

Cautions on using the Before-After-Control-Impact design in environmental effects monitoring programs

K.E. Smokorowski^{a*} and R.G. Randall^b

^aFisheries and Oceans Canada, Great Lakes Laboratory for Fisheries and Aquatic Sciences, 1219 Queen Street East, Sault Ste. Marie, ON, P6A 2E5, Canada; ^bFisheries and Oceans Canada, Great Lakes Laboratory for Fisheries and Aquatic Sciences, 867 Lakeshore Road, Burlington, ON L7S 1A1, Canada

*Karen.Smokorowski@dfo-mpo.gc.ca

Abstract

Often the Before-After-Control-Impact (BACI) design is suggested as being a statistically powerful experimental design in environmental impact studies. If the timing and location of the impact are known and adequate pre-data are collected, the BACI design is considered optimal to help isolate the effect of the development from natural variability. This paper presents 9 years of results from a long-term BACI experiment tested using a range of statistical models and post-impact monitoring designs. To explore suboptimal designs that are often utilized in environmental effects monitoring, the same data were also explored assuming either no control system was available (Before-After only), or that no pre-impact data were available (Control-Impact only). The results of the BACI design were robust to the statistical model used, and the BACI design was able to detect effects from the impact that the two suboptimal designs failed to detect. However, the BACI design demonstrated different conclusions depending on the number and configuration of post-impact years included in the analysis. Our results reinforce the idea that caution should be employed when using, or interpreting results from, a BACI design in an environmental impact study, but demonstrate that a well-designed BACI remains one of the best models for environmental effects monitoring programs.

Key words: anthropogenic impact detection, design and analysis considerations, long term monitoring, Before-After-Control-Impact

OPEN ACCESS

Citation: Smokorowski KE and Randall RG. 2017. Cautions on using the Before-After-Control-Impact design in environmental effects monitoring programs. FACETS 2: 212–232. doi:[10.1139/facets-2016-0058](https://doi.org/10.1139/facets-2016-0058)

Editor: Brett Favaro

Received: September 30, 2016

Accepted: December 5, 2016

Published: March 2, 2017

Copyright: © 2017 Her Majesty the Queen in right of Canada. This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/) (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

Published by: Canadian Science Publishing

Introduction

Environmental effects monitoring design and analyses have been highlighted in the scientific literature for decades ([Green 1979](#); [Stewart-Oaten et al. 1986](#); [Underwood 1991](#); [Chapman 1998](#)). Many challenges surround the ability to separate a human-induced change from natural variability, and addressing this real-world problem has been the focus of many sampling designs and statistical analyses. [Green \(1979\)](#) outlines a hierarchical tree of design options dependent on if the location and timing of the impact are known and/or if a control area is available. If a spatial control is lacking, the effects must be inferred from sampling both Before and After an impact (a design termed Intervention Analysis by [Stewart-Oaten and Bence 2001](#)), but this design necessarily assumes that an unaffected site would not have changed in a similar fashion. If the impact occurred before any monitoring began, the Impact site can be compared to a nearby (or upstream) Control

site, but this design is suboptimal because it assumes the Control and Impact sites were similar prior to the impact. If the timing and location of the impact are unknown, the Reference Condition Approach has been suggested as a method for quantifying impairment of an affected site, but specific and rigorous guidelines must be met for this approach to be considered sound (Bowman and Somers 2005). If the impact has not yet occurred, but the timing and location are known, the Before-After-Control-Impact (BACI) design was suggested as optimal (Green 1979). However, since the BACI design was first introduced by Green (1979) many criticisms and suggestions for improvement have been made.

The first criticism was from Hurlbert (1984) who suggested that the use of a single Control area and a single Impact area (with replicate samples) analyzed by analysis of variance (ANOVA) only demonstrated significant differences between locations, and not of the impact itself because the treatment could not be randomly assigned to the experimental plots. Stewart-Oaten et al. (1986) refuted Hurlbert's (1984) claim by suggesting that using sampling times as replicates with the Control and Impact sites sampled (near) simultaneously, and each sampling time represented by only one number (as the difference between the Impact and Control samples for that time), allows for the detection of the mean difference between time periods, using any standard two factor test. They do stress, however, that the assumptions of additivity and independence must be met in the data within a time period (Stewart-Oaten et al. 1986). If differences are not used the equivalent BACI test is the test for an interaction effect in a two-way ANOVA (specifically the $BA \times CI$ interaction), but the same assumptions apply (Underwood 1991; Smith et al. 1993).

Reviews of designs, new criticisms, and suggestions for design improvement were presented in a series of papers by Underwood (1991, 1992, 1993, 1994). The biggest concerns raised in these papers include the lack of ability to determine causation by the human impact, even with a significant BACI, and the need for multiple Control sites. An asymmetrical design using one Impact site and multiple Control sites was suggested to alleviate the problem of the need for simultaneous sampling, and for interpreting causation so long as the changes observed at the Impact site are greater than the patterns observed in the set of control locations (Underwood 1991, 1994). Stewart-Oaten and Bence (2001) refute Underwood's criticisms by clarifying that BACI Controls are not experimental controls, and are not chosen randomly but are instead deliberately chosen to be highly correlated with the Impact site to be useful covariates. As such, the variation among Control sites is irrelevant to assessing the problem, and thus multiple Control sites are unnecessary. In addition, the claim that causation cannot be inferred from the BACI model is refuted because of the fundamental reason that the Impact site is not randomly chosen, and therefore the analyses should not be considered that of a typical experiment (Stewart-Oaten and Bence 2001). Their final conclusion is that the BACI design is sound for use in impact assessments so long as statistical analyses are conducted carefully by checking assumptions, exploring sensitivity, inquiring where the "chance" comes from, checking models, and interpreting parameters and results (Stewart-Oaten and Bence 2001).

In Canada, any permanent alteration or destruction of fish habitat is assessed for impacts to affected fisheries to avoid violation of the revised *Fisheries Act*. Any residual serious harm to fish is addressed via an offsetting plan that includes a monitoring component with the objective of determining whether the proposed measures will be effective in offsetting the serious harm, and that such measures have stabilized and are functioning as intended (DFO 2013). Recent guidelines for effectiveness monitoring designs recommended the BACI design when possible, with a minimum of 3 years of pre-treatment monitoring and a range of options presented for post-term monitoring (Smokorowski et al. 2015). However, recognizing that there are often practical limitations on the duration, frequency, and intensity of monitoring that can be conducted as part of offsetting a

development project, a range of suboptimal designs may be adopted, with potential implications for the scientific soundness of the conclusions arising from such monitoring.

With the requirement for effectiveness monitoring now established in regulations, Fisheries and Oceans Canada (DFO) has an opportunity to guide proponent monitoring programs to be scientifically sound, clearly reported, and to contribute towards the long-term improvement of program effectiveness (Smokorowski et al. 2015). Conversely, poorly designed programs may hinder future decision-making ability by presenting, unknowingly, false conclusions about the ecological impacts of development or offsetting actions. With these important consequences in mind, we used data from a long-term BACI experiment as a case study to explore various design and analysis decisions. For illustration purposes, we used 9 years of total invertebrate and fish community biomass and diversity data (3 years pre and 6 years post) from two rivers, one with a hydroelectric facility that changed its operational regime, and one reference river. We used multiple models to explore the results of the BACI analyses using different combinations of years for the “After” period to test the implications of duration of monitoring and balanced vs. unbalanced designs. We used multiple statistical models to explore the effects of model choice on the results, and we examined the resulting confounding factors that can arise in the analyses and interpretation. Specific predictions tested include: (1) the BACI design results would be robust to the statistical model used; (2) the BACI design results would be robust to changes in continuous post-duration monitoring, but that intermittent sampling may affect results; and (3) that the BA or CI design may not produce the same results as the BACI models. Our intention is to show that caution should be employed when using, or interpreting results from, a BACI design in an environmental impact study, but that a well-designed BACI remains one of the best models for environmental effects monitoring programs.

Methods

Case study

To develop the scientific information necessary to design management tools that address both ecosystem integrity and energy efficiency, the DFO, the Ontario Ministry of Natural Resources and Forestry, Brookfield Renewable, and the University of Waterloo collaborated on a long-term, adaptive management experiment to test whether regulating ramping rates through hydroelectric turbines provided ecological benefits, while at the same time minimizing electricity production losses (Smokorowski et al. 2011). The main purpose of this experiment was to determine if removing all operational constraints on ramping rates from a hydroelectric facility was detrimental to the downstream riverine ecology. Aspects of river ecology that were monitored included hydrology, fluvial geomorphology, chemistry and plankton, benthic macroinvertebrates, fish populations and community structure, and food web ecology (via stable isotopes).

We used a BACI design for this experiment, which in this case involved comparing conditions on a regulated peaking river (Impact river; Magpie) to conditions on an unregulated river (Control river; Batchawana), Before and After implementing a change in ramping rates. The Control river had a natural flow regime and was carefully selected to be similar to the Impact system (Metcalfe et al. 2001). The intention was that this approach should allow detection of a change in the metrics of interest that were caused by the experimental ramping rate changes, as the Control river should reflect any changes in metrics resulting from regional environmental factors. Both rivers were similarly sampled during the entire field study, which included a Before period from 2002 to 2004 and an After period from 2005 onward (the planned final year of monitoring is 2017). The purpose of this paper is not to focus on the results of the Magpie project, but instead to use those results to examine the implications of using different BACI analyses methods and different post-impact monitoring designs.

Therefore, detailed methods on field sampling will not be provided, nor will a thorough analysis of results from a biological implications perspective.

Study sites

The experimental site was the Magpie River, Wawa, Ontario (48°0'N; 84°7'W) on the 40 km stretch between Steephill Falls and the Harris waterpower facilities (WPF) (Fig. 1). The reference river was the unregulated Batchawana River (47°0'N; 84°3'W), located approximately 60 km north of Sault Ste. Marie, Ontario. Both rivers were divided into transects spaced 500 m apart to provide a spatial reference and allow for the selection of sampling sites in a stratified random sampling design. The first 2.5 km below the dam on the Magpie River was characterized by large deep plunge pools and blasted channels that are atypical of a natural river system. For the Magpie, the most intensively sampled river segments were those considered the most vulnerable to ramping due to their riffle-run-pool riverine nature (total length 18 km, beginning 2.5 km downstream from the dam at Transect 5, Fig. 1) and these are the results presented here. The spatial and sampling coverage was similar for the Batchawana River.

Between 2002 and 2004, data were collected from the regulated Magpie River under the original restricted ramping rate regime: ramping rate could not exceed $1 \text{ m}^3 \cdot \text{s}^{-1} \cdot \text{h}^{-1}$ from 10 October to 15 November; $2 \text{ m}^3 \cdot \text{s}^{-1} \cdot \text{h}^{-1}$ from 16 November until spring freshet (early May); from May until early October, the dam was restricted to an increase or decrease of 25% of the previous hour's flow. From 2005 to 2010, data were collected with no restrictions on ramping and while the Steephill Falls plant operated in accordance with water availability and market forces (Fig. 2). During the majority of the study period (with one exception noted in the Results), through all seasons the Steephill Falls WPF could not release a discharge lower than $7.5 \text{ m}^3 \cdot \text{s}^{-1}$ as that was the regulated minimum flow. All sampling from the Batchawana River was done contemporaneously.

Field sampling

Hydrological data were obtained both from Brookfield Renewable (for data from the Steephill Falls Generating Station on the Magpie River), and from the Water Survey of Canada gauge on the Batchawana River (02BF001). Invertebrate data are from an annual sampling of the community in the thalweg of the river using rock bags. In each year at each site, five mesh rock bags were randomly placed in a riffle, ensuring a minimum distance of 3 m apart, and at a depth to maintain a sufficient flow over the bags throughout low water periods. The rock bags were constructed out of 5.1 cm net mesh, were 122 cm in circumference and 46 cm in length, and were filled with rocks of representative size found along the shoreline at the site of placement until each reached a weight of 7 kg (± 0.5 kg). The actual number of rocks used, their diameter, and the weight of each bag was recorded, as were the water depth and velocity (Marsh McBirney Flomate 2000 Portable Flow Meter) at each bag location. The bags were left in the river for a period of approximately 60 d (June to August), a sufficient length of time for full colonization to reach fluctuating taxa richness, abundance, and biomass (Mason et al. 1973; Shaw and Minshall 1980). Once bags were retrieved the rocks were cleaned and all invertebrates and debris were preserved in 70% ethanol. The entire sample was subsampled for identification to taxonomic level of family, and enumeration, although in each year a number of samples were identified in their entirety to allow for the calculation of accuracy and precision of subsampling procedure, which were always found to be within acceptable limits (defined as being within 20% of true counts, Elliott 1977).

Backpack electrofishing was used to examine relative fish population abundance, biomass, growth, species richness, diversity, and community composition on both rivers. Electrofishing was conducted in July of each year in the Batchawana River and in August of each year in the Magpie River. At each randomly selected transect, all areas ≤ 60 cm in depth were sampled using backpack electrofishing

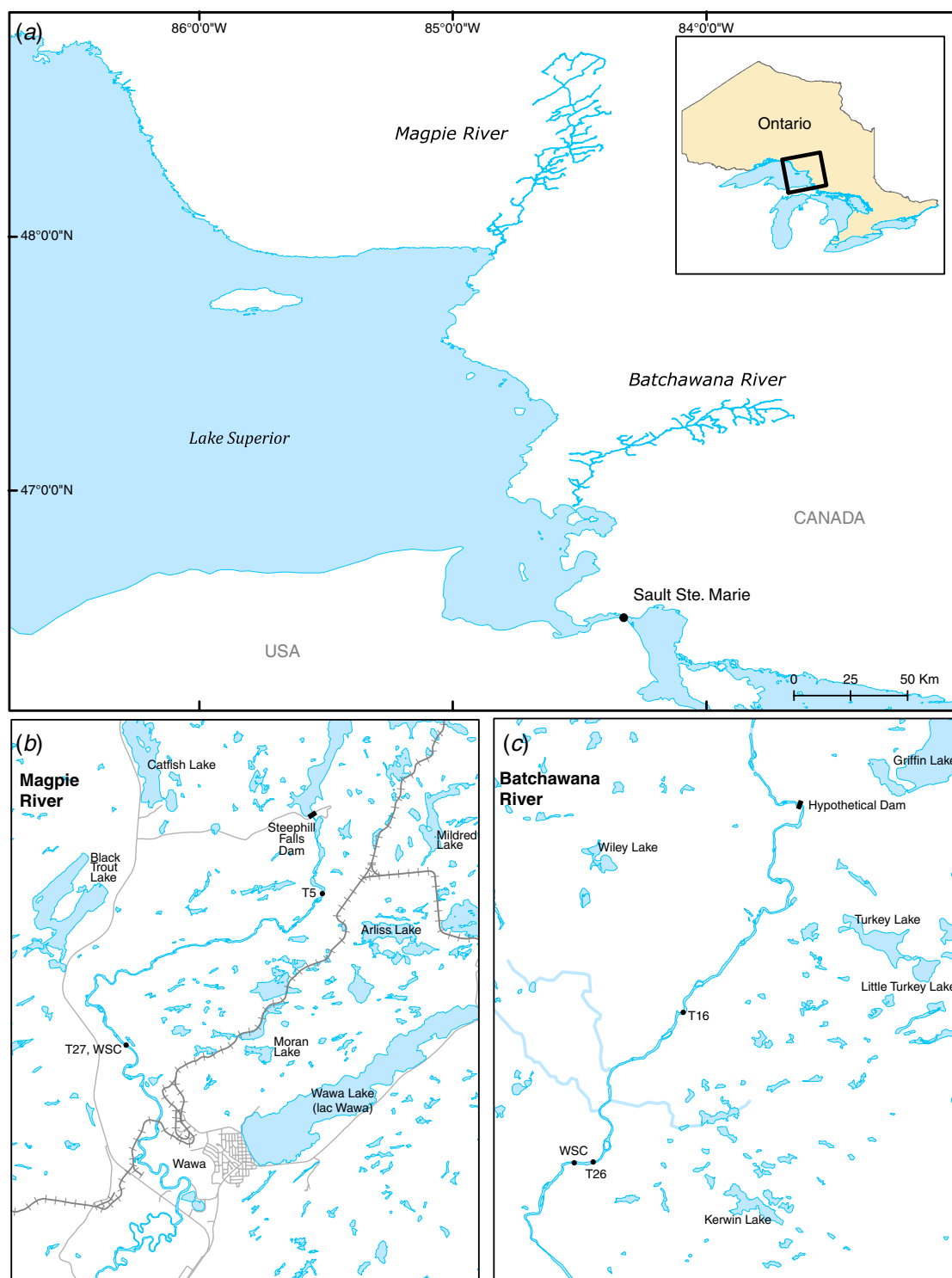


Fig. 1. Location of the (a) Magpie (altered) and Batchawana (natural) Rivers in Ontario, Canada, and detailed sections of the study reaches of the (b) Magpie and (c) Batchawana, showing Steephill Falls Waterpower Facility, select transect locations (e.g., T5 and T27), with Water Survey of Canada stream gauges shown for reference. Map modified from Smokorowski et al. (2011), and reproduced with permission from John Wiley & Sons, Inc.

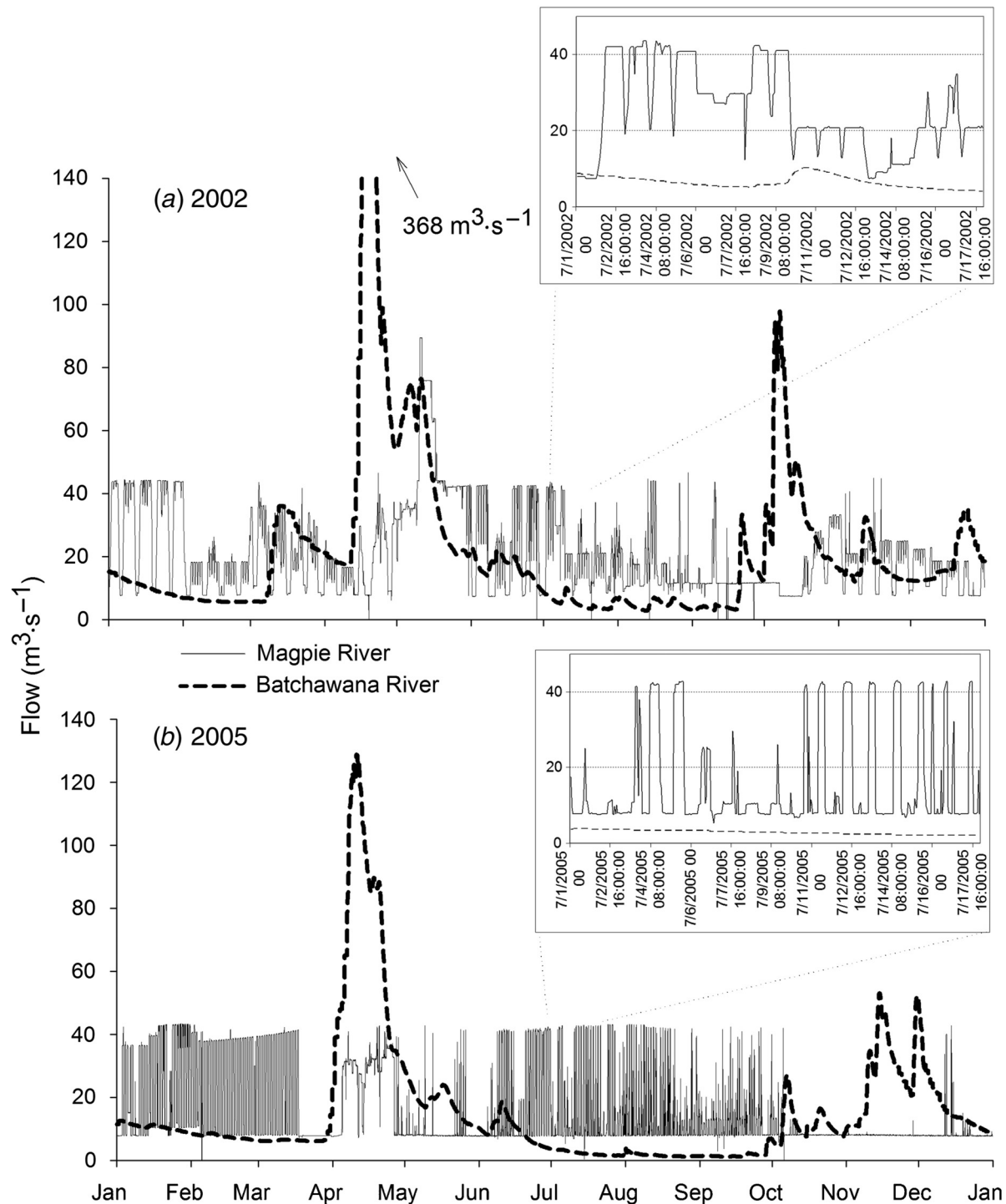


Fig. 2. Annual hydrograph of the Magpie (solid) and Batchawana (dotted) Rivers, (a) in 2002, before ramping change, and (b) in 2005, after unlimited ramping of the Magpie. Data for the Magpie from the Steephill Falls Waterpower Facility (provided by and used with permission from Brookfield Renewable); data for the Batchawana River obtained from the Water Survey of Canada, Environment Canada, gauge 02BF001. Insert plots highlight discharge patterns.

covering eight 100 m shoreline segments on each river. These segments were sampled according to habitat type (fast or slow) using a standard back-and-fourth electrofishing technique, and shockers were standardized by power (W), rate ($\text{s}\cdot\text{m}^{-2}$), area (m^2), and time (s). Fish were identified to species and enumerated by habitat area, which were averaged to the transect level. A random subsample of the captured fish was preserved for accurate length, weight and age data. Fish biomass per unit area ($\text{g}(100\text{ m}^2)^{-1}$) was calculated for each sampling site.

As backpack electrofishing was the exclusive method of fish capture, the fraction of the fish community vulnerable to capture were smaller individuals of larger species, or small species residing in shallower habitats (mean fish length captured was $50\text{ mm} \pm 26\text{ mm}$). The rationale was that relative changes observed in this proportion of the fish community would reflect early life history impacts on larger fish species. In addition, as home range size increases allometrically with body size (Minns 1995), sampling smaller body sizes ensured samples were representative of local conditions. Because the home range of fish 150 mm or smaller is $<500\text{ m}^2$ (Minns 1995), and the average distance between adjacent transects was 571 m (min–max = 200–5000 m) on rivers with approximate mean widths of $>40\text{ m}$, fishing transects were considered independent replicates.

Statistical analysis

All statistics were calculated using Statistica version 10 (StatSoft Inc., Tulsa, Oklahoma) software. Results were considered significant at p -values that were adjusted using the sequentially rejective Bonferroni test (or Holm–Bonferroni; Holm 1979), corrected for the number of tests within each data set and family of tests. We chose to use the Holm–Bonferroni test to minimize the family-wise error rate for multiple tests, because it is considered more powerful than the overly conservative Bonferroni method which runs the risk of not detecting real effects. We did not consider the full suite of tests in our corrections given that any one study would not normally conduct all analyses presented here, and we felt that correction would be overly conservative. Simply stated, with the Holm–Bonferroni test p -values for your family of tests (n) are ordered from smallest to largest, with the smallest p -value considered relative to α/n (where $\alpha = 0.05$), the same adjustment as the Bonferroni method. If the first test is found to be not-significant, the procedure stops there; but if it is significant, you compare the second lowest p -value to $\alpha/(n-1)$, and then continue the procedure until the sequentially larger p -value is found to be not significant, at which point all testing stops. Data were tested for parametric assumptions (normality and homogeneity of variance of the data) and transformed when necessary. Residuals from analyses of transformed data were tested for normality. Additivity was tested on transformed data by testing for a zero slope in the regression of the difference and averages of corresponding points between the Impact and Control rivers, which essentially corresponds to Tukey's test for additivity (Stewart-Oaten et al. 1986). Mean annual flow for each river was calculated as the mean discharge from January to January of each year, and a correlation analysis was run between rivers to test if changes in flow among years were similar.

Fish species and invertebrate families were used to calculate diversity (probability of interspecific encounter (PIE); Hurlbert 1971). PIE is an unbiased diversity measure that calculates the chance that two individuals drawn at random from a population represent different families:

$$\text{PIE} = \sum_{i=1}^s (n_i/n)[(n - n_i)/(n - 1)] \quad (1)$$

where n is the number of all individuals in the sample, n_i is the number of individuals of a family in the sample, and s is the number of families (Hurlbert 1971). PIE was selected over other diversity indices because it provides a statistically and biologically understandable probability (out of 100%; the higher the number the more diverse the community), unlike more traditional diversity measures (Gottelli and Graves 1996).

Mean annual invertebrate community abundance (numbers per rock bag), and mean annual fish biomass-per-unit-effort (by habitat type), and diversity for both taxon groups, were calculated and plotted to illustrate annual variability. To explore the effect of using different statistical models on the results, multiple approaches were used. First, to assess the effect of differing monitoring durations, two-way ANOVAs were run on the BACI data for fish and invertebrate biomass and diversity, using three options for the “after” period, each representing a different combination of continuous 3 year monitoring blocks: (1) 2005–2007 representing the first 3 years post-ramping change and a balanced design (short-term monitoring); (2) 2005–2010 representing a 6 year longer-term monitoring period but creating an unbalanced design, which is unavoidable with a fixed duration pre-period (complete long-term monitoring); and (3) 2008–2010 representing a long-term response period but keeping the design balanced for statistical purposes (balanced long-term monitoring). The main statistic of interest in a BACI analysis is the interaction term (Before-After \times Control-Impact), which would be significant when a change occurs at the impact site but not at the control site. To explore the effect of using an alternate statistical model, as recommended by [Stewart-Oaten et al. \(1986\)](#), differences between the Impact minus Control sites were calculated using means by transect and year, and a t test was run on the Before and After periods (including the three continuous block After durations outlined above) to assess if the BACI was significant. To test if the unbalanced design using the 6 year complete long-term monitoring had an effect on the BACI model validity, we compared the standard ANOVA result with the full data set to the same data using a restricted maximum likelihood estimation (REML) ANOVA, which is considered more robust for unbalanced designs ([Robinson 1987](#); [Fletcher and Underwood 2002](#)).

In addition, recognizing that there are often pressures to minimize the duration and frequency of post-monitoring requirements, and that there are often practical limitations on monitoring designs, our data were used to test a variety of possible post-impact sampling timetables, using a range of combinations of intermittent sampling in the BACI analyses. Finally, recognizing that sometimes the incorporation of a control system is not feasible, or that the development will occur without the opportunity to collect pre-treatment data, we examined the results from conducting a Before-After only analysis (BA: no Control), and a Control-Impact only analysis (CI: no Before), to examine the conclusions that would have resulted from these suboptimal designs.

Permitting

Fish for this project were collected under the Ontario Ministry of Natural Resources and Forestry Collection of Fish for Scientific Purposes permit number 1000784. Animal care permits for this project were approved by the Canadian Council for Animal Care certified GLLFAS/NWRI Animal Care Committee (permit number GLLFAS/NWRI 0907). Work was reviewed and permits issued on an annual basis.

Results

All abundance data were $\ln(x + 1)$ transformed and PIE was arcsine transformed to meet assumptions of additivity, normality, and homogeneity of variance. All transformed data met assumptions except for the invertebrate diversity measure (arcsinePIE), which did not meet the additivity assumption. In addition, residuals from the BACI analyses were not significantly different from normal with the exception of the fish diversity measure (arcsinePIE) in the slow habitat. We present the analyses for invertebrate and slow fish diversity despite these violations, but interpret the results with caution.

The natural hydrology of the Batchawana resulted in greater peak flows and lower minimum flows relative to the altered Magpie River ([Fig. 2](#)). In 2002, when ramping rate was restricted, the dam operated on a reduced peaking cycle. During the week when water supply was high, flow was “perched” on an elevated minimum, and on weekends when demand was low water levels did not reach full turbine flow ([Fig. 2a](#)). In 2005, however, full ramping from the maximum turbine

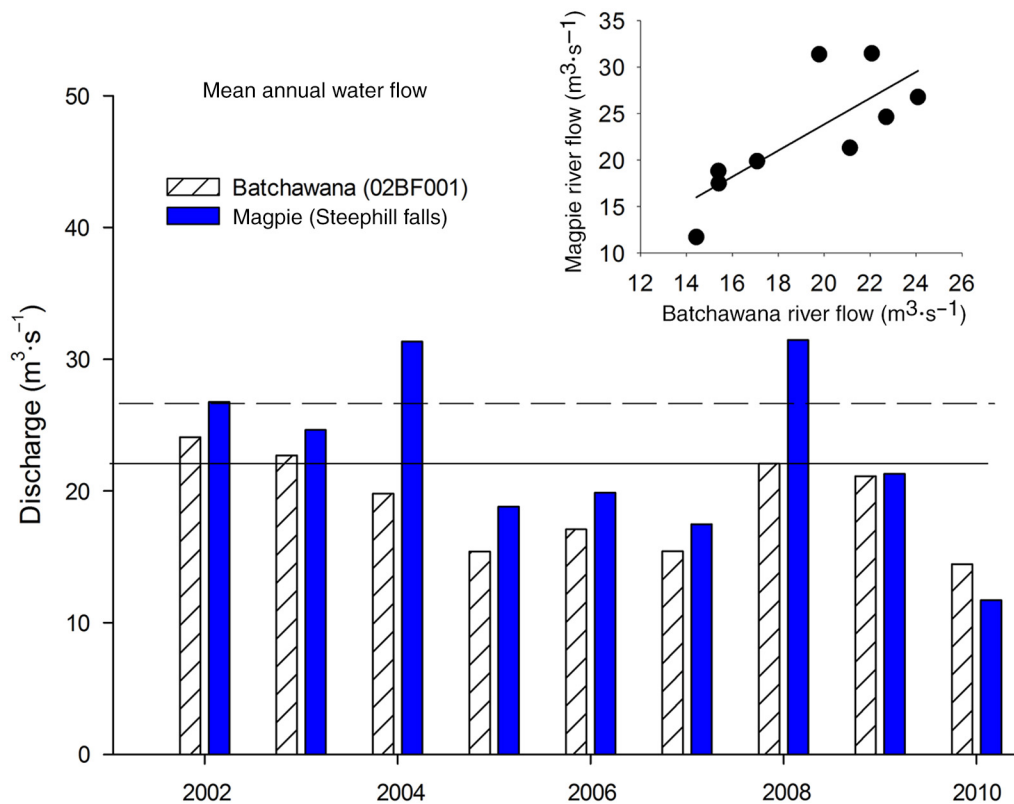


Fig. 3. Annual water flow (discharge in $\text{m}^3\cdot\text{s}^{-1}$) averaged from January to January for the Magpie and Batchawana Rivers from 2002 to 2010. The solid line represents the historical mean annual flow for the Batchawana ($22 \text{ m}^3\cdot\text{s}^{-1}$) and the dashed line represents the historical mean annual flow for the Magpie pre-dam ($27 \text{ m}^3\cdot\text{s}^{-1}$). Inset scatterplot correlation $r = 0.77$, $p = 0.01$. Data for the Magpie from the Steephill Falls Waterpower Facility (provided and used with permission from Brookfield Renewable); data for the Batchawana River obtained from the Water Survey of Canada, Environment Canada, gauge 02BF001.

discharge to minimum regulated flow occurred at a much greater frequency because the rate of change was unrestricted (Fig. 2b). Mean annual flow of the rivers tracked each other over time confirming the suitable use of the Batchawana River as a regional control for the Magpie River ($r = 0.78$, $p = 0.01$, Fig. 3). Coincidentally, 2005, which was the year corresponding to the change in unlimited ramping, also saw the beginning of a 3 year period of drought on both rivers. The return to normal annual flow levels was observed on both rivers in 2008 and 2009, but 2010 was again a relatively dry year (Fig. 3).

The average number of invertebrates per rock bag was generally greater for the Magpie than the Batchawana, and the pattern among years was roughly similar (Fig. 4a). Invertebrate diversity was initially greater for the Magpie River, but in 2005 the Batchawana diversity became higher than the Magpie until it switched back again in 2009 (Fig. 4b). The BACI analysis interaction was not significant for all year combinations for abundance (Table 1 and Fig. 5a), and was significant for diversity when both short and complete long-term monitoring were tested, but not when the balanced long-term monitoring was tested (Table 1 and Fig. 5b). Calculating the difference between Impact minus Control means by site and running the two sample t test yielded similar results for diversity, but different results for abundance, which was found to be significant against the adjusted p_{crit} values (Table 2). The results for the REML ANOVA for unbalanced

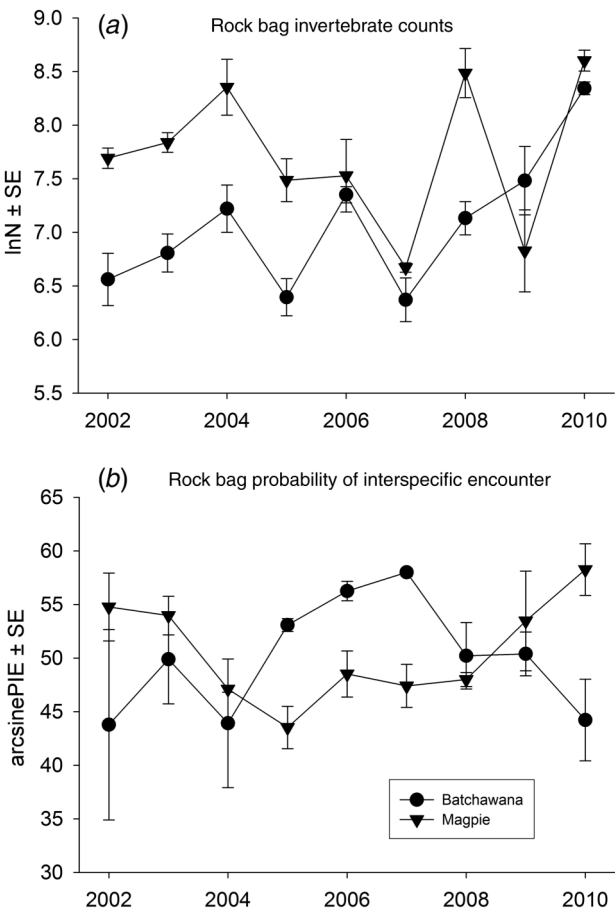


Fig. 4. Average (a) abundance (number per rock bag, ln transformed) and (b) diversity (PIE, arcsine transformed) of invertebrates per rock bag \pm standard error plotted as average across sites for each year 2002–2010.

Table 1. Before-After-Control-Impact (BACI) ANOVA interaction term Before-After (BA) \times Control-Impact (CI) sum of squares (SS), F value (degrees of freedom), and significance level p for invertebrate abundance (lnN) and diversity (arcsinePIE) in continuous monitoring designs.

After years included		SS	$F_{(df)}$	p	p_{crit}
Invertebrate lnN BACI					
1, 2, 3, 4, 5, 6	Long-term unbalanced	1.8	3.8 _(1,66)	0.05	—
1, 2, 3	Short-term balanced	1.1	4.3 _(1,48)	0.04	—
4, 5, 6	Long-term balanced	1.6	4.3 _(1,40)	0.04	0.0125
Invertebrate arcsinePIE BACI					
1, 2, 3, 4, 5, 6	Long-term unbalanced	412.5	6.7_(1,66)	0.01	0.01
1, 2, 3	Short-term balanced	824.1	13.4_(1,48)	<0.001	0.008
4, 5, 6	Long-term balanced	6.5	0.08 _(1,40)	0.78	—

Note: In all cases the Before period included 2002–2004. Years included in the After period varied as indicated, where 1 = 2005, 2 = 2006, 3 = 2007, 4 = 2008, 5 = 2009, and 6 = 2010. p_{crit} provides the critical p -value as adjusted by the Holm–Bonferroni method for determining significance, where in this case the total number of tests was 6. Bold text indicates a significant difference.

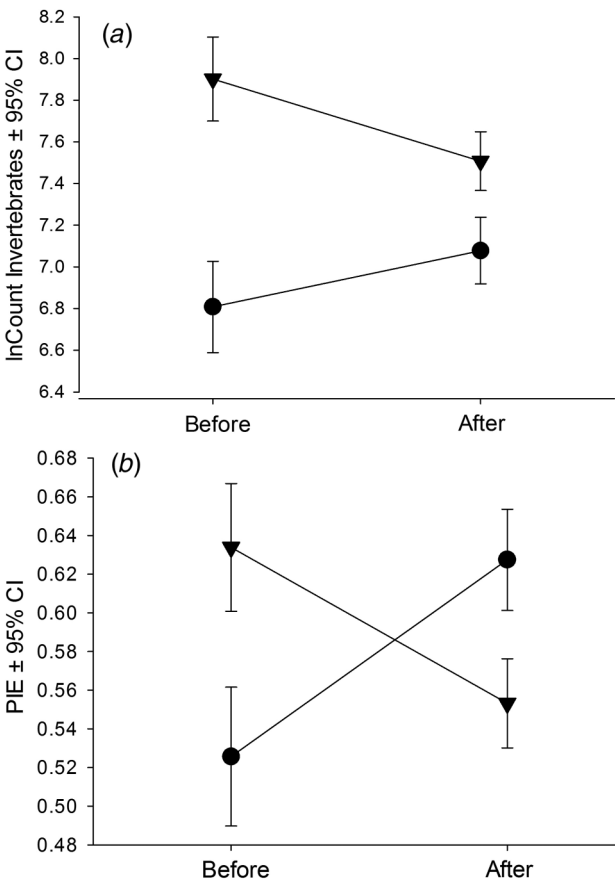


Fig. 5. Average (a) abundance (number per rock bag) and (b) diversity (probability of interspecific encounter) of invertebrates per rock bag \pm 95% confidence interval (CI) plotted as the Before-After-Control-Impact (BACI) plot, for the Impact (triangles) and Control (circle) rivers. Representative results are shown including the Before period (2002–2004) and the full After period (2005–2010).

Table 2. The *t* test results on differences between Impact minus Control means of invertebrate rock bags by site, using Before-After as the grouping variable in continuous monitoring designs.

Source of variation—invertebrates	Mean difference Before \pm SE (<i>n</i>)	Mean difference After \pm SE (<i>n</i>)	<i>t</i>	<i>p</i>	<i>p</i> _{crit}
lnN (1, 2, 3, 4, 5, 6)	1.1 \pm 0.13 (13)	0.43 \pm 0.17 (22)	2.7	0.01	0.0125
lnN (1, 2, 3)	1.1 \pm 0.13 (13)	0.51 \pm 0.16 (13)	2.8	<0.01	0.01
lnN (4, 5, 6)	1.1 \pm 0.13 (13)	0.32 \pm 0.35 (9)	2.4	0.03	0.025
arcsinePIE (1, 2, 3, 4, 5, 6)	6.5 \pm 3.4 (13)	−3.5 \pm 2.1 (22)	2.6	0.01	0.017
arcsinePIE (1, 2, 3)	6.5 \pm 3.4 (13)	−9.4 \pm 1.2 (13)	4.4	<0.001	0.008
arcsinePIE (4, 5, 6)	6.5 \pm 3.4 (13)	5.0 \pm 3.2 (9)	0.32	0.75	—

Note: In all cases the Before period included 2002–2004. Years included in the After period varied as indicated, where 1 = 2005, 2 = 2006, 3 = 2007, 4 = 2008, 5 = 2009, and 6 = 2010. *p*_{crit} provides the critical *p*-value as adjusted by the Holm–Bonferroni method for determining significance. Bold text indicates a significant difference from before to after the change in dam operations.

designs were the same as the standard BACI factorial ANOVA. As mentioned previously, invertebrate diversity statistics should be interpreted with caution due to the violation of the additivity assumption.

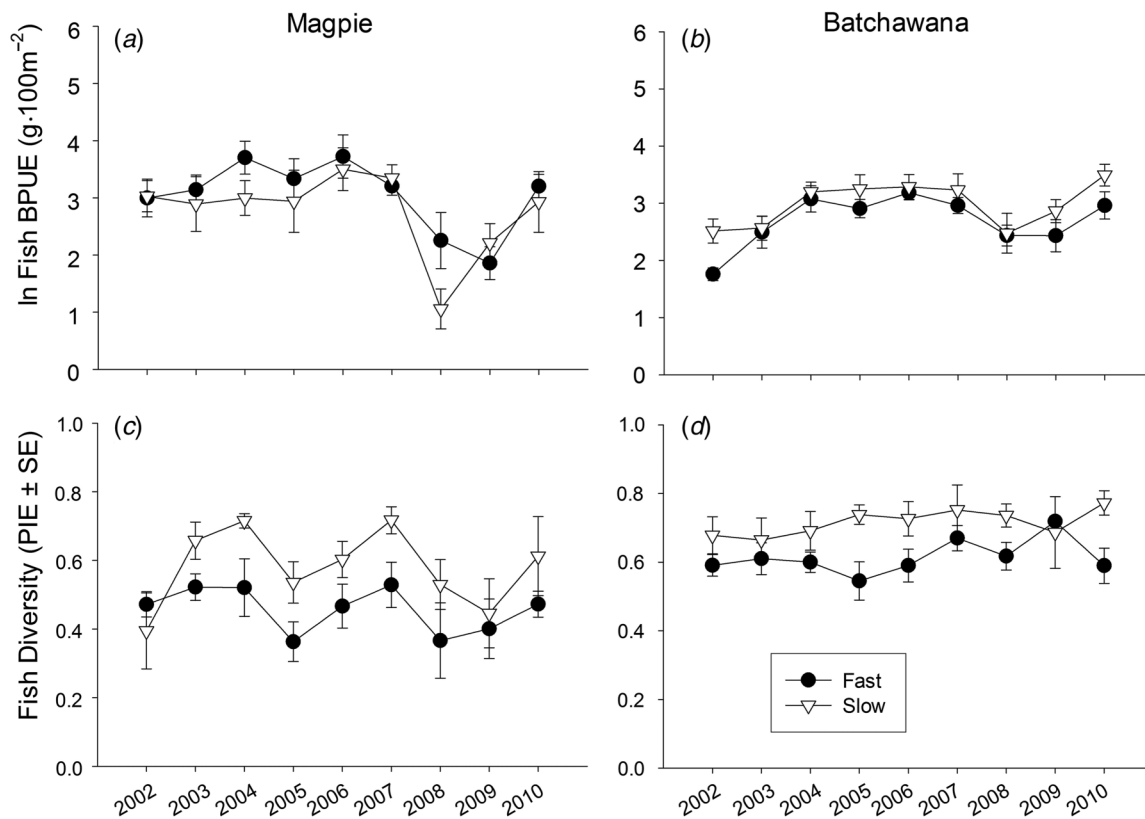


Fig. 6. Average fish biomass-per-unit-effort (lnBPUE) \pm standard error from 2002 to 2010 for (a) the Magpie River and (b) the Batchawana River, and average fish diversity (PIE) \pm standard error from 2002 to 2010 for (c) the Magpie River and (d) the Batchawana River.

Fish biomass per unit effort (BPUE) followed a similar pattern among years on both rivers, although overall interannual variability of the Batchawana was less than the Magpie (Figs. 6a, 6b). For the fish data analyses the total number of tests considered in the Holm–Bonferroni correction was 12, resulting in no significant results regardless of the years included in the design (Table 3 and Figs. 7a, 7b). Calculating the difference between Impact minus Control biomass means by transect and running the two sample *t* test found the difference between the Before and After period were not significant for either habitat type in the short- or long-term monitoring, but both habitat types were significant when testing the balanced long-term monitoring design (Table 4). Again, the REML ANOVA results were the same as the BACI ANOVA results.

Fish diversity in both habitat types was greater for the Batchawana than for the Magpie River (Figs. 6c, 6d). The interannual pattern of fish diversity in fast habitat was similar among years between the two rivers except in 2009. Fish diversity in the slow habitat was relatively consistent among years for the Batchawana River, whereas the interannual variability of the Magpie was great. The interaction term for fish diversity was not significant for any combination of years in either habitat type (Table 3 and Figs. 7c, 7d). The *t* test results similarly demonstrated that the difference in fish diversity from Before to After was not significant for any year combination (Table 4).

Exploring various intermittent sampling designs demonstrated that some significant impacts were no longer apparent, and very different conclusions could result depending on the years included in the monitoring program (Tables 5 and 6). The use of Before–After Magpie River only design

Table 3. Before-After-Control-Impact (BACI) ANOVA interaction term Before-After (BA) \times Control-Impact (CI) sum of squares (SS), F value (degrees of freedom), and significance level p for fish relative biomass (lnBPUE) and diversity (arcsinePIE) in continuous monitoring designs.

After years included	Design type	Habitat	SS	$F_{(df)}$	p	p_{crit}
Fish lnBPUE BACI						
1, 2, 3, 4, 5, 6	Long-term unbalanced	Fast	4.2	5.5 _(1,140)	0.02	—
1, 2, 3, 4, 5, 6	Long-term unbalanced	Slow	3.4	3.3 _(1,132)	0.07	—
1, 2, 3	Short-term balanced	Fast	1.1	2.0 _(1,92)	0.16	—
1, 2, 3	Short-term balanced	Slow	0.34	0.45 _(1,88)	0.50	—
4, 5, 6	Long-term balanced	Fast	6.1	7.9 _(1,92)	0.006	—
4, 5, 6	Long-term balanced	Slow	7.3	7.8 _(1,84)	0.006	0.004
Fish arcsinePIE BACI						
1, 2, 3, 4, 5, 6	Long-term unbalanced	Fast	123	0.95 _(1,140)	0.33	—
1, 2, 3, 4, 5, 6	Long-term unbalanced	Slow	311	1.4 _(1,132)	0.23	—
1, 2, 3	Short-term balanced	Fast	14.9	0.15 _(1,90)	0.69	—
1, 2, 3	Short-term balanced	Slow	95	0.57 _(1,88)	0.45	—
4, 5, 6	Long-term balanced	Fast	237	1.7 _(1,92)	0.19	—
4, 5, 6	Long-term balanced	Slow	453.8	1.7 _(1,84)	0.19	—

Note: In all cases the Before period included 2002–2004. Years included in the After period varied as indicated, where 1 = 2005, 2 = 2006, 3 = 2007, 4 = 2008, 5 = 2009, and 6 = 2010. p_{crit} provides the critical p -value as adjusted by the Holm–Bonferroni method for determining significance, where in this case the total number of tests was 12.

(no Control) was not able to detect a significant change in either the fish or invertebrate community in the impact river, although in most cases there was a decreasing trend in the metric from Before to After the change to unlimited ramping (Table 7). The Control-Impact design (no Before) was similarly not able to detect a difference in abundance between the Batchawana and Magpie Rivers, except that it did capture that both invertebrate and fish diversity was lower in the impact Magpie River relative to the control Batchawana River (Table 7).

Discussion

The purpose of this paper is to explore how results and conclusions can change depending on the design and analysis of the monitoring program, and to raise awareness of resulting implications. While it seems clear that the change in ramping rate regime had an effect on the biota of the Magpie River (details of the biotic response are discussed elsewhere), the interpretation of the data and conclusions of the statistical significance of those effects varied depending on the number and configuration of years included in the analysis, and the statistical model used to analyze the data. The computational procedure used to analyze the unbalanced BACI design resulting from continuous long-term monitoring with a fixed baseline phase did not have an effect on the interpretation of the data. However, it did appear that analyzing the difference between the reference and control before and after the impact occurred (using a t test as recommended by Stewart-Oaten et al. 1986), was a more powerful statistical test than the BACI ANOVA, as significant differences were declared despite the higher bar set by the Holm–Bonferroni adjusted p_{crit} values.

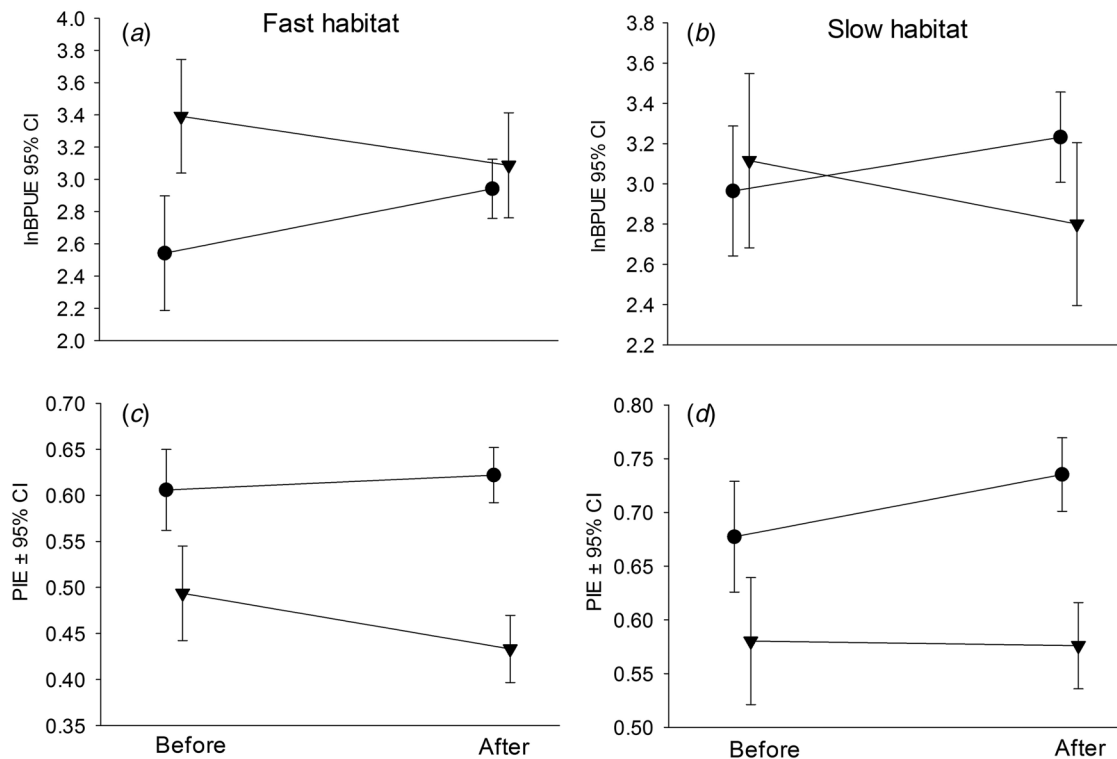


Fig. 7. Representative BACI plots (all data included 2002–2010) for the Control (circles) and Impact (triangles) sites of average fish biomass per unit effort (BPUE) \pm 95% confidence intervals for (a) fast habitat and (b) slow habitat; BACI plots of average probability of interspecific encounter (PIE) \pm 95% confidence intervals for the Control (circles) and Impact (triangles) sites for (c) fast habitat and (d) slow habitat.

Intermittent sampling designs often failed to detect effects that were found with continuous sampling, as did the suboptimal designs used when no Control system or Before data were available. We are unaware of any similar exploration of design and analysis options on environmental effects monitoring results and conclusions. Although others have examined optimal study designs for environmental effects or restoration programs (e.g., [Benedetti-Cecchi 2001](#); [Munkittrick et al. 2002, 2009](#); [Liermann and Roni 2008](#)), few have had the duration and consistency of data necessary for such analyses.

Using a balanced number of years before and after affecting a perturbation is usually considered important in statistical analyses due to its effect on the power of the test, the implications regarding the assumption of equality in variance, and in how you estimate parameters and partition the sum of squares (SS) to test the hypotheses of interest ([Shaw and Mitchell-Olds 1993](#)). Yet due to a myriad of factors, ecologists rarely have completely balanced data and need to face decisions about imposing balance by eliminating values (thus losing information), or to use computational methods that are specifically designed for unbalanced data ([Robinson 1987](#); [Shaw and Mitchell-Olds 1993](#)). For any long-term environmental effects monitoring study, the period of pre-treatment or “before” data is often limited, and therefore the longer duration of monitoring results in greater imbalance in the design. Yet long-term monitoring programs, although relatively rare especially in environmental effects assessments, are considered important to capture the full manifestation of management actions as system response times can be on the order of decades ([Burt et al. 2008](#); [Roni et al. 2008](#); [Lindenmayer and Likens 2009](#)). Therefore, reconciling the trade-off between long-term monitoring and balanced statistical analyses becomes necessary.

Table 4. The *t* test results on differences between Impact minus Control means of fish relative biomass (lnBPUE) and diversity (arcsinePIE) averaged by transect, using Before-After as the grouping variable in continuous monitoring designs.

Source of variation—fish lnBPUE or diversity (arcsinePIE)	Habitat	Mean difference Before \pm SE (<i>n</i>)	Mean difference After \pm SE (<i>n</i>)	<i>t</i>	<i>p</i>	<i>p</i> _{crit}
lnBPUE (1, 2, 3, 4, 5, 6)	Fast	0.84 \pm 0.20 (24)	0.12 \pm 0.16 (46)	2.6	0.01	0.005
	Slow	0.25 \pm 0.22 (22)	−0.43 \pm 0.20 (46)	2.0	0.05	—
lnBPUE (1, 2, 3)	Fast	0.84 \pm 0.20 (24)	0.40 \pm 0.19 (24)	1.6	0.12	—
	Slow	0.25 \pm 0.22 (22)	0.005 \pm 0.23 (24)	0.69	0.49	—
lnBPUE (4, 5, 6)	Fast	0.84 \pm 0.20 (24)	−0.17 \pm 0.26 (24)	3.1	0.003	0.0045
	Slow	0.25 \pm 0.23 (22)	−0.90 \pm 0.28 (22)	3.2	0.003	0.004
arcsinePIE (1, 2, 3, 4, 5, 6)	Fast	−8.5 \pm 2.6 (24)	−12.5 \pm 2.6 (48)	0.96	0.34	—
	Slow	−5.6 \pm 4.0 (22)	−12.1 \pm 3.1 (46)	1.2	0.22	—
arcsinePIE (1, 2, 3)	Fast	−8.5 \pm 2.6 (24)	−10.1 \pm 3.1 (24)	0.39	0.70	—
	Slow	−5.6 \pm 4.0 (22)	−9.7 \pm 2.5 (24)	0.88	0.38	—
arcsinePIE (4, 5, 6)	Fast	−8.5 \pm 2.6 (24)	−14.8 \pm 4.1 (24)	1.3	0.20	—
	Slow	−5.6 \pm 4.0 (22)	−14.7 \pm 5.8 (22)	1.3	0.20	—

Note: In all cases the Before period included 2002–2004. Years included in the After period varied as indicated, where 1 = 2005, 2 = 2006, 3 = 2007, 4 = 2008, 5 = 2009, and 6 = 2010. *p*_{crit} provides the critical *p*-value as adjusted by the Holm–Bonferroni method for determining significance. Bold text indicates a significant difference from before to after the change in dam operations.

Table 5. Before-After-Control-Impact (BACI) ANOVA interaction term Before-After (BA) \times Control-Impact (CI) sum of squares (SS), *F* value (degrees of freedom), and significance level *p* for invertebrate relative biomass (lnN) and diversity (arcsinePIE) in intermittent monitoring designs.

After years included		SS	<i>F</i> _(df)	<i>p</i>	<i>p</i> _{crit}
Invertebrate lnN BACI					
1, 3, 6	BA \times CI	0.91	1.9 _(1,46)	0.17	—
2, 4, 6	BA \times CI	0.82	3.0 _(1,42)	0.09	—
1, 3, 5	BA \times CI	1.8	7.0_(1,46)	0.01	0.01
Invertebrate arcsinePIE BACI					
1, 3, 6	BA \times CI	351.9	4.4 _(1,46)	0.04	0.0125
2, 4, 6	BA \times CI	104.3	1.3 _(1,42)	0.26	—
1, 3, 5	BA \times CI	556.8	7.9_(1,46)	0.007	0.008

Note: In all cases the Before period included 2002–2004. Years included in the After period varied as indicated, where 1 = 2005, 2 = 2006, 3 = 2007, 4 = 2008, 5 = 2009, and 6 = 2010. *p*_{crit} provides the critical *p*-value as adjusted by the Holm–Bonferroni method for determining significance. Bold text indicates a significant difference.

In our example, the original balanced design was termed “short-term” monitoring because it included only 3 years post-impact monitoring. While not having any statistical concerns due to imbalanced design, if sampling ended in 2007 we would have concluded that the unlimited ramping had no significant effect on relative fish biomass or diversity, but did have a significant effect on invertebrate

Table 6. Before-After-Control-Impact (BACI) ANOVA interaction term Before-After (BA) \times Control-Impact (CI) sum of squares (SS), F value (degrees of freedom), and significance level p for fish relative biomass (lnBPUE) and diversity (arcsinePIE) in intermittent monitoring designs.

After years included	Habitat	SS	$F_{(df)}$	p	p_{crit}
Fish lnBPUE BACI					
1, 3, 6	Fast	1.7	3.2 _(1,92)	0.08	—
1, 3, 6	Slow	1.3	1.7 _(1,86)	0.20	—
2, 4, 6	Fast	2.4	3.1 _(1,92)	0.08	—
2, 4, 6	Slow	4.0	3.6 _(1,86)	0.06	—
1, 3, 5	Fast	3.9	5.6 _(1,92)	0.02	0.004
1, 3, 5	Slow	1.5	1.9 _(1,86)	0.17	—
Fish arcsinePIE BACI					
1, 3, 6	Fast	25.6	0.28 _(1,92)	0.60	—
1, 3, 6	Slow	78.7	0.38 _(1,86)	0.54	—
2, 4, 6	Fast	21.6	0.16 _(1,92)	0.69	—
2, 4, 6	Slow	190.2	0.94 _(1,86)	0.34	—
1, 3, 5	Fast	213.5	2.1 _(1,92)	0.15	—
1, 3, 5	Slow	285.3	1.2 _(1,86)	0.27	—

Note: In all cases the Before period included 2002–2004. Years included in the After period varied as indicated, where 1 = 2005, 2 = 2006, 3 = 2007, 4 = 2008, 5 = 2009, 6 = 2010. p_{crit} provides the critical p -value as adjusted by the Holm–Bonferroni method for determining significance.

Table 7. One-way ANOVA results testing a Before-After Magpie only comparison (no Control), and a Control-Impact After only comparison (no Before) for invertebrate and fish abundance (lnN and lnBPUE respectively) and diversity (arcsinePIE) including all 6 years post monitoring.

Test	Habitat	lnN/Fish BPUE $F_{(df)}$, p , (trend)	arcsinePIE $F_{(df)}$, p , (trend)	p_{crit}
Invertebrates Before-After (no Control)		2.6 _(1,33) , 0.12, (↓)	2.6 _(1,33) , 0.11, (↓)	—
Invertebrates Control-Impact (no Before)		3.2 _(1,42) , 0.08, (↑)	4.0 _(1,42) , 0.05, (↓)	0.0125
Fish Before-After (no Control)	Fast	1.9 _(1,70) , 0.18, (↓)	0.32 _(1,70) , 0.57, (↓)	—
	Slow	0.87 _(1,66) , 0.35, (↓)	0.004 _(1,66) , 0.95, (↔)	—
Fish Control-Impact (no Before)	Fast	0.41 _(1,94) , 0.52, (↑)	24.0 _(1,94) , <0.001, (↓)	0.00625
	Slow	3.5 _(1,90) , 0.06, (↓)	15.6 _(1,90) , <0.001, (↓)	0.007

Note: The trend arrow indicates the directional difference (increasing or decreasing) between the mean values from Before to After or between Control and Impact. p_{crit} provides the critical p -value as adjusted by the Holm–Bonferroni method for determining significance, where in this case the total number of tests was 4 for invertebrates and 8 for fish.

abundance (using the difference tests) and diversity. After 6 years of post-impact monitoring differences in fish biomass were greater but still not significant using the adjusted p_{crit} values. Using a computational procedure specifically designed for unbalanced designs (REML ANOVA,

Robinson 1987) did not alter the significance of our BACI interaction term, or the interpretation of our data. If our main concern was achieving statistical balance, this was easily achieved by eliminating data and analyzing 3 year packages, which we refer to as balanced long-term monitoring. Losing years of data would result in detecting a statistically significant difference in fish biomass results, but eliminating significance in invertebrate diversity, and would have altered our conclusions. Therefore, achieving statistical balance by eliminating data should not be considered an option in environmental effects monitoring, preferentially using a computational procedure specific to unbalanced designs (available in most statistical packages) to ensure that statistical bias is not introduced by long-term environmental effects monitoring.

Smokorowski et al. (2015) provided recommendations for effectiveness monitoring designs that included: (1) 3 years before development at the impact site to establish a baseline, (2) 3 years sampling immediately post-change, (3) an additional 3 years of sampling at a later time (e.g., 4–6 years post or some later time), and (4) a revisit 10 years after project impact to capture longer-term changes to the site. These 3 year blocks of time were recommended to allow for the quantification of interannual variability, increasing the likelihood of being able to distinguish project effects from natural variability. However, when negotiating the duration and frequency of environmental effects monitoring programs, resource managers are often faced with proposals suggesting alternate years or multiple skipped years to increase the overall duration while minimizing the costs. For example, it may be suggested that years 1, 3, and 6 post-impact (or post-offsetting) should be monitored to capture a longer duration response but still minimize effort, but in this case no differences would have been found. Various scenarios could be envisioned and a number were tested with our data, which clearly demonstrated that the conclusions from intermittent designs would have been different, and would sometimes have failed to detect an effect that was found using continuous sampling. It is not unusual for long-term datasets to vary in design, with intermittent sampling as a common feature, which can have significant impacts on the ecological measure and change detection ability (Magurran et al. 2010).

When annual data are available, it is important that statistical model results should not be the sole basis for reaching conclusions of impact. A visual examination of the annual trend plots clearly shows why selecting different combinations of post-monitoring years could affect the significance of the statistical result. For example, the flip-flop between the Control and Impact sites for invertebrate diversity emphasize that caution is needed when interpreting the “significant impact” that was found in the long-term continuous design. In 2010, the invertebrate diversity at the Impact site was greater than at the Control site, despite continued unlimited ramping, indicating that invertebrate diversity is likely influenced by factors other than ramping rates, and the statistical significance was still being driven by 2005–2008 data. The violation of the additivity assumption for invertebrate diversity similarly calls any significant impact into question (Stewart-Oaten et al. 1986, 1992). The annual plots also highlight the potential pitfalls of intermittent sampling. For example, monitoring invertebrate abundance during 2006, 2008, and 2010 would have hit high abundance years in the Magpie River and concluded no impact; monitoring 2005, 2007 and 2009 would have hit low abundance years in the Magpie River and would have concluded the ramping change had a significant effect on invertebrate abundance. Thus, long-term monitoring programs that consider 1 year snapshots of data on an intermittent basis may have difficulty arriving at any conclusions, whether statistical or observational.

Invertebrate abundance and diversity were highly variable over the 10 years studied, and no one consistent trend was apparent in either abundance or diversity for the statistical models or range of designs tested. A quantitative review of the literature demonstrated mixed responses of macroinvertebrate abundance to changes in flow magnitude, similarly preventing the development of any robust

statistical relationships (Poff and Zimmerman 2010). The same review demonstrated a consistent negative response of fishes to alteration in flow magnitude (Poff and Zimmerman 2010). So perhaps our inconsistent invertebrate statistics are as much a reflection of the true (and inconsistent) invertebrate community response, and that this rapid-responding, ephemeral ecological group is not the best to use to as an indicator of flow alteration in environmental assessments.

The fish community appears to be a more consistent and potentially reliable indicator of environmental effects. Backpack shocking sampled the smaller size classes and species of fish in the wadable portion of the river. Thus, if response to environmental change at the population level is linked to life history duration, sampling these smaller species suggests that change, if it was to occur, would manifest in 3–5 years, or the maximum age reached by the majority of fishes sampled. Fish community biomass generally did not fluctuate greatly year-to-year, and the patterns between the Magpie and Batchawana were similar, with the greatest fluctuations observed in the Magpie close to the dam. This similarity reinforces the Batchawana as a good covariate for the Magpie in this BACI experiment (Stewart-Oaten and Bence 2001). Our BACI statistics indicate that *after* 3 years a negative impact on fish biomass was detectable; a time period that corresponds to when the majority of species will have completed one generation. The intermittent sampling analysis, however, failed to detect this effect. Fish diversity was never found to be significantly altered by the change to unlimited ramping no matter what combination of years were included in the analyses, but the Control-Impact analysis correctly detected that fish diversity was consistently greater in the Batchawana than the Magpie. If no Before data were available, the implication would have been that the lower diversity was caused by the unlimited ramping regime, highlighting the critical importance of Before data in helping to tease out causation.

Environmental impact assessments conducted in the real world are often limited by temporal and financial constraints, resulting in lack of adequate before data or multiple controls. This paper demonstrates that employing a long-term BACI design remains one of the best ways of detecting an effect from an environmental change, but as cautioned by Stewart-Oaten and Bence (2001), this is true only if statistical analyses are conducted carefully. The use of the BACI design with a minimum of 3 years of before data was suggested as the best scientific framework for assessing the effectiveness of offsetting activities required under the *Fisheries Act* (2012) in Canada (Smokorowski et al. 2015), and this paper further supports that recommendation. In addition, blocks of continuous monitoring after the impact provide more robust conclusions, and it is to the benefit of the project proponent, the regulator, and the aquatic system that poor designs and misleading statistical tests do not lead to flawed decisions. As far as the ramping rate experiment goes, exploring the mechanisms behind the observed biotic responses, if they continue with the inclusion of additional years of monitoring, requires an examination of environmental characteristics such as the flow and temperature regimes, and this is the subject of ongoing analyses. Also still to be considered is if the statistical significance translates into biological relevance for the impacted system (Arciszewski and Munkittrick 2015), but that question is beyond the scope of this paper.

Acknowledgements

We thank Fisheries and Oceans Canada, Ontario Ministry of Natural Resources and Forestry, and Brookfield Renewable for supporting this long-term monitoring program. We also thank Evan Timusk for assisting with data compilation and validation.

Author contributions

Conceived and designed the study: KES, RGR. Performed the experiments/collected the data: KES. Analyzed and interpreted the data: KES, RGR. Contributed resources: KES. Drafted or revised the manuscript: KES, RGR.

Competing interests

The authors have declared that no competing interests exist.

Data accessibility statement

All relevant data are within the paper.

References

- Arciszewski TJ, and Munkittrick KR. 2015. Development of an adaptive monitoring framework for long-term programs: an example using indicators of fish health. *Integrated Environmental Assessment and Management*, 11(4): 701–718. PMID:25781001. doi:10.1002/ieam.1636.
- Benedetti-Cecchi L. 2001. Beyond BACI: optimization of environmental sampling designs through monitoring and simulation. *Ecological Applications*, 11: 783–799. doi:10.1890/1051-0761(2001)011[0783:BBOES]2.0.CO;2.
- Bowman FB, and Somers KM. 2005. Considerations when using the reference condition approach for bioassessment of freshwater ecosystems. *Water Quality Research Journal of Canada*, 40: 347–360.
- Burt TP, Howden NJK, Worrall F, and Whelan MJ. 2008. Importance of long-term monitoring for detecting environmental change: lessons from a lowland river in south east England. *Biogeosciences*, 5: 1529–1535. doi:10.5194/bg-5-1529-2008.
- Chapman MG. 1998. Improving sampling designs for measuring restoration in aquatic habitats. *Journal of Aquatic Ecosystem Stress and Recovery*, 6: 235–251. doi:10.1023/A:1009987403481.
- DFO. 2013. Fisheries productivity investment policy: a proponent's guide to offsetting. Catalogue Number: Fs23-596/2013E-PDF. [online]: Available from <http://www.dfo-mpo.gc.ca/pnw-ppe/offsetting-guide-compensation/offsetting-guide-compensation-eng.pdf>.
- Elliott JM. 1977. Some methods for the statistical analysis of samples of benthic invertebrates. 2nd edition. *Freshwater Biological Association Scientific Publication No. 25*, Ambleside, UK. 159 p.
- Fletcher DJ, and Underwood AJ. 2002. How to cope with negative estimates of components of variance in ecological field studies. *Journal of Experimental Marine Biology and Ecology*, 273: 89–95. doi:10.1016/S0022-0981(02)00142-9.
- Gottelli NJ, and Graves GR. 1996. *Null models in ecology*. Smithsonian Institution Press, Herndon, Virginia. 368 p.
- Green RH. 1979. *Sampling design and statistical methods for environmental biologist*. Wiley Interscience, Chichester, UK.
- Holm S. 1979. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2): 65–70.
- Hurlbert SH. 1971. The nonconcept of species diversity: a critique and alternative parameters. *Ecology*, 52: 577–586. doi:10.2307/1934145.
- Hurlbert SH. 1984. Pseudoreplication and the design of ecological field experiments. *Ecological Monographs*, 54: 187–211. doi:10.2307/1942661.

- Liermann M, and Roni P. 2008. More sites or more years? Optimal study design for monitoring fish response to watershed restoration. *North American Journal of Fisheries Management*, 28(3): 935–943. doi:[10.1577/M06-175.1](https://doi.org/10.1577/M06-175.1).
- Lindenmayer DB, and Likens GE. 2009. Adaptive monitoring: a new paradigm for long-term research and monitoring. *Trends in Ecology & Evolution*, 24(9): 482–486. PMID:[19409648](https://pubmed.ncbi.nlm.nih.gov/19409648/). doi:[10.1016/j.tree.2009.03.005](https://doi.org/10.1016/j.tree.2009.03.005).
- Magurran AE, Baillie SR, Buckland ST, Dick JMCP, Elston DA, Scott EM, et al. 2010. Long-term datasets in biodiversity research and monitoring: assessing change in ecological communities through time. *Trends in Ecology & Evolution*, 25(10): 574–582. PMID:[20656371](https://pubmed.ncbi.nlm.nih.gov/20656371/). doi:[10.1016/j.tree.2010.06.016](https://doi.org/10.1016/j.tree.2010.06.016).
- Mason WT, Weber CI, Lewis PA, and Julian EG. 1973. Factors affecting the performance of basket and multiplate macroinvertebrate samplers. *Freshwater Biology*, 3: 409–436. doi:[10.1111/j.1365-2427.1973.tb00932.x](https://doi.org/10.1111/j.1365-2427.1973.tb00932.x).
- Metcalfe RA, House DA, and Jahncke R. 2001. Waterpower project—physical examination of possible reference study sites for the Magpie river ramping study, Watershed Science Centre Report: WSC.01.1, Ontario Ministry of Natural Resources; 18 p.
- Minns CK. 1995. Allometry of home range sizes in lake and river fishes. *Canadian Journal of Fisheries and Aquatic Sciences*, 52: 1499–1508. doi:[10.1139/f95-144](https://doi.org/10.1139/f95-144).
- Munkittrick KR, McGeachy A, McMaster ME, and Courtenay SC. 2002. Overview of freshwater fish studies from the pulp and paper environmental effects monitoring program. *Water Quality Research Journal of Canada*, 37(1): 49–77.
- Munkittrick KR, Arens CJ, Lowell RB, and Kaminski GP. 2009. A review of potential methods of determining critical effect size for designing environmental monitoring programs. *Environmental Toxicology and Chemistry*, 28(7): 1361–1371. PMID:[19199371](https://pubmed.ncbi.nlm.nih.gov/19199371/). doi:[10.1897/08-376.1](https://doi.org/10.1897/08-376.1).
- Poff NL, and Zimmerman JKH. 2010. Ecological responses to altered flow regimes: a literature review to inform the science and management of environmental flows. *Freshwater Biology*, 55: 194–205. doi:[10.1111/j.1365-2427.2009.02272.x](https://doi.org/10.1111/j.1365-2427.2009.02272.x).
- Robinson DL. 1987. Estimation and use of variance components. *The Statistician*, 36: 3–14. doi:[10.2307/2988267](https://doi.org/10.2307/2988267).
- Roni P, Hanson K, and Beechie T. 2008. Global review of the physical and biological effectiveness of stream habitat rehabilitation techniques. *North American Journal of Fisheries Management*, 28: 836–890. doi:[10.1577/M06-169.1](https://doi.org/10.1577/M06-169.1).
- Shaw DW, and Minshall GW. 1980. Colonization of an introduced substrate by stream macroinvertebrates. *Oikos*, 34: 259–271. doi:[10.2307/3544285](https://doi.org/10.2307/3544285).
- Shaw RG, and Mitchell-Olds T. 1993. ANOVA for unbalanced data: an overview. *Ecology*, 74(6): 1638–1645. doi:[10.2307/1939922](https://doi.org/10.2307/1939922).
- Smith EP, Orvos BW, and Cairns J Jr. 1993. Impact assessment using the Before-After-Control-Impact (BACI) model: concerns and comments. *Canadian Journal of Fisheries and Aquatic Sciences*, 50: 627–637. doi:[10.1139/f93-072](https://doi.org/10.1139/f93-072).

Smokorowski KE, Bradford MJ, Clarke KD, Clément M, Gregory RS, and Randall RG. 2015. Assessing the effectiveness of habitat offset activities in Canada: monitoring design and metrics. Canadian Technical Report of Fisheries and Aquatic Sciences: 3132: vi + 48 p.

Smokorowski KE, Metcalfe RA, Finucan SD, Jones N, Marty J, Power, et al. 2011. Ecosystem level assessment of environmentally-based flow restrictions for maintaining ecosystem integrity: a comparison of a modified peaking versus unaltered river. *Ecohydrology*, 4: 791–806. doi:[10.1002/eco.167](https://doi.org/10.1002/eco.167).

Stewart-Oaten A, and Bence JR. 2001. Temporal and spatial variation in environmental impact assessment. *Ecological Monographs*, 71: 305–339. doi:[10.1890/0012-9615\(2001\)071\[0305:TASVIE\]2.0.CO;2](https://doi.org/10.1890/0012-9615(2001)071[0305:TASVIE]2.0.CO;2).

Stewart-Oaten A, Bence JR, and Osenberg CW. 1992. Assessing effects of unreplicated perturbations: no simple solutions. *Ecology*, 73(4): 1396–1404. doi:[10.2307/1940685](https://doi.org/10.2307/1940685).

Stewart-Oaten A, Murdoch WW, and Parker KR. 1986. Environmental impact assessment: “Pseudoreplication” in time? *Ecology*, 67: 929–940. doi:[10.2307/1939815](https://doi.org/10.2307/1939815).

Underwood AJ. 1991. Beyond BACI: experimental designs for detecting human environmental impacts on temporal variations in natural populations. *Australian Journal of Marine and Freshwater Research*, 42: 569–587. doi:[10.1071/MF9910569](https://doi.org/10.1071/MF9910569).

Underwood AJ. 1992. Beyond BACI: the detection of environmental impacts on populations in the real, but variable, world. *Journal of Experimental Marine Biology and Ecology*, 161: 145–178. doi:[10.1016/0022-0981\(92\)90094-Q](https://doi.org/10.1016/0022-0981(92)90094-Q).

Underwood AJ. 1993. The mechanics of spatially replicated sampling programmes to detect environmental impacts in a variable world. *Australian Journal of Ecology*, 18: 99–116. doi:[10.1111/j.1442-9993.1993.tb00437.x](https://doi.org/10.1111/j.1442-9993.1993.tb00437.x).

Underwood AJ. 1994. On beyond BACI: sampling designs that might reliably detect environmental disturbances. *Ecological Applications*, 4(1): 3–15. doi:[10.2307/1942110](https://doi.org/10.2307/1942110).