UNIVERSITY of OTAGO

Te Whare Wānanga o Otāgo

Third party report for Environment Canada's "Second National Assessment of Environmetnal Effects Monitoring Data from Metal Mines Subjected to the *Metal Mining Effluent Regulations*" and for Stantec's evaluation of this assessment

Prepared by Shinichi Nakagawa, PhD
Department of Zoology, University of Otago, Dunedin, New Zealand

Prepared for the Mining Association of Canada

July 2012

Abbreviations used: CES (critical effect size), CI (confidence interval), EC (Environment Canada), MAC (Mining Association of Canada), EEM (Environmental Effects Monitoring), MMER (Metal Mining Effluent Regulations), MMTG (Metal Mining Technical Guidance), SAT (statistical assessment tool), SD (standard deviation)

PO Box 56, Dunedin, New Zealand.
Tel 64 3 479 5046 • Fax 64 3 479 7584
Email shinichi.nakagawa@otago.ac.nz • Web www.otago.ac.nz/Zoology/staff/academic/

Table of Contents

1.0 Background

Environment Canada (EC) recently released a report titled "Second National Assessment of Environmental Effects Monitoring Data from Metal Mines Subjected to the *Metal Mining Effluent Regulations*" (hereafter called, the 2nd assessment). In this report, EC presented several statistically significant and negative impacts of mining effluent on fish and invertebrate communities. In the light of the results reported in the 2nd assessment, Mining Association Canada (MAC) sought independent reviews. MAC asked Dr David Huebert (Stantec Consulting Ltd.) to write an independent review on the 2nd assessment. I (Dr Shinichi Nakagawa, University of Otago) have also been asked to provide an independent review on the 2nd assessment. At the same time, I have been asked to review Dr Huebert's evaluation (hereafter called, Stantec's assessment).

As a third party and independent review, the aim of the current review is to assess both the 2nd assessment and Stantec' assessment. My evaluation will focus on whether the analyses employed in these two reports were technically and statistically sound and also whether the stated conclusions of the two reports were appropriate given the results. As the two reports show, disagreements between these two parties (i.e. EC and Stantec) reside almost exclusively in their interpretations of the meta-analytic results, which EC presented in the 2nd assessment. Therefore, I concentrate my review on the meta-analytic methods EC used and the interpretations of meta-analytic results by these two parties. This report have three main parts: (i) an evaluation of the 2nd assessment, (ii) an evaluation of Stantec's assessment and (iii) a brief concluding section.

Before I start, I note that my knowledge regarding mining, ecological assessments and related matters is very limited, but I have considerable expertise in data analysis and statistics. Although I am a behavioural/evolutionary ecologist by training, I have published a number of papers on statistical practices (Nakagawa and Cuthill 2007; Nakagawa and Schielzeth 2010; Nakagawa and Hauber 2011). Notably, I have several publications on meta-analytic methods (Hadfield and Nakagawa 2010; Nakagawa and Santos 2012) along with numerous papers using meta-analyses (e.g. Cleasby and Nakagawa 2012; Nakagawa et al. 2012; for the full list of my publications see http://scholar.google.com/citations?user=9-pCZzcAAAAJ&hl=en). Therefore, I should be able to provide fair insights into and assessments of whether the meta-analyses by EC were conducted correctly and also whether the two parties interpreted the results appropriately.

2.0 Evaluation of the 2nd assessment

2.1 General comments

My general impression of environmental effects monitoring (EEM) under the Metal Mining Effluent Regulations (MMER) is very positive. EEM is a very carefully crafted, controlled and well thought-through programme, as I found out by reading relevant sections of the document titled "Metal Mining Technical Guidance for Environmental Effect Monitoring" (published by EC in 2012; referred to as MMTG hereafter), although EEM is not without its faults (Huebert et al. 2011). If one can assume that each mine participating in EEM follows the guidelines described in that document, the quality of data used in the 2nd assessment is indeed very high.

However, I have several major concerns on the 2nd assessment by EC. My prime concern is that virtually no details on meta-analytic methods were described so that I could not really evaluate the soundness of the meta-analyses conducted. Here I assume that EC used *MetaWin* because Rosenberg et al. (2000) was mentioned when the authors defined what meta-analysis is in the page 8. I note that I have read Booty et al. (2009), which described the statistical assessment tool (SAT), but SAT does not seem to have meta-analytic capacity.

In *MetaWin*, one can use either fixed-effect meta-analysis or random-effects meta-analysis. Fixed-effect models assume that there is one common meta-analytic mean among all the data points (i.e. the standardized mean difference, or Hedges' *d*, between a reference location and a treatment location across all the mining sites are the same apart from variation arising from sampling errors). Given such an unreasonable assumption, fixed-effect meta-analyses are not recommended and no longer used except special cases (Gurevitch and Hedges 1999; Higgins et al. 2009). Fixed-effect models usually provide narrower confidence intervals (CIs) than random-effects models. Hence, one is likely to find more statistical significance than when using random-effects models (cf. Jennions and Møller 2002) and as a result is likely to draw inappropriate statistical conclusions. My concern is that EC may have used fixed-effect models. Unfortunately, there is no way of telling this form the current version of the 2nd assessment.

Even if we assume that EC did use random-effects meta-analysis, which is an appropriate model for the data, the meta-analytic means and their CIs cannot be readily interpreted if significant heterogeneity exists in the data (i.e. effect inconsistence among studies/sites). Ordinarily, meta-analysis includes analysis of heterogeneity, which is usually quantified as either Cochran's Q (Hedges and Olkin 1985) or I^2 (Higgins and Thompson 2002; Higgins et al. 2003). However, the 2nd assessment does not provide any evaluation of heterogeneity. On one hand, statistical interpretations of meta-analytic means are fairly straightforward when there is low (and statistically non-significant) heterogeneity (i.e. there is consistency of effects observed among studies or data points included in a meta-analysis). On the other hand, when one finds high heterogeneity (inconsistencies among studies or data points), one needs to be cautious how meta-analytic means are interpreted because there may be some substructure among data points.

If high (and statistically significant) heterogeneity is found, it calls for meta-regression approaches (also called mixed-effects models; Hedges and Vevea 1998; Cheung 2008; see also Thompson and Higgins 2002); that is, one incorporates predictors (explanatory variables; or, in the terminology of meta-analysis, moderators) to explain such heterogeneity. Indeed, EC conducted some additional analyses incorporating habitats and ore types into meta-analyses of fish related endpoints (see Section 4.4 in the 2nd assessment). Given that they found large variation depending on habitat and ore types, it is likely that significant heterogeneity exists in the EEM data. Although it is good to see these additional analyses, meta-regression analyses were not carried out systemically across all the endpoints. The current approach lacks rigour and it may have missed some important moderators contributing to the variation in the data.

Furthermore, although *MetaWin* has been widely used in the field of ecology and evolution (Nakagawa and Santos 2012), its capability is limited. For example, *MetaWin* cannot account for data dependence (also referred to as pseudo-replication; Hurlbert 1984) and can take only one moderator (one predictor). I bring up these two points specifically (i.e. data dependence and only one moderator) here because they are highly relevant to the EEM data in the 2nd assessment. With regards to the first point, EC assumed all data points (studies) to be independent, but this assumption may not hold for near-by mines or mines operated by the same companies (e.g. the effect sizes from such mines may be closer to one another than other effect sizes). With regards to the second point, when heterogeneity exists, it is likely caused by multiple sources (e.g. habitats and ore types). These factors should be modelled in single meta-analytic model rather than fitting different models. Meta-analytic models (i.e. meta-regressions) incorporating multiple moderators can also examine how these moderators interact with each other (note that this is not possible using *MetaWin*). I will describe how these points can be addressed in the section '2.3 Recommendations' below.

Statistical models should always be accompanied by sensitivity analysis. In meta-analyses based on data from published literature, sensitivity analysis is usually in the form of analysis of publication bias, whereby statistically significant results are more likely to be published than otherwise (Rosenthal 1979; Rothstein et al. 2005). For the EEM data, publication bias should not be of concern because EEM is extensive. However, it is important to examine the distribution of data to see if any outlier data points are driving some unwarranted statistical significance. To achieve this, visual techniques employed for publication biases, such as funnel plots (Egger et al. 1997; Sterne et al. 2005; Peters et al. 2008) and the normal quantile plots (Wang and Bushman 1998) would be useful for the EEM data. Notably, funnel asymmetry found in funnel plots can indicate existence of heterogeneity (see Egger et al. 1997).

I also note that the use of Hedges' *d* in the EEM data may be inappropriate. The use of *d*-type effect sizes, including Hedges' *d,* in ecological experiments has been criticized in the ecological literature (Osenberg et al. 1997, Hedges et al. 1999). When the control and experimental groups have different variances, the use of *d* should be avoided because *d* is sensitive to such variance differences (Osenberg et al. 1997). Therefore, the most widely used effect size statistics for ecological experiments is now the response ratio (Nakagawa and Santos 2012). Some meta-analytic studies use both *d* and the response ratio, so as to ensure that their conclusions are robust (e.g. Isaksson 2010). The formulation and extension of the response ratio can be found in Hedges et al. (1999) and

Lajeunesse (2011). Obviously, my concern is that the response ratio may have been more appropriate for the EEM data than Hedge's *d*.

Above, I highlighted a number of shortcomings in the current methods used in the 2nd assessment: (1) the lack of descriptions regarding meta-analytic models and software used, (2) no analysis on heterogeneity undertaken, (3) inconsistent and limited applications of meta-regression, (4) the lack of controlling for possible non-independence, (5) no sensitivity analysis conducted and (6) no acknowledgement of the potential problems associated with the use of Hedges' *d* in ecological experiments. Given these concerns, unfortunately, I cannot really evaluate any of the specific conclusions drawn from the meta-analyses presented in the 2nd assessment. Therefore, I think that all the meta-analytic results should be seen as exploratory or, at the best, preliminary at this stage. Therefore, any conclusions drawn in the 2nd assessment should also be treated as preliminary or still unwarranted until analyses are repeated and reported in a statistically sound manner, as described in this section and in the following sections.

2.2 Specific comments

Page 7: The 2nd assessments states: "Some metal mines collected data from multiple areas. Data from more than one near-field area were pooled only if warranted based on inspection of pooling procedures used in the interpretative reports."

This sentence indicates to me that there may have been multiple data points from one study. Such data dependence needs to be statistically dealt with (see the section '2.3 Recommendations'). Also the information regarding each data point's variance (or standard deviation or precisions/weights) should be revealed. This can be easily achieved by incorporating funnel plots of raw data (e.g. a plot of precisions against effect sizes or variants of this). My reading of the 2nd assessment and MMTG suggests that there may have been some variation in sample sizes at each site (i.e. difference in sampling error variances). Also, how EC pooled related data points is important when working out variance values for effect sizes and the detail of this needs to be provided (see Chapter 25 in Borenstein et al. 2009; see also Lajeunesse 2011).

Page 8: The 2nd assessments states: "Meta-analysis is a set of statistical procedures used to quantitatively synthesize the results of a large number of independent studies (e.g., a meta-analysis of multiple studies of the effects of smoking to determine larger trends in the health impacts of smoking)."

This definition is probably a little obsolete, because recent meta-analyses using hierarchical (or multi-level) models can incorporate data dependences. Meta-analytic methods have improved drastically over recent years (see Sutton and Higgins 2008; Cooper et al. 2009; Nakagawa and Santos 2012).

Section 4.4 'Response Patterns – Additional Meta-analysis' – to determine statistical differences among different categories, contrast analyses (i.e. pair-wise comparisons) should be provided. The same can be said for the main meta-analyses as well; it will be helpful to know whether the results from Cycle 1 and Cycle 2 are significantly different.

Section 6.5 'Relationship to Effluent Flow' – the analyses in this section should be done in a meta-analytic framework (i.e. controlling for any differences in variances among data points).

2.3 Recommandations

I recommend that EC provide a detailed description of the meta-analytic methods in their report, accompanied by information regarding each data point or study (e.g. sample sizes, habitat, ore types etc.). Although I suspect that some of these details may be confidential, EC should at least be able to provide information on effect sizes, sample sizes and whether certain data points were obtained from the same mines and/or companies (i.e. information regarding data dependences).

I recommend the use of meta-analytic models that can incorporate correlated patterns of data; for example, meta-analysis using Bayesian hierarchical (multi-level) models can address this issue (e.g. Liermann and Hilborn 1997; see also Congdon 2003; Nakagawa and Santos 2012). Also, such models can incorporate spatial autocorrelations (Congdon 2003, 2007). These models are useful because mines geographically closer to one another may produce similar effect sizes (I note that in Figure A1 in the 2nd assessment, I can see some clustering of mines). Implementations of these models are possible using software packages such as *ASReml* (http://www.vsni.co.uk/software/asreml), *WinBUGS* (http://www.mrc-bsu.cam.ac.uk/bugs/) and *R* (http://www.r-project.org/); for *R*, the package *MCMCglmm* can be used (Hadfield 2010; Hadfield and Nakagawa 2010). Meta-analytic models implemented in these software packages are also able to take multiple moderators so that one can examine not only how each moderator accounts for variation in the meta-analytic data, but also how interactions between moderators (e.g. habitats and ore types) may explain variance in the data. Importantly, by using such a framework, EEM data from Cycle 1 and Cycle 2 can be analysed within one model and one can assess temporal trends in the data.

Furthermore, EC may want to consider multivariate (multi-response) meta-analysis (van Houwelingen et al. 2002; Jackson et al. 2011) because EEM data includes multiple endpoints for both fish and invertebrate communities. In a multivariate meta-analysis, related effect sizes (e.g. effect sizes for fish gonad weight, liver weight, condition weight-at-age and age) are modelled simultaneously. Such a model would provide degrees of correlations among these different effect sizes on various endpoints, as well as the parameter estimates one would expect from normal (univariate) meta-analysis. Of course, all other shortcomings I have listed above (e.g. data heterogeneity, the lack of sensitivity analysis and the use of response ratio) should also be addressed.

What I have suggested in this section would require some expertise and a good understanding of the statistics relevant to meta-analysis. This means that EC would probably like to commission the meta-analytic parts of their analyses to a professional statistician or meta-analyst.

3.0 Evaluation of Stantec's assessment

3.1 General comments

I have outlined the shortcomings of the current meta-analytic results in EC's assessment (the 2nd assessment). However, for the sake of argument, here I assume that all meta-analyses were conducted appropriately in order to interpret Stantec's assessment.

My overall impression of Stantec's assessment is that their case is well made, but it feels somewhat unbalanced toward negating the conclusions drawn by EC. In other words, I felt that the wording used in Stantec's assessment seems to be slightly inappropriate in some places in order to show discord with the EC's conclusions. This is particularly reflected in the rather inaccurate use of terms such as random errors and noise (see below; also explained in the 'Specific comments' section).

In Nakagawa and Cuthill (2007), my colleague and I used the term 'biological importance' to accompany the term 'statistical significance'. Biologists working in different sub-disciplines need to decide what magnitude of effect size should be considered to be biologically important given the best biological information available. My understanding, obtained from reading the 2nd assessment and MMTG, is that EC's critical effect sizes (CES) would be more appropriately treated as 'practical thresholds' or 'practical importance/significance' rather than what constitutes biological importance; here, I use practical importance as an acceptable threshold considering other types of factors than biological factors (cf. Kirk 1996; Thompson 2002a,b). For example, changes of 2 standard deviations (SD) in the endpoint measurements associated with invertebrate communities seems fairly large, and a change smaller than this threshold may be biologically meaningful or important for invertebrate communities in some (if not all) cases.

Stantec's assessment, to some extent, treated effect sizes below a threshold of 2 SD as biologically unimportant (when meta-analytic means and CIs were below the d values corresponding to CES; e.g. d = 0.9-1.0 or d = 1.8-2.0) or sometimes as random error/ noise (when Hedges' d is less than 0.2). I do not necessarily agree with such interpretation, because if meta-analytic means are statistically significant (i.e. with a CI that does not cross zero), it is likely that the associated biological effects are non-zero and are unlikely to be due to noise. I think, what Stantec's assessment should have said, is that if a particular overall effect size is below the threshold (and also statistically significant), such an effect should be interpreted as biologically negligible (e.g. if $d < 0.2$) or as unimportant in terms of the practical thresholds set by EC ($d <$ 0.9-1.0 or 1.8-2.0, depending on the endpoints).

Also, I am not certain about the validity of comparing CES values to the meta-analytic means. I feel that CES should be applied to each site (study) separately. Imagine a hypothetical case where one mine is having a large effect on endpoints (above CES values), whereas the impacts of the other mines are much below CES. Relying on meta-analytic means will not reveal the effect of this one particular mine (note that this is why I mentioned earlier the importance of quantifying heterogeneity and checking for outliers using funnel plots).

Nonetheless, I think that the alternative interpretations of EC's meta-analytic results put forward by Stantec are well made and do indeed seem to be more appropriate in relation to EC's CES thresholds. Also, Statntec's assessment rightly states that EC failed to calibrate Hedges' *d* values with regards to their CES cut-offs. Therefore, EC should take the points and concerns raised in Stantec's assessment seriously.

3.2 Specific comments

Section 2.1.1 'Calibration Options' – Stantec's assessment says: "Absent any understanding of biological significance and as a first approximation, it has long been considered (Cohen 1988 cited in Nakagawa and Cuthill 2007) that a Hedges 'd' value below 0.2 is indicative of 'no effect'. Values below this cutoff are indistinguishable from background 'noise,' and within the EEM context, indicate there is no discernible effect of effluent discharge on downstream receiving environments."

I do not think that Hedges' *d* below 0.2 is considered 'no effect' or 'noise'. Depending on the contexts, $d = 0.1$ or smaller effects (if statistically significant) can be considered as important; for example, a 3% increase in cancer rates would be a clinically significant effect.

Section 2.1.1 'Calibration Options' – Stantec's assessment says: "These average values provide a cutoff benchmark for interpretation of the results. If the Hedges 'd' values for 'exposure' versus 'reference' sites were the same as, or less than, the 'reference' versus 'reference' sites, then it could be considered that there was no effect of mine effluent on downstream fish health and benthic invertebrate communities… The mean reference Hedges 'd' was approximately 0.8 for density and 1.2 for richness (Figure 3-1)."

This argument on the reference sites is interesting. However, I point out that the values of 0.8 and 1.2 are probably overestimates. If pairs of random references sites are taken and then Hedges' *d* values were calculated (randomly assigning which reference group will first be entered into the equation for Hedges' *d*), one expects that the mean Hedges' *d* should be close to zero. I assume that these values in Stantec's assessment (i.e. 0.8 and 1.2) were calculated using the absolute values of Hedges' *d*. If so, I am not sure how comparable these values are to the meta-analytic results presented by EC. Provided that there are no particular reasons to suspect that all exposure sites are intrinsically different from reference sites in the EMM data (my reading of MMTG suggests not), Hedges' *d* estimate should not be biased.

Section 2.1.1 'Calibration Options' – Stantec's assessment says: "For the benthic invertebrate community-structure data, the Critical Effect Size (CES) is ±2 Standard Deviations, which corresponds to a Hedges 'd' of approximately 1.8-2.0. For the fish-health data, the CES is 25% for all parameters but fish condition, which has a stated CES of 10%. Assuming a coefficient of variation of 10% for fish condition, and 25% for the other endpoints (as stated in Environment Canada 2011), this is a Hedges 'd' of approximately 0.9-1.0. It can be considered that Hedges 'd' values below these benchmarks are not biologically meaningful in the context of the EEM program"

There should be no disagreement of how 2 SD translates into Hedges' *d*. However, how % differences translate into Hedges' *d* may need some attention because cut-off values will vary depending on what one assumes as the pool standard deviation (SD) in the Hedges' *d* formula. I recommend that EC should explicitly state how their CES translate into Hedge's *d* in their report for each endpoint (note that Hedges' *d* may not be an appropriate effect size statistic for the EEM data; see above).

Section 2.3 'INFERRED MEANING FROM "NON-SIGNFICANT" DATA' – Stantec's assessment says: "These conclusions arose because the authors inexplicably inferred meaning from statistically non-significant data (Table 2-2). Even though sole reliance on 'statistical significance' is inappropriate in the context of a meta-analysis, at a minimum endpoints should at least be significantly different than Hedges 'd' = 0 if they are to be discussed in terms of effluent effects."

This criticism is correct, but it is little harsh to say: "the authors inexplicably inferred meaning from statistically non-significant data". I point out that EC states, "*with more of a tendency to reduced gonad size*". In the literature, researchers do talk about 'tendencies' without statistical significance, especially if results are in the expected directions (whether this is appropriate in this context or not). See also Section 3.1 for the same criticism.

Section 3.8 'BRAY CUTIS INDEX (BCI)' – Stantec's assessment says: "The Hedges 'd' values for the BCI are inconsistent with the Hedges 'd' values for density and richness. This inconsistency is of importance because the BCI is calculated from density and richness data and it would be expected that meta-analysis would result in similar effect sizes for all three endpoints. However, the inconsistency of the BCI metadata is perhaps not surprising, since within the prescribed EEM methods the BCI is always incorrectly calculated, with a strong bias towards increasing the difference between 'reference' and 'exposure' values (Huebert et al. 2010; Huebert et al. 2011; Huebert et al. 2012b). This suggests that the Hedges 'd' values for the BCI are biased upwards."

I have read Huebert et al. (2011) and Huebert (2012a,b) to ensure the validity of what is described here. As far as I am concerned, BCI does seem upwardly biased in the EEM data, as described in Huebert et al. (2011).

4.0 Conclusion

I have highlighted the methodological problems in the meta-analyses conducted by EC (the 2nd assessment), whereas Stantec's assessment focused on the interpretational issues of EC's meta-analytic results (e.g. the problems associated with the calibration of CES into Hedges' *d*). I certainly feel that EC needs to address both sets of concerns (from Stantec's assessment and my review) before EC's conclusions regarding the relationships between mining effluents and biological endpoints are established in terms of statistical significance and of practical significance. I think that EC has an excellent dataset, which can potentially address what they set out to do (page 4 in the 2nd assessment). Therefore, I hope they will follow the recommendations outlined above. Also, EC should take into account what Stantec's assessment described, especially regarding the calibration of which values of Hedges' *d* are considered to be practically important. Until further (appropriate) analysis is undertaken, the meta-analytic results in the 2nd assessment should be taken as exploratory or preliminary. Therefore, the conclusions in the 2nd assessment are not supported by EC's current results and they should be revised when new results from such appropriate analysis are obtained.

5.0 References

Booty, W.G., Wong, I., Lam, D. & Resler, O. (2009) A decision support system for environmental effects monitoring. *Environmental Modelling and Software*, **24**, 889-900.

Borenstein, M., Hedges, L.V., HigginS, J.P.I. & Rothstein, H.R. (2009) *Introduction to meta-analysis*. Wiley, Chichester.

Cheung, M.W.L. (2008) A model for integrating fixed-, random-, and mixed-effects meta-analyses into structural equation modeling. *Psychological Methods,* **13,** 182-202.

Cleasby, I.R. & Nakagawa, S. (2012) The influence of male age on within-pair and extra-pair paternity in passerines. *Ibis,* **154,** 318-324.

Congdon, P. (2003) *Applied bayesian modelling*. Wiley, Chichester.

Congdon, P. (2007) *Bayesian statistical modelling,* 2nd edn. Wiley, Chichester.

Cooper, H., Hedges, L.V. & Valentine, J.C. (2009) *The handbook of research synthesis and meta-analysis* 2nd edn. Russell Sage Foundation, New York.

Egger, M., Smith, G.D., Schneider, M. & Minder, C. (1997) Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal,* **315,** 629-634.

Gurevitch, J. & Hedges, L.V. (1999) Statistical issues in ecological meta-analyses. *Ecology,* **80,** 1142-1149.

Hadfield, J.D. (2010) MCMC methods for multi-response Generalised Linear Mixed Models: the MCMCglmm R package. *Journal of Statistical Software,* **33,** 1-22.

Hadfield, J.D. & Nakagawa, S. (2010) General quantitative genetic methods for comparative biology: phylogenies, taxonomies and multi-trait models for continuous and categorical characters. *Journal of Evolutionary Biology,* **23,** 494-508.

Hedges, L. & Olkin, I. (1985) *Statistical methods for meta-analysis*. Academic Press, New York.

Hedges, L.V., Gurevitch, J. & Curtis, P.S. (1999) The meta-analysis of response ratios in experimental ecology. *Ecology,* **80,** 1150-1156.

Hedges, L.V. & Vevea, J.L. (1998) Fixed- and random-effects models in meta-analysis. *Psychological Methods,* **3,** 486-504.

Higgins, J.P.T. & Thompson, S.G. (2002) Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine,* **21,** 1539-1558.

Higgins, J.P.T., Thompson, S.G., Deeks, J.J. & Altman, D.G. (2003) Measuring inconsistency in meta-analyses. *British Medical Journal,* **327,** 557-560.

Higgins, J.P.T., Thompson, S.G. & Spiegelhalter, D.J. (2009) A re-evaluation of random-effects meta-analysis. *Journal of the Royal Statistical Society Series A-Statistics in Society,* **172,** 137-159.

Huebert, D.B. (2012a) Rebuttal to Lowell et al. (2012) regarding Canada's environmental effects monitoring (EEM) program. *Integrated Environmental Assessment and Management,* **8,** 568-569.

Huebert, D.B. (2012b) Rebuttal to the response of Bosker et al. (2012) to "Canada's Environmental Effects Monitoring Program: areas for improvement." *Integrated Environmental Assessment and Management,* **8,** 382-383.

Huebert, D.B., McKernan, M. & Paquette, C. (2011) Canada's environmental effects monitoring program; areas for improvement. *Integrated Environmental Assessment and Management,* **7**.

Hurlbert, S.H. (1984) Pseudoreplication and the design of ecological field experiments. *Ecological Monographs,* **54,** 187-211.

Isaksson, C. (2010) Pollution and its impact on wild animals: a meta-analysis on oxidative stress. *Ecohealth,* **7,** 342-350.
Jackson, D., Riley, R. & White, I.R. (2011) Multivariate meta-analysis: potential and promise. *Statistics in Medicine,* **30,** 2481-2498.
Jennions, M.D. & Møller, A.P. (2002) Publication bias in ecology and evolution: an empirical assessment using the 'trim and fill' method. *Biological Reviews,* **77,** 211-222.
Kirk, R.E. (1996) Practical significance: a concept whose time has come. *Educational and Psychological Measurement,* **56,** 746-759.
Lajeunesse, M.J. (2011) On the meta-analysis of response ratios for studies with correlated and multi-group designs. *Ecology,* **92,** 2049-2055.
Liermann, M. & Hilborn, R. (1997) Depensation in fish stocks: a hierarchic Bayesian meta-analysis. *Canadian Journal of Fisheries and Aquatic Sciences,* **54,** 1976-1984.
Nakagawa, S. & Cuthill, I.C. (2007) Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biological Reviews,* **82,** 591-605.
Nakagawa, S. & Hauber, M.E. (2011) Great challenges with few subjects: Statistical strategies for neuroscientists. *Neuroscience and Biobehavioral Reviews,* **35,** 462-473.
Nakagawa, S., Lagisz, M., Hector, K.L. & Spencer, H.G. (2012) Comparative and meta-analytic insights into life-extension via dietary restriction. *Aging Cell,* **11,** 401-409.
Nakagawa, S. & Santos, E.S.A. (2012) Methodological issues and advances in biological meta-analysis. *Evolutionary Ecology,* **doi: 10.1007/s10682-012-9555-5**.
Nakagawa, S. & Schielzeth, H. (2010) Repeatability for Gaussian and non-Gaussian data: a practical guide for biologists. *Biological Review,* **85,** 935-956.
Osenberg, C.W., Sarnelle, O. & Cooper, S.D. (1997) Effect size in ecological experiments: the application of biological models in meta-analysis. *American Naturalist,* **150,** 798-812.
Peters, J.L., Sutton, A.J., Jones, D.R., Abrams, K.R. & Rushton, L. (2008) Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. *Journal of Clinical Epidemiology,* **61,** 991-996.
Rosenberg, M.S., Adams, D.C. & Gurevitch, J. (2000) MetaWin: statistical software for meta-analysis. Sinauer, Sunderland, MA.
Rosenthal, R. (1979) The "file drawer problem" and tolerance for null results. *Psychological Bulletin,* **86,** 638-641.
Rothstein, H., Sutton, A.J. & Borenstein, M. (2005) Publication bias in meta-analysis: prevention, assessment and adjustments. John Wiley, Chichester.
Sterne, J.A.C., Becker, B.J. & Egger, M. (2005) The funnel plot. *Publication bias in meta-analysis : prevention, assessment and adjustments* (eds H. Rothstein, A.J. Sutton & M. Borenstein), pp. 75-98. John Wiley, Chichester.
Sutton, A.J. & Higgins, J.P.I. (2008) Recent developments in meta-analysis. *Statistics in Medicine,* **27,** 625-650.
Thompson, B. (2002a) "Statistical", "practical", and "clinical": how many kinds of significance do couselors need to consider. *Journal of couseling & development,* **80,** 64-71.
Thompson, B. (2002b) What future quantitative social science research could look like: confidence intervals for effect sizes. *Educational Researcher,* **31,** 25-32.

Thompson, S.G. & Higgins, J.P.T. (2002) How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine,* **21,** 1559-1573.
van Houwelingen, H.C., Arends, L.R. & Stijnen, T. (2002) Advanced methods in meta-analysis: multivariate approach and meta-regression. *Statistics in Medicine,* **21,** 589-624.
Wang, M.C. & Bushman, B.J. (1998) Using the normal quantile plot to explore meta-analytic data sets. *Psychological Methods,* **3,** 46-54.

6.0 Closure

This review was prepared for the Mining Association of Canada. This report may not be relied upon by any other people or organizations without the explicit written consent of myself (Shinichi Nakagawa) and the Mining Association of Canada.

Should third parties make use of the contents in this report, or make decisions relying upon the contents of this report, I (Shinichi Nakagawa) will not accept any responsibility for damages, if any, incurred by such actions and/or decisions.

The information and contents in this report were based on published studies and documents. The conclusions and recommendations made in this report represent the best judgement of myself (Shinichi Nakagawa) when this report was prepared.

Written by (20 July 2012):

Shinichi Nakagawa, PhD
Senior Lecturer at the Department of Zoology, University of Otago, New Zealand