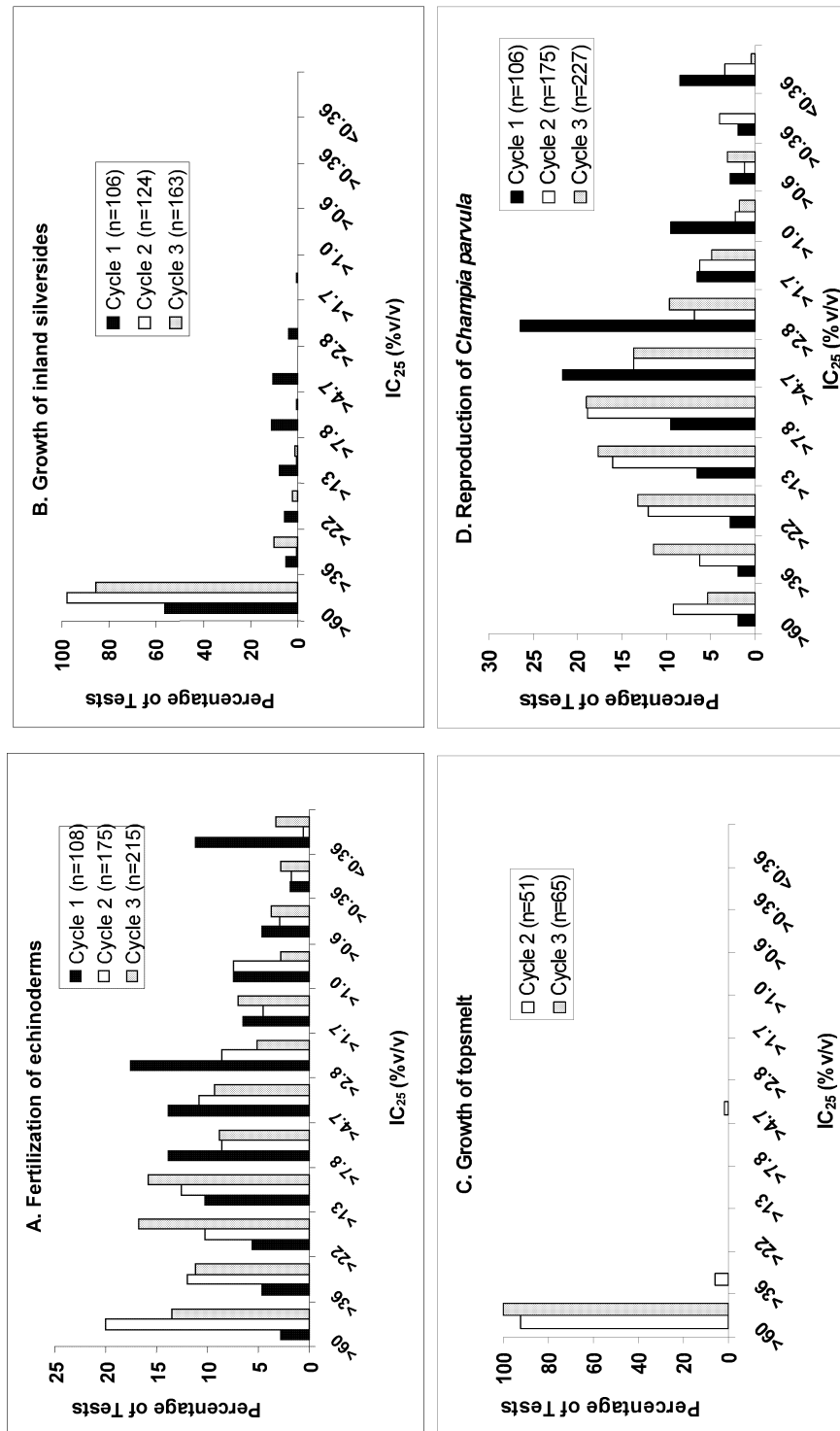


Figure 13: Comparison of sublethal toxicity to marine species in Cycles 1, 2 and 3.

For the echinoderm fertilization test, there was a slight shift in the data spread from no toxicity in Cycle 2 to a moderately toxic effect (i.e., in the range of <22% v/v to $\geq 7.8\%$ v/v) in Cycle 3. The reason for this shift will not be known without further analysis of the individual mill data. There may be higher variability in the echinoderm toxicity data because of the different species that were used for this test between cycles. For Cycle 3, 85% of test species were sea urchins, including *Lytechinus pictus*, *Arbacia punctulata* and *Strongylocentrotus purpuratus*, and 15% of test species were sand dollars (*Dendraster excentricus*). In contrast, *Arbacia* was not used in Cycle 2.

Looking at both the freshwater and marine histograms over all three cycles (Figs. 12 and 13), the tests measuring reproductive endpoints (*C. dubia*, echinoderms, *C. parvula*) were shifted the farthest to the right, indicating the higher sensitivity of those tests.

In Figure 14, the percentages of tests with IC₂₅s showing no sublethal response in the highest test concentration are compared for Cycles 1, 2 and 3. For most of the tests, the percentages of tests reporting no sublethal response increased from Cycle 1 to 2 and showed little further change in Cycle 3 (as discussed above, methodological difficulties were encountered with the rainbow trout test).

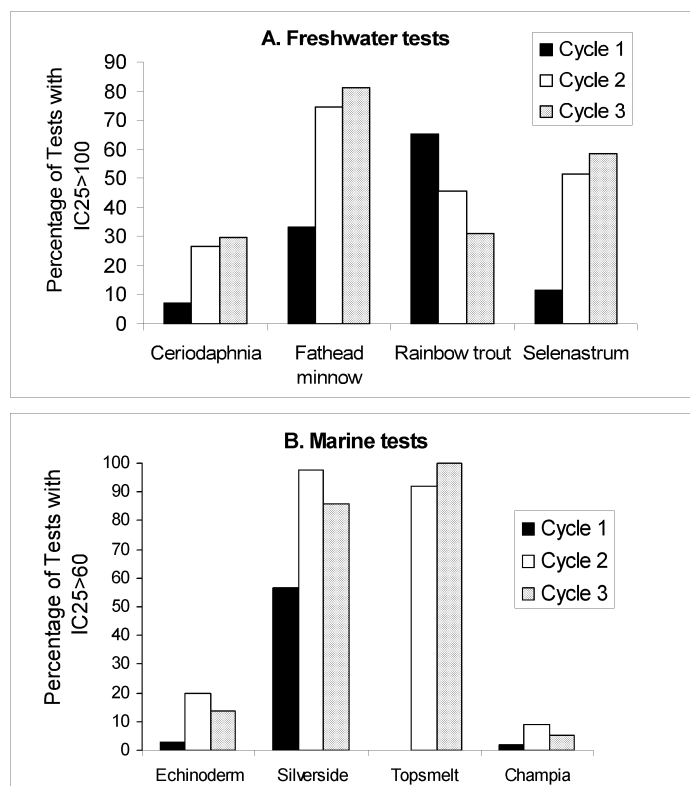


Figure 14: Percentages of sublethal tests showing no effect at highest concentration.

7.3 Summary and Future Considerations

Overall, the national analysis of the Cycle 3 sublethal toxicity data was consistent with the analysis of the Cycle 2 sublethal toxicity data and confirmed the improvement in effluent quality at Canadian pulp and paper mills from Cycle 1 noted in the Cycle 2 national EEM assessment by Lowell *et al.* (2003). This improvement in effluent quality following Cycle 1 was likely linked to upgrades in effluent treatment in response to the PPER. No further improvement in effluent quality, however, was found for Cycle 3.

The pattern of distributions for the Cycle 2 versus Cycle 3 sublethal toxicity results was similar for *C. dubia*, fathead minnow, *S. capricornutum*, topsmelt, inland silverside, *C. parvula* and, to a lesser degree, echinoderms. The rainbow trout development test results were inconsistent with the results of the other test species.

Although the field surveys found nutrient enrichment to be widely associated with pulp mill effluent exposure for both Cycle 2 and Cycle 3, this type of effect was not assessed using current practices for EEM toxicity testing. It is recommended that future algal toxicity testing on *S. capricornutum* and *C. parvula* in the EEM program report both stimulatory and inhibitory effects.

Results of some of the sublethal toxicity tests (particularly topsmelt) indicated that a high proportion of tests were not responsive to effluent toxicity ($IC_{25} > 100\%$ or no effect at the highest concentration). This clear trend of no sublethal effect at the highest concentration, consistent between Cycle 2 and Cycle 3, may be indicative of the need to reassess the cost-effectiveness and usefulness of some of the tests included in the battery of sublethal toxicity testing used to monitor effluent quality. On the other hand, it is noteworthy that the greatest test sensitivity was observed for those endpoints measuring reproductive endpoints.

8. Summary and Conclusions

Having completed its third cycle of data collection and reporting, the EEM program has provided an extensive database and a fairly robust picture of the nature of the effects of pulp and paper mill effluents on receiving water biota across the country.

The response patterns observed for effluent-exposed fish and benthic invertebrates during Cycle 3 were, for the most part, similar to those measured during Cycle 2. The national average response pattern measured for fish in both Cycles 2 and 3 was consistent with one of nutrient enrichment overlaid by metabolic disruption. That is, exposed fish have consistently shown evidence of increased food availability or increased food absorption (fatter, faster growing, with larger livers) together with disruption of allocation of resources to reproduction (smaller gonads). This latter metabolic disruption may include some aspect of endocrine disruption associated with problems in producing sufficient sex steroid hormones. Other observed response patterns for fish have included nutrient enrichment without measurable metabolic disruption, nutrient limitation and chemical toxicity. These findings agree with a number of research studies that have been

carried out at mills across the country (reviewed by Munkittrick *et al.* 1991, 1994, 2000, Lowell *et al.* 2003, 2004).

The national average response for benthic invertebrate communities in both Cycles 2 and 3 was consistent with one of eutrophication, ranging from mild to more pronounced, partly depending on habitat type. More specifically, benthic invertebrate communities exposed to pulp mill effluent have commonly exhibited increases in abundance, together with some combination of increases, decreases or no change in taxon richness, depending on the degree of eutrophication. Other observed benthic invertebrate response patterns have included toxicity or smothering effects. These response patterns and the mechanisms leading to them have been confirmed by a number of experimental studies conducted at mills across Canada (reviewed by Culp *et al.* 2000, Lowell *et al.* 2000, 2003).

The sublethal toxicity data showed clear improvements in effluent quality from Cycle 1 to Cycle 2, with, for the most part, no further changes in effluent quality in Cycle 3. It should be noted, however, that the sublethal toxicity reporting methodology and selection of tests do not track all aspects of effluent quality. In particular, they do not currently measure nutrient enrichment effects of the effluent (the most commonly observed effect in the field surveys), although the tests could be modified to provide some information on nutrient enrichment effects. Also of note, and consistent with the reductions in gonad size measured in the fish field survey, the most sensitive sublethal toxicity tests were those that measured a reproductive endpoint (although these did not include fish tests).

Some possible shifts in the patterns of effects measured in the fish and benthic invertebrate field surveys between Cycles 2 and 3 were also observed. The fish data have suggested some overall lessening of nutrient enrichment effects. On a habitat-specific basis, the benthic invertebrate data have shown evidence of a possible lessening of toxicity/smothering effects in marine-type habitats, as well as more pronounced eutrophication in freshwater depositional habitats. These shifts for fish and invertebrates may be due to a variety of possible causes. For example, the average reduction in number of invertebrate taxa observed in Cycle 3 may have led to a reduction in food availability for sentinel fish at some mills (resulting in reduced nutrient enrichment in fish), if the taxa that were lost were also taxa that the fish would normally use as a food source. Alternatively, the shifts in fish and invertebrate responses may have been influenced by changes between cycles in study designs and selection of study areas. It is also possible that the fish and benthic invertebrate response patterns will continue to show slight shifts from cycle to cycle due to natural variability in the receiving environment or to anthropogenic changes through time that are unrelated to exposure to pulp and paper mill effluent. Under this extrinsic variability scenario, these shifts may include either increases or decreases in individual endpoints, sometimes changing direction from cycle to cycle. In this situation, the underlying, time-averaged response pattern may remain approximately the same, unless these extrinsic factors show a long-term, unidirectional trend.

The sublethal toxicity data suggest that effluent quality has not changed between Cycle 2 and Cycle 3. However, as noted above, not all aspects of effluent quality are currently measured by the sublethal toxicity tests (particularly nutrient enrichment), so it is possible that changes in effluent quality may have contributed to the field survey shifts outlined above. It should also be noted that some of these shifts may have resulted from gradual recovery from (or, in some cases, added accumulation of) historical effluent deposits in the receiving environment. Further, some of the shifts may have resulted from lag effects. That is, recovery from (or worsening of) some effluent effects may not be evident until some time after release of the effluent causing the effects, as effects are gradually transmitted through complex trophic food webs. Additional information is required to fully address these issues. The results obtained in Cycle 4 and beyond, together with more focused studies at some mills, will help to answer these questions.

In the meantime, the national analyses have identified a subset of mills and kinds of effects toward which more extensive monitoring should be directed. Finer-resolution evaluations will be made at the regional level, but initial estimates are that approximately 20% of the monitored mills are expected to conduct only sublethal toxicity testing in Cycle 4, while roughly another 20% are would be expected to progress to determinations of the extent and magnitude of effects or, for some mills, investigations of the causes of the effects.

Despite the mainly small shifts in response patterns noted above, the effects measured in the fish and benthic invertebrate field surveys were, for the most part, quite similar between Cycles 2 and 3. Most notably, both cycles have shown 1) evidence of overall nutrient enrichment or eutrophication (with the degree of response partly linked to habitat type) and 2) a reduction in fish gonad size that has remained virtually unchanged over two cycles of data collection.

9. Glossary

Benthic invertebrate community – The interacting populations of small animals (excluding fish and other vertebrates), living at the bottom of a water body, on which fish may feed. Measuring changes in invertebrate communities helps to understand changes in aquatic habitats and provides an evaluation of the aquatic food resources available to fish.

Bray-Curtis index – An index that measures the degree of difference in community structure (especially community composition) between sites. This measure helps to evaluate the amount of dissimilarity between benthic invertebrate communities at different sites.

Condition factor – A measure of the physical condition of fish that describes the relationship between body weight and body length. Essentially, the condition factor measures how “fat” fish are at each area.

Control/impact design – A study design consisting of no less than one reference area, usually upstream from the mill or situated in a different watershed, and a series of downstream exposure areas. This study design was most commonly used in rivers.

Depositional – Section of a riverine (or other) habitat where the flow of water tends to be slower and therefore where sediment tends to deposit. The bottom substrate in these areas tends to be softer and more granular in nature.

Effect – In the context of the EEM program, an effect is a statistically significant difference between measurements taken from the exposure area and from the reference area or measurements taken from sampling areas that have gradually decreasing effluent concentrations.

Endocrine – The endocrine system controls a number of internal body functions via hormonal secretions, which are transported throughout the body in the blood.

Endpoint – A particular measurement that is used as an indicator of potentially important effluent effects on receiving water biota. Examples of endpoints are gonad weight, liver weight, condition factor, age and weight at age for fish or abundance, taxon richness, Simpson’s evenness index and Bray-Curtis index of dissimilarity for benthic invertebrates.

Erosional – Section of a riverine (or other) habitat where the flow of water tends to be fast and turbulent. In these areas, sediments are usually carried downstream. Generally, the bottom substrate in these sections tends to be made up of larger sediments, rocks and boulders.

Eutrophication – The process of over fertilization of a body of water by nutrients that often results in excessive production of organic biomass and is typified by large numbers of organisms and, when pronounced, few species. Eutrophication can be a natural

process, or it can be accelerated by an increase of nutrient loading to a water body by human activity.

Exposure area – A sampling area where fish and benthic invertebrates are exposed to pulp and paper mill effluent. This area may extend through a number of receiving environments and contain a variety of habitat types.

Gradient design – Generally, sampling is done along a gradient of decreasing effluent concentration, starting with exposure areas close to the mill and progressing towards less exposed areas farther from the mill. This study design was sometimes used in situations where rapid effluent dilution was a factor.

Metabolic disruption – Metabolism is a mechanism used by the body whereby complex substances are synthesized from simple ones or complex substances are broken down. The disruption of this system can occur from exposure to deleterious substances in the environment and can cause important imbalances in the maturation, sexual behaviour, growth, etc. of the organism.

Nutrient enrichment – The effect of adding large quantities of organic and inorganic nutrients to the environment.

Reference area – A sampling area that has no effluent exposure from the pulp and paper mill in question and natural habitat features that are similar to those of the exposure area, including anthropogenic impacts.

Simpson's evenness index – A measure of how evenly individuals are distributed among taxa. This measure helps to evaluate changes in the relative abundance of taxa.

Smothering – The overaccumulation of organic matter derived from pulp mill effluent at the bottom of a water body, impeding the functioning of organisms and sometimes causing death.

Sublethal toxicity – In the context of EEM, sublethal toxicity tests usually measure the proportion of organisms affected by their exposure to specific concentrations of pulp mill effluent in a laboratory setting. A sublethal toxicity test measures what is detrimental to the organism (e.g., effects on growth or reproduction), but below the level that directly causes death within the test period.

Taxon – Organisms are classified into categories based on similarities and evolutionary relationships between them. Each of these categories (e.g., species, genus, family, phylum, etc.) is called a taxon (plural taxa).

Taxon richness – The total number of different taxonomic categories collected at a sampling station.

Total abundance – The total number of individuals of all taxonomic categories collected at the sampling station, expressed per unit area (i.e., density).

Weight at age – A measurement of the rate of growth of fish described by the relationship of size (weight) to age. Over the entire life span of a fish, the rate of increase in size may decline as the fish ages.

10. Acronyms / Abbreviations

Adj. mean	– adjusted mean
ANOVA	– analysis of variance
ANCOVA	– analysis of covariance
C2	– cycle 2
C3	– cycle 3
CES	– critical effect size
EEM	– Environmental Effects Monitoring
IC ₂₅	– inhibiting concentration for 25 % effect
LS	– longnose sucker
MU	– mummichog
NWRI	– National Water Research Institute
PPER	– <i>Pulp and Paper Effluent Regulations</i>
RAPPER	– <i>Regulations Amending the Pulp and Paper Effluent Regulations</i>
SAT	– statistical assessment tool
SD	– standard deviation
WS	– white sucker
YP	– yellow perch

11. References

- Bailey, R.C., R.H. Norris, and T.B. Reynoldson. 2001. Taxonomic resolution of benthic macroinvertebrate communities in bioassessments. *J.N. Am. Benthol. Soc.* 20: 280-286.
- Bowman, J.F. and R.C. Bailey. 1997. Does taxonomic resolution affect the multivariate description of the structure of freshwater benthic macroinvertebrate communities? *Can. J. Fish. Aquat. Sci.* 54: 1802-1807.
- Chambers, P.A., A.R. Dale, G.J. Scrimgeour, and M.L. Bothwell. 2000. Nutrient enrichment of northern rivers in response to pulp mill and municipal discharges. *J. Aquat. Ecosyst. Stress Recov.* 8: 53-66.
- Courtenay, S.C., K. Munkittrick, H.M.C. Dupuis, R. Parker, and J. Boyd. 2002. Quantifying impacts of pulp mill effluent on fish in Canadian marine and estuarine environments: problems and progress. *Water Qual. Res. J. Can.* 37(1): 77-99.
- Culp, J.M., R.B. Lowell and K.J. Cash. 2000. Integrating mesocosm experiments with field and laboratory studies to generate weight-of-evidence risk assessments for large rivers. *Environ. Toxicol. Chem.* 19: 1167-1173.
- Culp, J.M., M.E. Wiseman, R.C. Bailey, N.E. Glozier, R.B. Lowell, T.B. Reynoldson, L. Trudel, and G.D. Watson. 2003. New requirements for benthic community assessments at Canadian metal mines are progressive and robust: reply to Orr *et al.* *SETAC Globe* 4: 31-32.
- Damstra, T., S. Barlow, A. Bergman, R. Kavlock and G. Van Der Kraak (eds.). 2002. Global Assessment of the State-of-the-Science of Endocrine Disruptors. WHO/PCS/EDC/0.2.2; World Health Organization, International Programme on Chemical Safety.
- Doherty, C.A., R.A. Currey and K.R. Munkittrick. 2004. Adult white sucker show limited mobility near point source discharges in a large Canadian river. *In* D.L. Borton, T.J. Hall, R.P. Fisher and J.F. Thomas (eds.), *Pulp and Paper Mill Effluent Environmental Fate and Effects*. DEStech Publications, Lancaster, PA. pp. 123-132.
- Environment Canada. 1997. Fish population survey expert working group final report: Recommendations from Cycle 1 review. National EEM Office, Science Policy and Environmental Quality Branch, Environment Canada, Ottawa, Canada (EEM/1997/6).
- Environment Canada. 2004a. Revised Technical Guidance for Conducting Aquatic Environmental Effects Monitoring for pulp and paper mills. Environment Canada, National EEM Office, National Water Research Institute, Gatineau, QC., Canada.

Environment Canada. 2004b. Understanding the Pulp and Paper Environmental Effects Monitoring Program. Environment Canada, National EEM Office, National Water Research Institute, Gatineau, QC., Canada.

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. Can.* 37: 251–278.

Gurevitch, J. and L.V. Hedges. 2001. Meta-analysis. Combining the results of independent experiments. Pages 347-369 in *Design and Analysis of Ecological Experiments*, S.M. Scheiner and J. Gurevitch (eds.), Oxford University Press, New York.

Hedges, L.V. and I. Olkin. 1985. Statistical methods for meta-analysis. Academic Press, New York, N.Y.

Kilgour, B.W., K.M. Somers and D.E. Matthews. 1998. Using the normal range as a criterion for ecological significance in environmental monitoring and assessment. *Ecoscience* 5: 542-550.

Lenat, D.R. and V.H. Resh. 2001. Taxonomy and stream ecology – The benefits of genus- and species- level identifications. *J.N. Am. Benthol. Soc.* 20: 287-298.

Lowell, R.B. 1997. Discussion paper on critical effect size guidelines for EEM using benthic invertebrate communities. National EEM Office, Science Policy and Environmental Quality Branch, Environment Canada, Ottawa, Ontario (EEM/1997/8).

Lowell, R.B. and J.M. Culp. 2002. Implications of sampling frequency for detecting temporal patterns during environmental effects monitoring. *Water Qual. Res. J. Can.* 37: 119–132.

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. *Environ. Toxicol. Chem.* 19: 1182–1190.

Lowell, R.B., J.M. Culp and F.J. Wrona. 1995. Stimulation of increased short-term growth and development of mayflies by pulp mill effluent. *Environ. Toxicol. Chem.* 14: 1529–1541.

Lowell, R.B., K. Hedley and E. Porter. 2002. Data interpretation issues for Canada's Environmental Effects Monitoring Program. *Water Qual. Res. J. Can.* 37: 101–117.

Lowell, R.B., K.R. Munkittrick, J.M. Culp, M.E. McMaster, and L.C. Grapentine. 2004. National response patterns of fish and invertebrates exposed to pulp and paper mill effluents: metabolic disruption in combination with eutrophication and other effects. Pages 147-155 in *Pulp and Paper Mill Effluent Environmental Fate and Effects*, D.L. Borton, T.J. Hall, R.P. Fisher, J.F. Thomas, eds., DEStech Publications, Lancaster, PA.

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, Environment Canada, Gatineau, QC.. NWRI Contribution 03-521.

Munkittrick, K.R., S.A. McGeachy, M.E. McMaster, and S.C. Courtenay. 2002. Review of cycle 2 freshwater fish studies from the pulp and paper environmental effects monitoring program. *Water Qual. Res. J. Can.* 37(1):49-77.

Munkittrick, K. R., M.E. McMaster, G. Van Der Kraak, C. Portt, W.N. Gibbons, A. Farwell, and M. Gray. 2000. Development of methods for effects-based cumulative effects assessment using fish populations: Moose River Project. Society of Environmental Toxicology and Chemistry (SETAC) Press, Pensacola, Florida.

Munkittrick, K. R., C. Portt, G.J. Van Der Kraak, I. Smith and D. Rokosh. 1991. Impact of bleached kraft mill effluent on population characteristics, liver MFO activity and serum steroid levels of a Lake Superior white sucker (*Catostomus commersoni*) population. *Can. J. Fish. Aquat. Sci.* 48: 1371-1380.

Munkittrick, K.R., G.J. Van Der Kraak, M.E. McMaster, C.B. Portt, M.R. van den Heuvel, and M.R. Servos. 1994. Survey of receiving water environmental impacts associated with discharges from pulp mills. 2. Gonad size, liver size, hepatic EROD activity and plasma sex steroid levels in white sucker. *Environ. Toxicol. Chem.* 13: 1089-1101.

Rosenberg, M.S., D.C. Adams, and J. Gurevitch. 2000. MetaWin: Statistical software for meta-analysis. Version 2.0. Sinauer Associates, Sunderland, Massachusetts.

Rosenthal, R. 1991. Meta-analytic Procedures for Social Research (revised ed.). Sage, Newbury Park, CA.

Van Der Kraak, G.J., K.R. Munkittrick, M.E. McMaster, C.B. Portt and J.P. Chang. 1992. Exposure to bleached kraft pulp mill effluent disrupts the pituitary-gonadal axis of white sucker at multiple sites. *Toxicol. Appl. Pharmacol.* 115: 224-233.

**NATIONAL WATER RESEARCH INSTITUTE
ENVIRONMENT CANADA**

