Re-Analysis of Data from Metal Mining Effluent Regulations Environmental Effects Monitoring Program

MEND Report 4.7.2b

This work was done on behalf and sponsored by the Mine Environment Neutral Drainage (MEND) Program

May 2015

RE-ANALYSIS OF DATA FROM METAL MINING EFFLUENT REGULATION ENVIRONMENTAL EFFECTS MONITORING PROGRAM

Laurie Marczak, PhD. R2 Resource Consultants Inc. &
Carla Davidson, PhD. Endeavour Scientific Inc.

For: MMER 10-Year Review Environmental Caucus

May 2015

Executive Summary

Environment Canada (EC) is currently undertaking a 10-year review of the Federal Metal Mining Effluent Regulations (MMER) that guide the effluent deposition activities of 105 metal mines active in Canada (as of 2010.) Roughly 60 additional new mines are in various stages of approval and will likely be added to the program in coming years.

The MMER not only stipulate contaminant discharge limits for arsenic, copper, cyanide, lead, nickel, zinc, total suspended solids, and radium 226, but require mines to perform Environmental Effects Monitoring (EEM) field biological studies every three years. In these studies, mines test endpoints such as fish age, weight, condition, and gonad and liver size downstream of mines to assess whether their activities are having an effect on the receiving environment. In particular, the EEM studies are designed to detect whether there's an effect on the downstream environment, even when the mines are in compliance with the regulations. The results from these studies, when considered on the national level, allow for continuous improvement of the MMER guidelines.

The EEM studies address two important but separate questions:

1) Is any one particular mine having an effect on the receiving environment?; and,
2) are metal mines nationally having an effect on downstream environments?

As part of the 10-year review of the MMER, Environment Canada convened a multi-stakeholder group that has been tasked with providing advice to the Minister of Environment regarding changes to the Metal Mines Effluent Regulations.

To address this question, Environment Canada has undertaken two major reviews of environmental data gathered by mining companies over the course of the last decade. In 2010, it was determined that the metal mining sector was in compliance with MMER permitted discharge limits for arsenic, copper, nickel, zinc, radium 226 and pH 99% of the time (Environment Canada, 2012c). However, in the Second National Assessment of Environmental Effects Monitoring Data from Metal Mines Subjected to the Metal Mining Effluent Regulations (Environment Canada, 2012b) a meta-analysis of all EEM studies found that fish downstream of metal mines were, on average, thinner, older, and slower growing.

In the original EEM program, an effect is defined as a simple statistically significant difference between a reference and exposed site. The current technical guidance document requires only one reference and exposed site. Because it is pseudo-replicated, with sufficient power it is very likely that the study will find a statistically significant difference between the reference and exposed sites. Thus, EC recommended the implementation of Critical Effects Sizes, or a threshold difference between reference and exposed sites. The 2012 Discussion Paper states: "As statistically significant difference may not necessarily be indicative of the level of risk to the environment, from the conception of the EEM program, it was envisioned that increased monitoring efforts such as investigative studies should be focused on facilities demonstrating effects of greatest concern. Critical Effect Sizes (CES), defined as thresholds above which effects may be indicative of a potential higher risk to the environment, have been developed for the fish population and benthic invertebrate community components of the MMER-EEM program." (Environment Canada 2012c).

Given these challenges, there have been questions as to whether the conclusions drawn from the Second National Assessment are sufficiently informative, and what changes can and should be made to the EEM program.

The Mining Association of Canada undertook two further studies on the EEM data. Though neither study replicated the meta-analysis of the Second National Assessment, both questioned the methodology of the EEM studies, and suggested that the findings of the Second National Assessment were due to confounding factors. Both reports supported implementing CES, but also suggested other changes to the EEM studies, such as utilizing a Bonferroni Correction on the level of significance to account for performing multiple tests.

In order to inform our policy recommendations, we undertook our own meta-analysis of these same data to determine whether the MMER guidelines are protective of the receiving environment nationally, and how methodological changes suggested by Environment Canada and the Mining Association of Canada would affect interpretation of monitoring results. Our meta-analysis results largely confirm Environment Canada's findings regarding reduced fish condition downstream of base metal (Phase 1, 2 and 3/4) and precious metal (Phase 2 and 3/4) mines, though we find reduced weight-at-age only in Phase 2 data in base metal mines (Table 1). Contrary to Environment Canada's finding in the Second National Assessment, we found enlarged livers in Phase 1 and 2 in base metal mines and in Phase 2 in precious metal mines. The effects on gonad size were variable between phases (Table 1). We were significantly hampered by insufficient data to assess iron ore, uranium, and other metal mines in Phases 1 and 2, though results from Phase 3/4 show no effects across endpoints for other metal or uranium mines (incomplete data to assess gonad size and weight-at-age), and reduced liver size and increased condition in iron ore mines. When pooled across mine type and phase, our analysis shows lowered condition, reduced weight-at-age, and increased liver size. This effect is likely driven by base metal and precious metal mines. That the effects are detectable given the wide variety of receiving environments that are subject to MMER and thus included in this study is striking. However, caution must be afforded in interpreting these results: we were hampered by a number of data quality issues, and some mines entered the program more recently, thus our data set is under-represented by iron ore, uranium and other metal mines.

Our analysis also indicates that utilizing Critical Effects Sizes would reduce the number of failed tests by more than half. This will have a large impact on the rate of false positives in the EEM program. It has been suggested that the Bonferroni correction should also be utilized in order to correct for the inflation of false positives due to undertaking multiple tests. However, our analysis does not support the conclusion that the rate of endpoint failure of mines across Canada is driven by chance, and thus require a Bonferroni correction. Most mines that fail EEM studies do so by more than one end-point, and the number of endpoint failures, even once corrected for CES, is much higher than would be expected by chance. Furthermore, the Bonferroni correction is known to be very conservative, particularly as the number of tests increases, and inflates Type II error.

We conclude that there are measurable biological effects downstream of metal mines across Canada, particularly for base and precious metal mines. We also conclude that implementing Critical Effects Sizes will do a great deal to address the rate of false positives in EEM studies, but that there is no evidence to suggest that the rate of endpoint failure in EEM studies is driven by stochasticity, therefore implementing a Bonferroni correction is likely overly conservative and would result in the reduction of protection for the environment.

Table 1. Summary of findings of meta-analysis, by mine type. Symbols indicate whether meta-analysis found significant increase (+), decrease(-), or no change (N) in exposed areas relative to reference sites. NA indicates that there was not enough data to support a comparison. Due to low numbers, the data for phases 3 and 4 were combined.

MINE TYPE	ENDPOINT	PHASE		
		1	2	3/4
BASE METAL	CONDITION	-	-	-
	LIVER SIZE	+	+	N
	GONAD SIZE	-	+	-
	WEIGHT-AT-AGE	N	-	N
	AGE	+	N	N
PRECIOUS METAL	CONDITION	N	-	-
	LIVER SIZE	N	+	N
	GONAD SIZE	+	N	-
	WEIGHT-AT-AGE	N	N	N
	AGE	N	N	+
IRON ORE	CONDITION	N	NA	+
	LIVER SIZE	N	NA	N
	GONAD SIZE	N	NA	-
	WEIGHT-AT-AGE	NA	NA	N
	AGE	NA	NA	N
OTHER METAL	CONDITION	NA	NA	N
	LIVER SIZE	NA	NA	N
	GONAD SIZE	NA	NA	N
	WEIGHT-AT-AGE	NA	NA	N
	AGE	NA	NA	N
URANIUM	CONDITION	NA	NA	N
	LIVER SIZE	NA	NA	N
	GONAD SIZE	NA	NA	NA
	WEIGHT-AT-AGE	NA	NA	NA
	AGE	NA	NA	N

Résumé

Environnement Canada (EC) mène actuellement une revue décennale du *Réglement fédéral sur les effluents des mines de métaux* (REMM) régissant les activités de rejet des effluents des 105 mines métallifères actives au Canada (à compter de 2010). Environ 60 nouvelles mines supplémentaires sont en processus d'approbation et seront vraisemblablement ajoutées au programme dans les années à venir.

Le REMM stipule non seulement les limites de décharge de contaminants pour l'arsenic, le cuivre, le cyanure, le plomb, le nickel, le zinc, le total des solides en suspension et le radium 226, mais il exige aussi que les mines procèdent à des Études de suivi des effets sur l'environnement (ESEE) sur le terrain aux trois ans. Dans le cadre de ces études biologiques, les mines examinent, en aval de leur site d'exploitation, des paramètres comme l'âge, le poids et l'état des poissons, ainsi que la taille de leurs gonades et la grosseur de leur foie, pour évaluer si leurs activités ont des effets sur le milieu récepteur. Les ESEE sont spécialement conçues pour détecter les effets éventuels sur le milieu en aval, et ce, même si les mines respectent la réglementation en vigueur. Les résultats de ces études, lorsqu'ils sont évalués au niveau national, favorisent l'amélioration continue des lignes directrices du REMM.

Les ESEE portent sur deux questions distinctes essentielles :

1) Une mine donnée a-t-elle des effets sur le milieu récepteur?
2) Les mines de métaux ont-elles des effets d'ordre national sur les milieux en aval?

Dans le cadre de la revue décennale du REMM, Environnement Canada a créé un groupe multilatéral chargé d'offrir des conseils au ministre de l'Environnement sur les changements à apporter au *Règlement sur les effluents des mines de métaux*.

Pour se pencher sur cette question, Environnement Canada a entrepris deux revues d'envergure des données environnementales recueillies par des sociétés minières au cours de la dernière décennie. En 2010, le secteur d'extraction des métaux a été jugé conforme aux limites de décharge permises par le REMM pour l'arsenic, le cuivre, le nickel, le zinc, le radium 226 et le pH dans 99 % des cas (Environnement Canada, 2012c). Durant la Deuxième évaluation nationale des données des études de suivi des effets sur l'environnement des mines de métaux visées par le Règlement sur les effluents des mines de métaux (Environnement Canada, 2012b), une méta-analyse de toutes les ESEE a cependant permis de découvrir que les poissons en aval des mines de métaux étaient, en moyenne, plus maigres et plus vieux, et que leur croissance était plus lente.

Dans le cadre du programme d'ESEE d'origine, un effet est simplement défini comme une différence significative entre une zone de référence et un site exposé. Le guide technique actuel n'exige qu'un site de référence et qu'un site exposé seulement. Puisqu'elle est pseudo-dupliquée, il est très probable que l'étude, avec suffisamment de pression, mette à jour une différence significative entre les sites de références et les sites exposés. EC a donc recommandé que des seuils critiques des effets, ou une différence seuil entre les sites de références et les sites exposés, soient mis en application. Le document de travail de 2012 déclare ce qui suit :

« Dans la mesure où une différence statistiquement significative n'indique pas nécessairement un niveau de risque pour l'environnement, selon la conception du Programme d'études de suivi des effets sur l'environnement, on a envisagé que des efforts de suivi accrus (comme des études approfondies)

doivent être axés sur les mines démontrant les effets les plus préoccupants. Des seuils critiques d'effet (définis comme étant des seuils au-delà desquels les effets peuvent indiquer un risque potentiel plus élevé pour l'environnement) ont été mis au point pour les volets d'étude portant sur la population de poissons et la communauté d'invertébrés benthiques du Programme d'études de suivi des effets sur l'environnement du Règlement sur les effluents des mines de métaux » (Environnement Canada, 2012c).

À la lumière de ces défis, on se demande si les conclusions de la Deuxième évaluation nationale sont suffisamment informatives et quels changements peuvent et doivent être apportés au programme d'ESEE.

L'Association minière du Canada a lancé deux autres études sur les données d'ESEE. Même si aucune n'a dupliqué la méta-analyse de la Deuxième évaluation nationale, ces deux études ont toutefois remis la méthodologie des ESEE en cause et suggéré que les résultats de la Deuxième évaluation nationale découlaient de facteurs de confusion. Les deux rapports appuyaient la mise en application des SCE, tout en suggérant que d'autres changements soient apportés aux ESEE, notamment l'utilisation de la correction de Bonferroni sur le niveau de signification pour justifier l'exécution de tests multiples.

Pour renseigner nos recommandations en matière de politique, nous avons entrepris notre propre méta-analyse desdites données pour déterminer si les lignes directrices du REMM protègent le milieu récepteur à l'échelle nationale et comment les changements méthodologiques suggérés par Environnement Canada et l'Association minière du Canada influenceraient l'interprétation des résultats des études de suivi. Les résultats de notre méta-analyse confirment largement les constatations d'Environnement Canada relatives à la détérioration de l'état des poissons en aval des mines de métaux communs (phases 1, 2 et 3/4) et précieux (phases 2 et 3/4), bien que nous n'ayons observé une baisse du poids selon l'âge que dans les données de phase 2 des mines de métaux communs (tableau 1). Contrairement aux résultats d'Environnement Canada tirés de la Deuxième évaluation nationale, nous avons observé des hypertrophies du foie durant les phases 1 et 2 pour les mines de métaux communs et la phase 2 pour les mines de métaux précieux. Les effets sur la taille des gonades variaient d'une phase à l'autre (tableau 1). Nous avons été fortement ralentis par le manque de données à disposition pour évaluer les mines de minerai de fer, d'uranium et d'autres métaux durant les phases 1 et 2 de l'analyse, tandis que les résultats de phase 3/4 n'ont démontré aucun effet sur les paramètres pour les mines d'uranium ou d'autres métaux (données incomplètes pour évaluer la taille des gonades et le poids selon l'âge) et confirmé la réduction de la taille du foie et l'amélioration de l'état pour les mines de minerai de fer. Une fois regroupées par type de mine et phase, les données de notre analyse démontrent la détérioration de l'état, la réduction du poids selon l'âge et la hausse de la grosseur du foie. Cet effet est vraisemblablement provoqué par les mines de métaux communs et précieux. Il est frappant de noter que les effets sont détectables malgré la vaste gamme de milieux récepteurs qui sont soumis au REMM et conséquemment inclus dans la présente étude. Ces résultats doivent cependant être interprétés prudemment : nous avons été ralentis par un certain nombre de problèmes relatifs à la qualité des données et quelques mines ne sont que depuis récemment incluses dans le programme. Notre ensemble de données est ainsi sous-représenté pour les mines de minerai de fer, d'uranium et d'autres métaux.

Notre analyse indique aussi que l'utilisation des seuils critiques des effets (SCE) réduirait de plus de moitié le nombre d'essais comportant une cote d'échec. Cela aura une grande incidence sur le taux de résultats faussement positifs obtenus dans le cadre du programme d'ESEE. L'utilisation de la correction de Bonferroni a aussi été suggérée pour corriger le gonflement des résultats faussement positifs associé à l'exécution de tests multiples. Notre analyse n'appuie cependant pas la conclusion voulant que le non-respect des paramètres par les mines au Canada soit attribuable au hasard, requérant ainsi l'utilisation d'une correction de Bonferroni. La plupart des mines jugées non conformes dans le cadre d'ESEE le sont par rapport à plus d'un paramètre et le nombre d'échecs connexes, même corrigés pour correspondre aux SCE, est grandement supérieur à celui qu'on pourrait attribuer au hasard. De plus, la correction de Bonferroni est réputée très conservatrice, particulièrement lorsque le nombre de tests augmente et gonfle l'erreur de type II.

Nous concluons ainsi qu'il y a bel et bien des effets biologiques mesurables en aval des mines de métaux à l'échelle du Canada, surtout des mines de métaux communs et précieux. Nous concluons aussi que la mise en pratique des seuils critiques des effets influencera avantageusement le taux de résultats faussement positifs des ESEE, mais qu'aucune preuve ne suggère que le non-respect des paramètres des ESEE est dicté par la stochasticité. La mise en pratique de la correction de Bonferroni serait ainsi vraisemblablement trop conservatrice et provoquerait la baisse de la protection du milieu.

Tableau 1. Sommaire des résultats de la méta-analyse par type de mine. Les symboles indiquent si la méta-analyse a permis de mesurer une hausse (+) ou une baisse (-) importante, ou bien nul changement (N) dans les régions exposées par rapport aux sites de référence. S.O. signifie que les données étaient insuffisantes pour procéder à une comparaison. Leur nombre étant insuffisant, les données des phases 3 et 4 ont été combinées.

TYPE DE MINE	PARAMÈTRE	PHASE		
		1	2	3/4
MÉTAUX COMMUNS	ÉTAT	-	-	-
	GROSSEUR DU FOIE	+	+	N
	TAILLE DES GONADES	-	+	-
	POIDS SELON L'ÂGE	N	-	N
	ÂGE	+	N	N
MÉTAUX PRÉCIEUX	ÉTAT	N	-	-
	GROSSEUR DU FOIE	N	+	N
	TAILLE DES GONADES	+	N	-
	POIDS SELON L'ÂGE	N	N	N
	ÂGE	N	N	+
MINERAI DE FER	ÉTAT	N	S.O.	+
	GROSSEUR DU FOIE	N	S.O.	N
	TAILLE DES GONADES	N	S.O.	-
	POIDS SELON L'ÂGE	S.O.	S.O.	N
	ÂGE	S.O.	S.O.	N
AUTRE MÉTAUX	ÉTAT	S.O.	S.O.	N
	GROSSEUR DU FOIE	S.O.	S.O.	N
	TAILLE DES GONADES	S.O.	S.O.	N
	POIDS SELON L'ÂGE	S.O.	S.O.	N
	ÂGE	S.O.	S.O.	N
URANIUM	ÉTAT	S.O.	S.O.	N
	GROSSEUR DU FOIE	S.O.	S.O.	N
	TAILLE DES GONADES	S.O.	S.O.	S.O.
	POIDS SELON L'ÂGE	S.O.	S.O.	S.O.
	ÂGE	S.O.	S.O.	N

Background and Problem Description

Under the Metal Mining Effluent Regulations (2002) of the Fisheries Act, individual mines are required to undertake periodic Environmental Effects Monitoring (EEM). These assessments are intended to evaluate whether any one particular mine is having an effect on the receiving environment, but also whether the MMER, as drafted, are sufficiently protective of the environment at the national level. These studies include surveys of fish populations, benthic invertebrate communities and mercury accumulation in fish tissues upstream and downstream of mines. EEM plans are drafted following the guidelines developed by Environment Canada (Environment Canada, 2012a) and each individual study plan and any proposed variant, is reviewed by EC.

As of 2014, Environment Canada has completed two national level assessments of the data generated by this monitoring program and these reports suggest that there are observable impacts of metal mine effluent on target biological endpoints in receiving waters. Revisions to end of pipe discharge limits for some parameters may be justified if the 3rd National Assessment currently underway concludes that the MMER, as currently written, do not adequately protect the environment.

Both the Mining Association of Canada and the several environmental NGOs are concerned that any decisions made during the 10 year review of these regulations need to be based on the best available data as significant concerns over data quality and analysis have been raised during prior reviews.

This report presents a re-analysis of the fish endpoint data from all available phases of Environmental Effects Monitoring that have been undertaken by mining companies regulated under the MMERs over the past decade. We intended to investigate four questions:

1. Are there consistent differences in EEM end-points downstream of metal mine outfalls?
2. Are there correlations between changes in particular end-points?
3. Will changing decision-making thresholds in the EEM process for individual mines result in a reduction in protection to the environment?
4. Do significantly affected end-points correlate with studies of acute or sub-lethal toxicity, or with compounds released by mines?

In the present analysis, we restricted our assessment to fish population data only – invertebrate community analysis and fish toxicology results are not examined. We evaluated the available data for mine effects on fish growth, reproduction, condition and survival using the five biological endpoints and their Critical Effects Sizes (CES) identified by Environment Canada:

- Size-at-age (body weight relative to age) CES = +/-25%
- Relative gonad size (gonad weight to body weight) CES = +/-25%
- Condition (body length to body weight) CES = +/-10%
- Relative liver size (liver weight to body weight) CES = +/-25%
- Age CES = +/-25%

Previous Re-evaluations

In the first National Assessment (Lowell et al. 2007), Environment Canada concluded from its meta-analysis of effect sizes that the liver size of fish was reduced in exposure areas. The second National Assessment (Environment Canada, 2012b), which combined data from two cycles of monitoring, also concluded that liver size of fish was reduced in exposure areas and also that fish in exposure areas were significantly older, thinner and slower growing with a trend of smaller gonad size.

Previous reviews questioned some of the methods employed by Environment Canada in their meta-analysis. In particular, the question of whether effect sizes needed to be calibrated (Huebert and McKernan 2014; Nakagawa 2012) and that a multivariate approach to the analysis of effect sizes would be more appropriate (Nakagawa 2012). Kilgour & Associates Ltd (2014) conducted an evaluation of the national assessment reports for the Mining Association of Canada but focused on the methodology and study design underlying the data architecture.

A recent evaluation by Huebert and McKernan (2014) of Stantec Consulting Ltd. was not able to duplicate the methods of the meta-analysis due to a lack of pairing in the data they received which eliminated the potential to construct effect sizes in a formal meta-analysis. Instead Huebert and McKernan used the pooled mean values for each fish endpoint to determine whether a statistically significant difference (for the same endpoint and in the same direction) is confirmed by two consecutive studies. Effect sizes were calculated for each fish endpoint by dividing the absolute difference between the two adjusted means by the weighted average for that endpoint – each effect size is thus a percentage of the pooled mean. This procedure gives effect sizes that represent the magnitude of a given change but not whether the change was an increase or a decrease. Effect sizes were subsequently standardized by dividing by the EEM critical effect size value for each endpoint. This standardization is meant to “calibrate” the effect sizes so that they can be directly compared to each other.

Methods and Data QC

Huebert and McKernan (2014) used the means or adjusted means from the ANOVA or ANCOVA results within tables provided by Environment Canada. They also identified a number of data quality issues that hampered their re-evaluation, most critically, the absence of markers indicating which means were for reference or exposure sites. We were able to obtain the raw data including identifiers of site type (reference or exposure) and base our analysis on this instead of the output from previous analyses.

At the same time, the analyses undertaken here (and presumably those conducted by the first and second National Assessment reports) were hindered by some substantial deficiencies in the raw data. An extensive QA/QC process was undertaken to address as many of these issues as possible, however, the quality of the remaining data are almost certainly patchy. Data issues included: inadequate reporting and/or missing data, obvious data entry errors, presence of extreme outliers, inclusion of immature and unidentifiable fish in the data, insufficient sample sizes, inclusion of multiple fish species at certain sites (above the mandated two sentinel species) and the apparent retention of dummy or placeholder data in submitted fields. We made the following edits:

- We utilized only the two most abundant species per mine.
- Removed immature fish
- Removed studies with fewer than 12 fish per species
- Where fish length was not recorded, used tail length (as per Environment Canada)
- Used “nearfield” sites as exposed, deleted “farfield”
- Removed all non-lethal records.
- Endpoints that showed significant interaction in ANCOVA removed from effect size analysis.

- The four mines which have done Phase 4 studies (two precious metal, one uranium and one base metal mine) were merged with the Phase 3 studies.
- Obvious data entry errors, such as dummy variables, or physiologically impossible gonad or liver sizes, were removed from analysis.

In discussions with Environment Canada it became clear that a critical source of information for addressing these issues is access to the interpretive reports submitted by the consultants undertaking the studies. The data for some of the MMER studies still reside in paper copies, and EC was able to assess the interpretive reports as a source of data as well as use these reports to reconcile discrepancies in the data. Therefore, differences between our findings and those of Environment Canada may be due to these data issues which we could not reconcile.

Additional data transformation or rectification was also required. Mines reported a finer grain of habitat categories than were assessed during the second National Assessment. We combined Stream and Creek; River Erosional and River Depositional; Lake Erosional and Lake Depositional; and Estuary and Marine Intertidal categories.

Given the issues with the underlying data, for our preliminary analysis of these data we modified our approach to ensure that our methods paralleled those of EC as much as possible. This allowed us to compare our results to those obtained by EC's First and Second National Assessments by restricting discrepancies between our results and those of Environment Canada to differences in underlying data structure rather than data editing decisions. As a result, we made the following amendments to our original approach to mirror those of EC:

1. In our initial approach we determined the relationship between tail length and fish length for studies which reported both for individual fish and then interpolated the missing data in studies which reported only one measure of length. In consultation with Environment Canada we determined that their analyses had used a simple substitution of one measure for the other where data gaps occurred. Once fish from non-lethal surveys were excluded few instances of missing tail length occurred, however, we recommend that in future analyses our original approach be used.
2. We chose to utilize Hedge's d as the metric for this analysis, rather than response ratios, in order to maintain comparability with EC. In general terms, the response ratio (natural log of the mean of the exposure sites over the reference sites or lnR) is often preferable as it is more naturally interpretable – as the proportionate change resulting from experimental manipulation – and has preferential statistical properties. However, Hedge's d has not been completely invalidated as an approach, and many commenters, including in previous reviews of these data, recommend using both Hedge's d and response ratios. In future work, once the identified data issues are addressed, we recommend utilizing both approaches to better understand national trends in Metal Mine Effluent Effects.
3. Comparisons that showed a significant interaction were removed from the meta-analysis.
4. We used the same analysis software as EC (MetaWin 2.0) which has the limitation of not allowing mixed effects modeling or the inclusion of interactions between terms.

ANCOVAs were performed on data on adult fish for each species, sex and phase at each mine, for the endpoints Liver Weight, Gonad Weight, Length (Condition), and Weight at Age, using Fish Weight as a covariate. ANOVAs were used to assess Age. Least squared means and standard deviation from ANCOVAs were generated for adult male and adult female fish of each species recorded at each mine site in both reference and exposure sites for the endpoints Liver Weight, Gonad Weight, Condition, and Weight at Age. These least squared means and standard deviations were then used to calculate effect sizes as the standardized mean difference between the exposure and reference sites or Hedge's d for all groups with a sample size ≥12 per the national assessment guidelines.

Question 1: Are there consistent differences in EEM endpoints downstream of metal mine outfalls?

Statistical assumptions in meta-analysis

We used a combination of funnel plots, normal quantile plots and normal histograms to confirm that all data used met standard statistical assumptions for meta-analyses. The cumulative effect size represents the overall magnitude of the effect present in the studies entered in a meta-analysis and together with its confidence interval can be used to determine the significance and magnitude of support for a particular hypothesis. A cumulative effect size is necessarily an average of underlying effects and as such the total heterogeneity of a sample of effect sizes, Q_T, should be calculated to determine if a given sample is homogeneous. Tested against a χ2-distribution – a significant Q_T indicates that variance in effect sizes is greater than expected due to sampling error alone and that additional explanatory variables should be explored (Rosenberg et al. 2000). In the present analysis, the cumulative effect size across all endpoints and all study phases was Q_T = 5414.7, df = 626 and this effect contained significant variance indicative of further variability (P(χ2) <0.0001).

Are there significant effects for identified fish endpoints?

We subdivided the cumulative mean for all studies between biological endpoints and examined the resulting effect sizes. However all endpoints have Q_W (analogous to the between group variance in ANOVA, or heterogeneity not explained by the defined parameters) suggestive of further data partitioning (Table 2).

Table 2. Mean effect sizes (E+) and confidence intervals for each endpoint cumulative for all time periods (phases). Endpoints highlighted in red suggest that exposed fish have endpoints that are significantly decreased relative to reference fish, while those in green have endpoints that are increased.

FISH ENDPOINT	NUMBER OF STUDIES	Q_W	DF	PROB (X^2)	E+	95% CONFIDENCE INTERVAL
RELATIVE LIVER SIZE	109	839.96	108	<0.0001	0.1107	0.0723 to 0.1491
CONDITION (RELATIVE BODY SIZE)	100	718.16	99	<0.0001	-0.15	-0.2447 to -0.0486
RELATIVE GONAD SIZE	108	530.54	107	<0.0001	-0.029	-0.0664 to 0.0092
AGE	99	527.05	98	<0.0001	0.0185	-0.1033 to 0.1462
WEIGHT AT AGE	96	370.74	95	<0.0001	-0.05	-0.0920 to -0.0074

The assessment conducted by Environment Canada during the First National Assessment (Lowell et al. 2007) concluded that, on average, exposure area fish exhibited significantly lowered condition and relative liver size while there was no significant effect documented for the other endpoints (95% CI overlapped zero). The assessment conducted here (Table 1) draws the same conclusion with respect to fish body condition (fish in exposure areas were thinner) but enlarged livers for fish measured in exposure sites. Effect sizes for relative gonad size, weight at age and mean age had 95% CI that overlapped zero.

Where the Second National Assessment (Environment Canada, 2012b) reported that exposure area fish showed significantly reduced condition and relative liver size, our assessment concluded that there was a significant effect on condition while relative liver size was larger for exposure area fish. The Second National Assessment identified further lowered growth (weight-at-age) and older fish in exposure areas. Our assessment found similar lowered values for weight-at-age but no trend for older fish in exposure areas.

We further subdivided the variance in effect sizes between the same categorical groupings explored by the national assessment reports – mine type; receiving habitat type and the sex of adult fish – and Table 3 contains a comprehensive listing of effect sizes with associated 95% confidence intervals. Results at this level of categorical assessment are again variable. Overall trends indicate negative effects on condition, increased liver size, and variable effects on gonad size. Reduced size and enlarged livers are most pronounced in base metal and precious metal mines, and mines that deposit in rivers and lakes. Reduced condition is prevalent in studies from all phases, liver enlargement is present in Phase 1, but most common in Phase 2 studies, and not discernable in studies from Phase 3 or 4. Gonad size varies between significantly larger and significantly smaller from phase to phase, and between mine types. Finally, a significant reduction in weight-at-age is discernible in Phase 2, but not the other phases (Table 3). Taken in total, these results suggest that there is evidence for an impact on condition and liver size nationally, primarily in base and precious metal mines.

Table 3. Meta-analysis results by mine, mine type, phase, sex and habitat.

PHASE 1						
CONDITION	Group	#Studies	E+	df	95% CI up	95% CI down
Mine Type	Base Metal	13	-0.1913	12	-0.3186	-0.0641
	Precious Metal	14	0.0179	13	-0.0872	0.123
	Iron Ore	4	-0.1333	3	-0.2143	0.5947
	Other Metal	-	-	-	-	-
	Uranium	-	-	-	-	-
Sex	Adult Female	15	-0.0668	14	-0.1701	0.0366
	Adult Male	16	-0.023	15	-0.1374	0.0914
Habitat	River	14	-0.0807	13	-0.2051	0.0438
	Lake	8	0.0582	7	-0.1255	0.2418
	Creek	2	0.1708	1	-1.8725	2.2141
	Marine/Estuary	4	0.1262	3	-0.2413	0.4936
	Pond	3	-0.1938	2	-0.5006	0.1129
	Wetland	-	-	-	-	-
PHASE 2						
CONDITION	Group	#Studies	E+	df	95% CI up	95% CI down
Mine Type	Base Metal	9	-0.5607	8	-0.7052	-0.4162
	Precious Metal	13	-0.2333	12	-0.3380	-0.1286
	Iron Ore	-	-	-	-	-
	Other Metal	-	-	-	-	-
	Uranium	-	-	-	-	-
Sex	Adult Female	13	-0.2352	12	-0.3526	-0.1177
	Adult Male	9	-0.4741	8	-0.5985	-0.3497
Habitat	River	9	-0.6928	8	-0.8504	-0.5351
	Lake	8	0.0003	7	-0.1605	0.1610
	Creek	3	-0.4421	2	-0.7826	-0.1015
	Marine/Estuary	-	-	-	-	-
	Pond	-	-	-	-	-
	Wetland	2	-0.2567	1	-1.541	1.0276
PHASE 3 AND 4						
CONDITION	Group	#Studies	E+	df	95% CI up	95% CI down
Mine Type	Base Metal	10	-0.3856	9	-0.5434	-0.2279
	Precious Metal	23	-0.0778	22	-0.1533	-0.0023
	Iron Ore	9	0.2287	8	0.0920	0.3655
	Other Metal	2	-0.7632	1	-2.1841	0.6576
	Uranium	3	-0.1794	2	-0.6056	0.2469
Sex	Adult Female	25	-0.1007	24	-0.1746	-0.0269
	Adult Male	22	-0.114	21	-0.1959	-0.0321
Habitat	River	20	-0.3578	19	-0.4597	-0.2558
	Lake	23	-0.0133	22	-0.0833	0.0568
	Creek	4	0.0802	3	-0.2015	0.3618
	Marine/Estuary	-	-	-	-	-
	Pond	-	-	-	-	-
	Wetland	-	-	-	-	-

PHASE 1						
RELATIVE LIVER SIZE	Group	#Studies	E+	df	95% CI up	95% CI down
Mine Type	Base Metal	15	0.1845	14	0.0711	0.2979
	Precious Metal	18	-0.0228	17	-0.1318	0.0862
	Iron Ore	3	-0.2957	2	-0.9247	0.3332
	Other Metal	-	-	-	-	-
	Uranium	-	-	-	-	-
Sex	Adult Female	17	0.07	16	-0.0382	0.1783
	Adult Male	19	0.0424	18	-0.0632	0.148
Habitat	River	16	-0.0658	15	-0.1822	0.0507
	Lake	13	0.0612	12	-0.0641	0.1865
	Creek	2	0.0698	1	-1.9728	2.1124
	Marine/Estuary	2	0.4161	1	-1.7933	2.6255
	Pond	3	0.4361	2	-0.0769	0.9491
	Wetland	-	-	-	-	-
PHASE 2						
RELATIVE LIVER SIZE	Group	#Studies	E+	df	95% CI up	95% CI down
Mine Type	Base Metal	10	0.6842	9	0.5423	0.8261
	Precious Metal	10	0.2148	9	0.0672	0.3623
	Iron Ore	-	-	-	-	-
	Other Metal	-	-	-	-	-
	Uranium	-	-	-	-	-
Sex	Adult Female	13	0.3662	12	0.2452	0.4871
	Adult Male	7	0.6409	6	0.4503	0.8315
Habitat	River	8	0.7598	7	0.5952	0.9243
	Lake	8	0.0920	7	-0.0760	0.2601
	Creek	4	0.5813	3	0.2357	0.9269
	Marine/Estuary	-	-	-	-	-
	Pond	-	-	-	-	-
	Wetland	-	-	-	-	-
PHASE 3 AND 4						
RELATIVE LIVER SIZE	Group	#Studies	E+	df	95% CI up	95% CI down
Mine Type	Base Metal	9	-0.0525	8	-0.2204	0.1153
	Precious Metal	31	0.0076	30	-0.0629	0.0780
	Iron Ore	9	-0.0097	8	-0.1605	0.1411
	Other Metal	2	0.3772	1	-1.2853	2.0396
	Uranium	2	0.1417	1	-1.2810	1.5644
Sex	Adult Female	26	-0.0029	25	-0.0813	0.0755
	Adult Male	27	0.0415	26	-0.0356	0.1186
Habitat	River	19	-0.2093	18	-0.3178	-0.1007
	Lake	20	-0.0349	19	-0.1185	0.0486
	Creek	14	0.3205	13	0.2124	0.4287
	Marine/Estuary	-	-	-	-	-
	Pond	-	-	-	-	-
	Wetland	-	-	-	-	-

PHASE 1						
GONAD SIZE	Group	#Studies	E+	df	95% CI up	95% CI down
Mine Type	Base Metal	14	0.2778	13	0.1539	0.4017
	Precious Metal	15	-0.2564	14	-0.3767	-0.1361
	Iron Ore	4	-0.0308	3	-0.4372	0.3756
	Other Metal	-	-	-	-	-
	Uranium	-	-	-	-	-
Sex	Adult Female	14	-0.0718	13	-0.1957	0.052
	Adult Male	19	0.0606	18	-0.0474	0.1685
Habitat	River	13	0.023	12	-0.1059	0.152
	Lake	13	0.0216	12	-0.1135	0.1566
	Creek	2	0.0606	1	-1.9787	2.1
	Marine/Estuary	4	-0.0014	3	-0.3695	0.3666
	Pond	-	-	-	-	-
	Wetland	-	-	-	-	-
PHASE 2						
GONAD SIZE	Group	#Studies	E+	df	95% CI up	95% CI down
Mine Type	Base Metal	9	0.4625	8	0.3185	0.6064
	Precious Metal	9	-0.0243	8	-0.1819	0.1333
	Iron Ore	-	-	-	-	-
	Other Metal	-	-	-	-	-
	Uranium	-	-	-	-	-
Sex	Adult Female	11	0.1402	10	0.01	0.2704
	Adult Male	7	0.4071	6	0.2236	0.5907
Habitat	River	5	0.3996	4	0.1844	0.6147
	Lake	8	0.3376	7	0.1642	0.511
	Creek	3	-0.252	2	-0.7682	0.2642
	Marine/Estuary	-	-	-	-	-
	Pond	-	-	-	-	-
	Wetland	2	0.033	1	-1.7884	1.8545
PHASE 3 AND 4						
GONAD SIZE	Group	#Studies	E+	df	95% CI up	95% CI down
Mine Type	Base Metal	10	-0.2229	9	-0.3797	-0.0660
	Precious Metal	33	-0.0748	32	-0.1397	-0.0100
	Iron Ore	10	-0.2397	9	-0.3794	-0.1000
	Other Metal	3	-0.0900	2	-0.5092	0.3292
	Uranium					
Sex	Adult Female	28	-0.0949	27	-0.1679	-0.0219
	Adult Male	29	-0.1471	28	-0.2195	-0.0748
Habitat	River	19	-0.2441	18	-0.354	-0.1343
	Lake	23	-0.1391	22	-0.2125	-0.0657
	Creek	15	0.0171	14	-0.0865	0.1207
	Marine/Estuary	-	-	-	-	-
	Pond	-	-	-	-	-
	Wetland	-	-	-	-	-

PHASE 1						
AGE	Group	#Studies	E+	df	95% CI up	95% CI down
Mine Type	Base Metal	9	0.3155	8	0.1032	0.5277
	Precious Metal	15	-0.0274	14	-0.1916	0.1367
	Iron Ore	2	0.4047	1	-2.7633	3.5727
	Other Metal	-	-	-	-	-
	Uranium	-	-	-	-	-
Sex	Adult Female	13	0.1984	12	0.0224	0.3743
	Adult Male	13	0.0571	12	-0.1199	0.2342
Habitat	River	8	0.4088	7	0.1614	0.6563
	Lake	-	-	-	-	-
	Creek	11	0.0964	10	-0.0953	0.2882
	Marine/Estuary	3	0.0342	2	-0.7168	0.7853
	Pond	4	-0.2708	3	-0.7414	0.1998
	Wetland	-	-	-	-	-
PHASE 2						
AGE	Group	#Studies	E+	df	95% CI up	95% CI down
Mine Type	Base Metal	13	-0.0236	12	-0.1920	0.1448
	Precious Metal	7	-0.1770	6	-0.4425	0.0885
	Iron Ore	-	-	-	-	-
	Other Metal	-	-	-	-	-
	Uranium	-	-	-	-	-
Sex	Adult Female	12	-0.0034	11	-0.1832	0.1763
	Adult Male	8	-0.1803	7	-0.4139	0.0533
Habitat	River	9	-0.1286	8	-0.3304	0.0732
	Lake	7	-0.0823	6	-0.3617	0.1971
	Creek	3	0.2313	2	-0.5210	0.9837
	Marine/Estuary	-	-	-	-	-
	Pond	-	-	-	-	-
	Wetland	-	-	-	-	-
PHASE 3 AND 4						
AGE	Group	#Studies	E+	df	95% CI up	95% CI down
Mine Type	Base Metal	8	0.1817	7	-0.0617	0.4250
	Precious Metal	29	0.1376	28	0.0326	0.2427
	Iron Ore	10	-0.2961	9	-0.5141	-0.0780
	Other Metal	2	-0.3742	1	-2.7491	2.0007
	Uranium	4	-0.3553	3	-0.7413	0.0308
Sex	Adult Female	28	-0.0376	27	-0.1472	0.0720
	Adult Male	25	0.0485	24	-0.0654	0.1624
Habitat	River	16	0.0774	15	-0.0835	0.2382
	Lake	23	-0.0453	22	-0.1615	0.0709
	Creek	14	0.0194	13	-0.1396	0.1785
	Marine/Estuary	-	-	-	-	-
	Pond	-	-	-	-	-
	Wetland	-	-	-	-	-

PHASE 1						
WEIGHT AT AGE	Group	#Studies	E+	df	95% CI up	95% CI down
Mine Type	Base Metal	12	-0.106	11	-0.2505	0.0384
	Precious Metal	15	0.1006	14	-0.0182	0.2195
	Iron Ore	3	0.1327	2	-0.4835	0.7488
	Other Metal	-	-	-	-	-
	Uranium	-	-	-	-	-
Sex	Adult Female	12	0.1311	11	-0.0142	0.2763
	Adult Male	18	-0.0411	17	-0.1497	0.0676
Habitat	River	10	0.0592	9	-0.093	0.2114
	Lake	12	-0.043	11	-0.1803	0.0944
	Creek	2	0.0415	1	-4.0073	4.0904
	Marine/Estuary	3	-0.1005	2	-0.7156	0.5147
	Pond	3	0.2442	2	-0.2703	0.7587
	Wetland	-	-	-	-	-
PHASE 2						
WEIGHT AT AGE	Group	#Studies	E+	df	95% CI up	95% CI down
Mine Type	Base Metal	10	-0.2689	9	-0.4014	-0.1363
	Precious Metal	11	-0.1046	10	-0.261	0.0518
	Iron Ore	-	-	-	-	-
	Other Metal	-	-	-	-	-
	Uranium	-	-	-	-	-
Sex	Adult Female	10	-0.2114	9	-0.3596	-0.0632
	Adult Male	11	-0.1925	10	-0.3303	-0.0547
Habitat	River	9	-0.2882	8	-0.4393	-0.1371
	Lake	8	-0.202	7	-0.3816	-0.0224
	Creek	3	0.0415	2	-0.5008	0.5838
	Marine/Estuary	-	-	-	-	-
	Pond	-	-	-	-	-
	Wetland					
PHASE 3 AND 4						
WEIGHT AT AGE	Group	#Studies	E+	df	95% CI up	95% CI down
Mine Type	Base Metal	24	-0.0061	23	-0.0884	0.0763
	Precious Metal	9	0.0763	8	-0.0864	0.239
	Iron Ore	2	-0.4646	1	-2.1575	1.2284
	Other Metal	2	-0.3478	1	-1.9544	1.2587
	Uranium					
Sex	Adult Female	23	-0.0385	22	-0.1242	0.0473
	Adult Male	22	-0.0042	21	-0.0954	0.087
Habitat	River	16	-0.0594	15	-0.1797	0.061
	Lake	19	-0.016	18	-0.1069	0.075
	Creek	10	0.0095	9	-0.1314	0.1504
	Marine/Estuary	-	-	-	-	-
	Pond	-	-	-	-	-
	Wetland	-	-	-	-	-

As detailed in the Second National Assessment (Environment Canada, 2012b), investigations of each endpoint between Phases shows differences in magnitude and direction of effect size. We confirm this general finding, and note consistent reductions in condition, a reduction in weight at age in phase 2 studies, and enlargement of liver sizes in phase 1 and phase 2 studies. To examine the distributions of effects more closely, we plotted the distribution of ANCOVA calculated response ratios, using weight as a covariate, (ln(exposed/reference)) for all studies (Figure 1.)

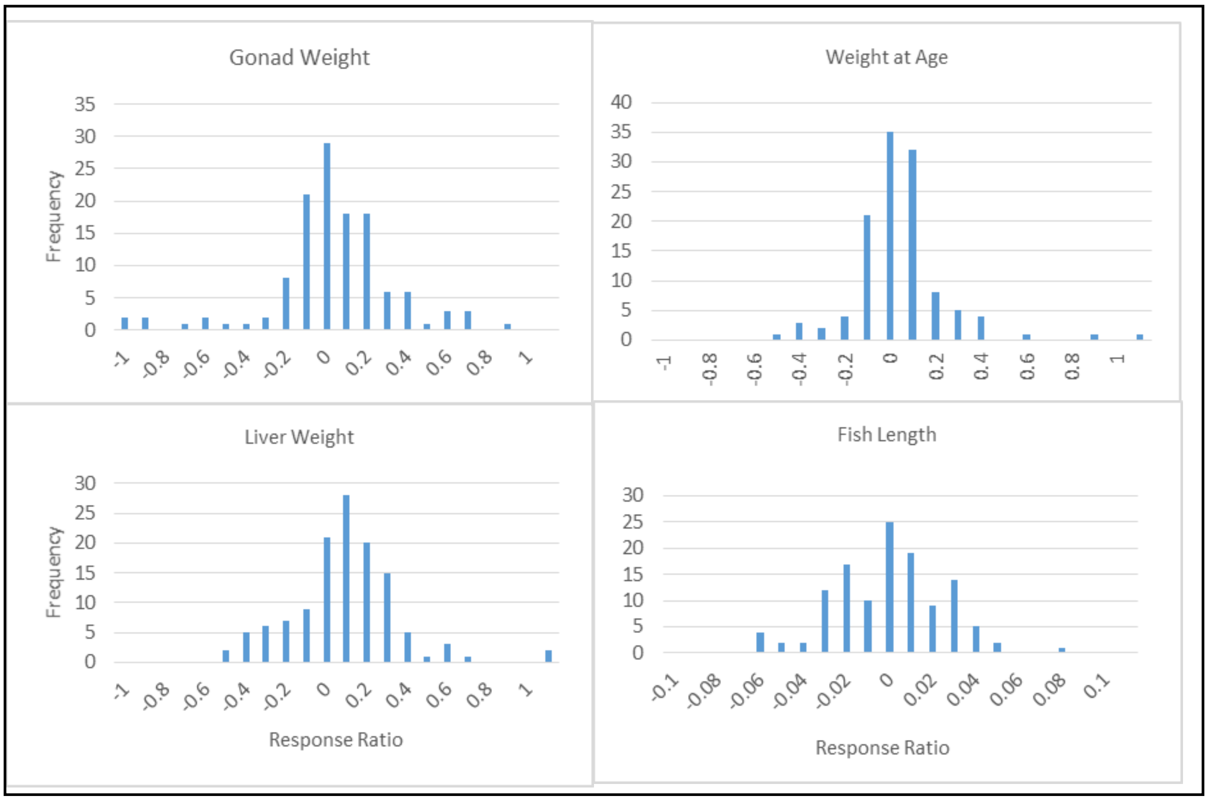

Figure 1. Frequency Histograms of response ratios (ln(exposed/reference)) ancova calculated (weight as covariate) effects sizes for measures: Gonad Weight, Fish Length (Condition), Weight at Age, and Liver Weight. The range of values for Fish Length is narrowed relative to other measures to better show the distribution.

Consistent with our meta-analysis results, we find that gonad weight remains centered around 0 though with few higher ratios (median = -0.027), and weight at age is also strongly centered around 0 (median = -0.030), but that the liver weight shows a general trend towards larger ratios (median = 0.035), indicating the presence of a number of mines whose fish show larger livers in exposed regions. Finally, the distribution of fish length (Condition) is strongly clustered around 0 (median = -0.0029), reflecting this end-point's relative stability.

Differences remain between this analysis and that of the Second National Assessment, primarily in our finding that liver size in enlarged in exposed areas rather than smaller. It is difficult to determine the extent to which these differences in outcome are due to differences in the underlying data sets employed (see discussion in methods). We don't believe that further consultation on the data sets with

Environment Canada to resolve data issues will address the discrepancies; likely the availability of interpretive reports would address some differences, but these data are not available to us.

Question 2: Are There Correlations Among Results at Endpoints?

Concern has been expressed about the problem of Type I errors resulting from multiple testing of the endpoints, or False Discovery. This possibility is not inconsiderable, however, we were unable to address the question at the meta-analysis level due to insufficient sample size. The question of whether the multiple testing is inappropriate partly depends on whether the tests are independent. The fish endpoints often respond in interpretable patterns. For example, a pattern of increased liver size, decreased condition and decrease in gonad weight may indicate chemical toxicity (Environment Canada, 2012a). Because fish must allocate resources among reproduction, growth, and stress response, the endpoints are necessarily correlated.

This correlation among endpoints has implications for the potential to apply the Bonferroni correction. Where tests are strongly correlated, applying the Bonferroni correction can distort conclusions and be overly conservative (MacDonald 2014). We computed Pearson Correlation Coefficients for the five endpoints, but found no pattern of widespread correlation among endpoints (data not shown). This is likely due to insufficient data to separate out mines by types that have effects (base and precious metals) and other mines. This is not surprising, given the variety of species, mine types, receiving bodies, and that the analysis included both mines showing an effect, and those not showing an effect. Therefore we have insufficient data to see whether mines that do show a downstream effect also have an interpretable pattern of endpoint effects.

This pattern of endpoint failure is one way of looking at the EEM results that would help distinguish mines that truly have an effect from those random test failures. However, it is important to note that the program requires that not only do mines have a test failure, but that they show the same test failure in a consecutive round of monitoring, before being required to undertake Investigation Of Cause (IOC). Again, we had insufficient data to test whether there are recurrent EEM end point effects detected in consecutive phases of monitoring.

Question 3: Will Changing Decision Making Thresholds in the EEM Process for Individual Mines Result in a Reduction in Protection to the Environment?

At the individual mine level, we analyzed whether mines were failing multiple endpoints. With an α=0.10 and 20 tests per mine (5 endpoints, 2 species, 2 sexes), for each individual study, one would expect two tests to be positive by chance. This is an unacceptable level of false positives. To address this issue, Environment Canada has proposed that the program adopt Critical Effect Sizes (Table 4), such that a simple significant difference between reference and exposure is no longer sufficient to trigger Investigation of Cause. We analyzed the effect of this change on the existing data, using results for the ANCOVAs (Table 5).

Table 4. Environment Canada's Proposed Critical Effects Sizes (CES) for fish endpoints (Environment Canada 2012c).

ENDPOINT	CRITICAL EFFECT SIZE (CES)
GONAD SIZE	± 25%
LIVER SIZE	± 25%
CONDITION	± 10%
AGE	± 25%
WEIGHT AT AGE	± 25%

Table 5. Number of contrasts (using EEM endpoints) that fail by a significance level of 0.10, and those that fail by this level and by a magnitude greater than the proposed Cumulative Effects Size.

ENDPOINT	FAIL BY SIGNIFICANCE	FAIL BY CES AND SIGNIFICANCE	PERCENT OF MINES FAILING BY SIGNIFICANCE AND CES
GONAD WEIGHT	47	25	25.00%
LIVER WEIGHT	58	31	27.58%
CONDITION	67	0	0%
WEIGHT-AT-AGE	45	16	16.22%
AGE	74	17	23.73%

This analysis indicates that utilizing Critical Effects Sizes (CES) would reduce the number of test failures by more than half (Table 5). It is likely that this will do a great deal to eliminate the number of test failures that are driven by stochasticity. We are in agreement with Huebert and McKernan that implementing CES would address a great deal of the problem of false positives.

Huebert and McKernan (2014) also suggested that EEM studies should also use a Bonferroni correction to address the issue of a high false-positive rate due to multiple tests. This correction may be warranted if the distribution of failed tests approximates what one would expect if failure were due to random chance. The number of mines with *n* number of failed tests is given by:

Expected Number of Mines = α^n x Total Number of Mines.

For 78 mines, this is:

$E = \alpha^n \times 78$

Therefore, we plotted the distribution of failed tests, compared to the expected distribution of failed tests, to determine whether stochasticity was driving the number of failed endpoints in the EEM studies (Figure 2). If test failures were a product of random chance, one would expect to see many mines with few failures, and then a decreasing distribution. In other words, one would expect that if random chance were driving the number of failed tests in the EEM program, most mines would fail by only one endpoint. Due to limitations in the data, only 1 species per mine were considered, giving a total possible of ten tests (one species, two sexes and five endpoints.) Figure 2 shows that the actual distribution of failed tests does not follow the pattern expected if endpoint failure were driven by chance. Though there are a number of mines that failed by one endpoint, there are many more that failed by multiple endpoints.

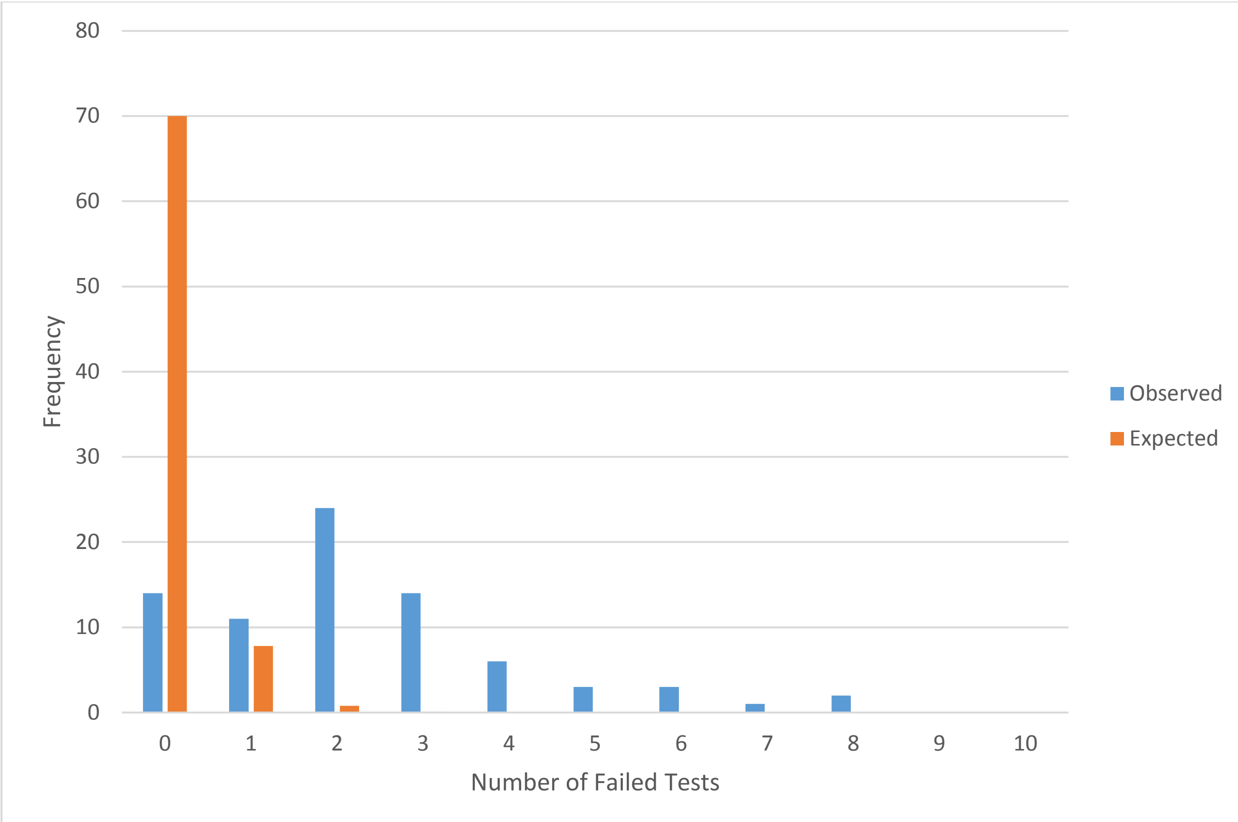

Figure 2. Number of tests failed both by statistical significance, both of contrasts and interaction, and by difference greater than Critical Effects Sizes. The number of tests possible is ten: one species, two sexes, and five endpoints (gonad weight, liver weight, length (condition), weight at age, and age.). There are 78 EEM studies in this data set. The Expected number of mines with n failed tests is given by $E = 0.1^n \times 78$

These results suggest that where a mine fails one endpoint, it is likely to fail another, particularly once interactions are considered.

It is important to note here that calculating the expected number of failed tests assumes that a failure is simple significance, whereas the failures in the observed data set are defined as significance plus magnitude greater than Critical Effect Sizes. Furthermore, as the data in Figure 2 were not corrected for the number of tests per mine, it is likely that Figure 2 represents an underestimate of the number of multiple failures, as a number of mines did not have complete data sets. As incorporating CES reduced the number of failed tests by more than half, we anticipate that this will remove the bulk of the false positives. It is known that applying the Bonferroni correction can result in an unacceptable rate of false negatives, and in cases where one is testing a family of hypotheses, it is unclear whether this simple adjustment makes sense (McDonald 2014). Though we did not test the distribution of p-values, for these data, application of the Bonferroni correction will likely result in the elimination of a number of tests that show truly large changes in magnitude between reference and exposure. An examination of the analysis of Huebert and McKernan (2014, Appendix A) supports this conclusion.

Question 4: Do significantly affected end-points correlate with studies of acute or sub-lethal toxicity, or with compounds released by mines?

Addressing the question of whether any of the acute lethality or sublethal toxicity tests are predictive of effects in the receiving environment are beyond the scope of the data that were supplied by Environment Canada. We only received data for the field component of the EEM studies. It was also not possible to relate the data in the file “MMER Review – Summary of MMER Monitoring Results – 2014-09-26.xlsx” (Summary Table) to data provided for this analysis, as there were no common mine identifying information between the two tables. Furthermore, there is no way to determine if tests failed in the same phase, as there is no phase information. Therefore, it is not possible to address Question 4 using the data at hand.

Conclusion

We analyzed the fish endpoints for the EEM data provided to the Mining Association of Canada in two previous requests. After significant conferring with Environment Canada, we were able to organize the data to perform a meta-analysis to determine whether there were pervasive, national-level effects on the downstream environment. Our meta-analysis reports similar findings as the Second National Assessment: there is a trend towards smaller fish downstream of mines, although with enlarged (rather than smaller) livers. There’s also a tendency towards reduced weight-at-age. These results are found primarily in base metal and precious metal mines. We used methods consistent with those of the Second National Assessment, thus the reasons for the differences between our analysis and that of Environment Canada are likely due to discrepancies between underlying data rather than data handling procedures. However, our results do agree with those of Environment Canada that there is a detectable effect of base and precious metal mine effluent on the receiving environment. This requires further investigation.

We also evaluated whether proposed changes to decision making thresholds would have an effect on the outcome of the studies. We found that applying Critical Effects Sizes to the fish endpoints reduced the number of significant effects by more than half. This will likely eliminate the majority of false positive tests that have encumbered MMER studies. In addition, we note that proceeding to Investigation of Cause does not simply require finding one significant difference in one round of monitoring. Instead, it requires finding the same effect in consecutive rounds of monitoring (Environment Canada, 2012a). Thus it is not simply the demonstration of an effect that leads to IOC, but an interpretable pattern of changes confirmed over two monitoring cycles. This requirement, combined with implementing Critical Effects Sizes, greatly reduces the probability of a mine proceeding to IOC based on stochasticity. We were unable to evaluate the available data to determine how many mines had an interpretable pattern of effects over CES in two consecutive rounds of monitoring, as there were too few mines with data in consecutive phases.

We also evaluated whether there was reason to incorporate a Bonferroni correction to account for the inflation of Type I error caused by multiple testing, a phenomenon termed the False Discovery Rate. If the prevalence of test failures in the EEM program were strongly influenced by the False Discovery Rate, there would be a large number of EEM studies for which there are only one failed test. However, we find that mines who fail one EEM test typically fail others (Figure 2). Therefore it is unlikely that the degree of effects reflected by these studies is strongly influenced by False Discovery. Furthermore, it's unclear how or whether to apply a correction such as the Bonferroni to a family of hypotheses (McDonald 2014) such as the EEM endpoints. The EEM endpoints are chosen to reflect relative contributions of energy to reproduction, stress response, or growth, therefore different patterns of test results may indicate potential sources of a problem (Environment Canada 2012a). The EEM tests therefore constitute a family of hypotheses, whose results are correlated. Finally, though the False Discovery Rate is a general problem of major concern in environmental and medical studies, most correction factors assume that the cost of a Type I error is much higher than the cost of a Type II error. In the context of science for making regulatory and compliance decisions, this is not appropriate, and in fact the Technical Guidance Document (Environment Canada, 2012a) recommends that α and β be equivalent. Experience with the Pulp and Paper Effluent Regulations has shown that in practice, with an α of 0.05, 11.6%, 39.0% and 31.9% of EEM tests failed to reach 80% power for condition, gonad size and liver size respectively (Mudge et al. 2012). The Bonferroni is the most conservative of techniques designed to reduce the False Discovery Rate, and is known to increase the Type II error when the number of tests is large (Colquhoun 2014, Bland and Altman 1995, Perneger 1998.) Thus applying a Bonferroni correction would be needlessly conservative, and will likely have an unacceptable increase in the rate of false negatives.

The data provided to us were insufficient to test whether there were correlations between sublethal toxicity and acute lethality tests, the concentration of effluent or effluent components, and the results of the field studies.

A final comment must be made regarding the quality of these data. There are a number of studies in this data set with insufficient number of fish, or obvious data entry errors. The patchy quality of the data made our analyses quite difficult, and we required a great deal of conferral with Environment Canada to obtain the data we have. Thus we encourage Environment Canada to improve the guidance of EEM studies, or enforcement of the guidelines that do exist. It may be that consultants return from the field before dissecting fish, thus find upon their return that they do not have enough adult fish for a full data set. This represents a considerable waste of financial and natural resources. It may also be that there are too few fish in the receiving environment to support a full EEM fish study. In this case an alternative study must be recommended to Mine operators.

For this particular set of analyses we chose to emulate EC's methodology as far as possible, to provide a starting point for comparison. This included using Hedge's d, removing comparisons that showed significant interactions, and substituting fork length for fish length when one measure was missing. This allowed us to demonstrate general agreement with EC's findings in the Second National Assessment, and identify differences that are likely the result of differences in base data. In particular, we are missing early studies, and the Second National Assessment did not consider Phase 3 or 4 studies.

Further work is required to address: effects of including interaction terms in study, effects of more nuanced data interpolation including inference of length from fork length where total length is not available, and utilizing response ratios in addition to Hedge's d. In addition, by performing the same analysis on the benthic data we could assess whether effects are detected in these communities. This would be very important to informing management decisions, particularly given that the quality of the fish data makes drawing firm conclusions difficult. Finally, no one has yet been able to assess the combination of critical effects sizes, statistical significance, effects in consecutive rounds of monitoring, and findings of Investigation of Cause. This type of analysis would go a long way to determining whether mines are having an effect on the receiving environment.

Furthermore, if, as suggested (Environment Canada 2012c), the frequency of EEM studies is reduced in future, it will result in collection of data with even larger gaps in time, making a national analysis of EEM results even more difficult. If Environment Canada mandates that mines undertake studies, then it has a responsibility to ensure that they are done in a manner sufficient to generate data that can answer the question for which they were designed. Thus we recommend a number of changes to the guidance document, including mandating, rather than recommending, multiple exposure and reference sites, and enforcing the requirement for catching sufficient fish to undertake a proper study. If sites are unproductive such that there are too few fish present or too much effort required to capture sufficient samples to support an EEM study, then alternatives should be designed.

A national analysis is essential to ensure that our national regulations are doing the job for which they were designed. This is an important goal, and the data submitted must be sufficient for this task. Therefore, we recommend that the frequency of EEM studies not be reduced.

Acknowledgements

The authors would like to acknowledge the following people for their assistance and helpful comments on this work:

Provided through the MEND Secretariat: Carrie Rickwood, Natural Resources Canada; Sylvie Richard, Environment Canada; Alan Willsie, Environment Canada; Kenneth Olsen, Environment Canada. On behalf of the Mining Association of Canada: Dave Huebert, Stantec; Bruce Kilgour, Kilgour & Associates; and a member of a Mining Association of Canada company. On behalf of Cameco Corp.: Cassandra Rees and Patrick Carrier of Canada North Environmental Services.

On behalf of the Environmental Caucus of the Metal Mining Effluent Regulations 10-Year Review: John Werring, David Suzuki Foundation; Sue Moodie, Council for Public Health in Mining Communities; and Ugo Lapointe, MiningWatch Canada. The Environmental Caucus of the Metal Mining Effluent Regulations 10-Year Review (Ugo Lapointe, Mining Watch Canada; Ramsey Hart, MiningWatch Canada; Sheila Cole, Nova Scotia Environmental Network; Sue Moodie, Council for Public Health in Mining Communities; John Werring, David Suzuki Foundation and others) helped to guide the initial questions.

We would also like to thank the kind advice and input from Kelly Munkittrick, Canadian Oil Sands Innovation Alliance, and a great deal of assistance from Alan Willsie and Rick Lowell, Environment Canada, in understanding the data. We would also like to thank John Werring, David Suzuki Foundation, for assistance in finding support for this initiative, and the financial support of the Mine Environment Neutral Drainage (MEND) program for funding this report, and for their kind assistance and coordination of the review process.

References

Bland JM and Altman DG. 1995. Multiple significance tests: the Bonferroni method. British Medical Journal. 310 (6973): 170

Colquhoun D. 2014. An investigation of the false discovery rate and the misinterpretation of p-values. Royal Society Open Science. 1(3): 140216

Environment Canada. 2012a. Metal Mining Technical Guidance for Environmental Effects Monitoring. Gatineau (QC): National EEM Office, Environment Canada.

Environment Canada. 2012b. Second National Assessment of Environmental Effects Monitoring Data from Metal Mines Subjected to the Metal Mining Effluent Regulations. Gatineau (QC): National EEM Office, Environment Canada.

Environment Canada. 2012c. 10-Year Review of Metal Mining Effluent Regulations: Discussion Paper.

Kilgour & Associates 2014. Review and Analysis of the 3rd Metal Mining Environmental Effects Monitoring National Assessment. Submitted to: The Mining Association of Canada.

Huebert D and McKernan M. 2014. Analysis of Environmental Effects Monitoring Data. Stantec Consulting Ltd. Prepared for: The Mining Association of Canada.

Lowell RB, Tessier C, Walker SL, Willie A, Bowerman M and Gautron D. 2007. National Assessment of Phase 1 Data from the Metal Mining Environmental Effects Monitoring Program. National EEM Office, Environment Canada.

McDonald JH. 2014. Handbook of Biological Statistics (3rd ed). Sparky House Publishing, Baltimore, Maryland.

Mudge JF, Barrett TJ, Munkittrick KR and Houlahan JE. 2012. Negative consequences of using $\alpha = 0.05$ for environmental monitoring decisions: a case study from a decade of Canada's Environmental Effects Monitoring Program. Environmental Science and Technology. 46: 9249.

Nakagawa S. 2012. Third party report for Environment Canada's "Second National Assessment of Environmental Effects Monitoring Data from Metal Mines Subjected to the Metal Mining Effluent Regulations" and for Stantec's evaluation of this assessment. Report prepared for the Mining Association of Canada.

Perneger TV. 1998. What's wrong with Bonferroni adjustments. British Medical Journal. 316(7139): 1236.

Rosenberg, MS, Adams DC and Gurevitch J. 2000. MetaWin statistical software for Meta-Analysis version 2. Sinauer Associates Inc. Sunderland, Massachusetts.