RIVER RESEARCH AND APPLICATIONS

River Res. Applic. **27**: 967–975 (2011)

Published online 7 June 2010 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/rra.1415

LEARNING FROM DAM REMOVAL MONITORING: CHALLENGES TO SELECTING EXPERIMENTAL DESIGN AND ESTABLISHING SIGNIFICANCE OF OUTCOMES

K. M. KIBLER,[a*,†] D. D. TULLOS[b‡] and G. M. KONDOLF[c§]

[a] *Water Resources Engineering, Oregon State University, Corvallis, Oregon, USA*
[b] *Biological & Ecological Engineering, Oregon State University, Corvallis, Oregon, USA*
[c] *Environmental Planning and Geography, University of California, Berkeley, California, USA*

ABSTRACT

As the decommissioning of dams becomes a common restoration technique, decisions about dam removals must be based on sound predictions of expected outcomes. Results of past and ongoing dam removal monitoring are an important source of information that practitioners may utilize to evolve predictive and decision-making tools, emphasizing the need for thorough and defensible documentation of dam removal outcomes. However, as dam removals challenge many basic assumptions of conventional experimental designs and data analysis techniques, the quality of information available to aid decision-making may be questionable or misleading.

Nevertheless, some study design principles and analysis procedures may be robust to the challenges presented by dam removal research. To assist managers in undertaking dam removal monitoring, this article discusses the assets and limitations of monitoring and analysis options available for dam removal studies, with emphasis on selecting a rigorous experimental design and determining significance of results.

As the chosen monitoring design will influence the appropriateness of applying standard analytical methods, particularly statistical hypothesis testing, researchers should carefully consider constraints inherent to dam removal studies when designing a monitoring plan and assigning significance to observed changes. Ecological significance is often the most justifiable method for framing significance of dam removal outcomes, though it may be complicated by identification of environmentally significant thresholds. Another alternative is evaluation of the practical significance of results, when observed changes exceed measurement error and background variability. Establishing practical significance may be informative when statistical and ecological significance is inappropriate or impossible to determine.

KEY WORDS: dam removal; experimental design; statistical significance; ecological significance

Received 1 December 2009; Revised 15 April 2010; Accepted 22 April 2010

INTRODUCTION

Ecological restoration is a billion dollar per annum industry in the United States alone (Bernhardt *et al*., 2005), and dam removal is increasingly implemented as a river restoration technique (Hart *et al*., 2002; Heinz Center, 2002; Doyle *et al*., 2003b). To maximize benefits from ecological restoration activities, such as dam removals, it is fundamental that the restoration community sustain a practice of vigilant evaluation with respect to implemented restoration practices. Such assessment will both demonstrate project effectiveness, and for the purpose of continued learning, enhance the success of future restoration efforts. Bernhardt *et al*. (2007) recently reported that, while some degree of monitoring occurs at most restoration sites, monitoring designs tend to lack rigor and often these data are not used to determine success of the restoration projects. Rather, many restoration efforts are evaluated by qualitative monitoring such appearance of the project site or public opinion (Bernhardt *et al*., 2007).

The restoration community has acknowledged the need for more thorough monitoring and analysis of ecological restoration efforts (Kondolf and Micheli, 1995; Michener, 1997; Bernhardt *et al*., 2005; Roni *et al*., 2005; Bernhardt *et al*., 2007), as well as the opportunities for enhanced learning that can be presented by monitoring dam removals (Kibler *et al*., in review). Given the number of dams that are approaching the end of their working lives (Federal Emergency Management Agency, 2009) and the increasing popularity of dam removals (Pohl, 2002), this class of projects would seem ideal for post-project appraisal and learning opportunities. Yet the nature of dam removal poses challenges to application of traditional experimental designs and standard methods of assigning significance to outcomes.

*Correspondence to: K. Kibler, 116 Gilmore Hall, Corvallis, OR 97331-3906, USA. E-mail: kiblerk@engr.orst.edu
†Student
‡Assistant Professor
§Professor

Because incorrect application of experimental design principles and significance testing to environmental monitoring may lead to flawed management and conservation decisions (Underwood, 1994a; Dayton, 1998), restoration practitioners should give specific attention to the design of monitoring studies, choice of data analysis techniques and approaches to verifying practical, ecological and statistical significance of results. To aid managers in monitoring dam removal projects, this article reviews conventional experimental designs applied to environmental research, considering whether and how they can be applied to dam removal monitoring and discusses obstacles to and options for reporting significance of dam removal outcomes.

STUDY DESIGNS TO ENHANCE LEARNING FROM DAM REMOVALS

Although dam removals often push the limits of validity for many accepted experimental designs, some study designs are better suited than others to the constraints inherent to dam removal. For the following summary of dam removal study design options, we distinguish *diachronous* study designs, where observations are collected through time (before and after removal), from *synchronous* designs, in which pre-removal data are unavailable (Piégay and Schumm, 2003). For clarity, we refer readers to Table I, which outlines key characteristics of the study designs discussed below.

Table I. Characteristics of potential experimental designs for dam removal studies

Study design	Minimum spatial requirements	Minimum temporal requirements	Strengths	Limitations
BA	One monitoring site—impact reach	Two years of monitoring, pre- and post-impact	Requires comparatively few resources to implement monitoring	Changes observed cannot be differentiated between dam removal and other external factors
BACI	Two monitoring sites—impact and one control reach	Two years of monitoring, pre- and post-impact	Allows differentiation between changes related to dam removal and changes caused by other factors	Impact and control sites must 'track' consistently through time which is difficult to ensure with only one control site; temporal heterogeneity between impact and control sites may confuse true effects; not robust to serial correlation of time-series data; selection of appropriate, independent control sites can be challenging; more resource-intensive than BA
BACIPS	Two monitoring sites—impact and one control reach	Many years of pre- and post-impact monitoring	Allows differentiation between changes related to dam removal and changes caused by other factors; effect of serial correlation can be evaluated and controlled for	Impact and control sites must 'track' consistently through time which is difficult to ensure with only one control site; temporal heterogeneity between impact and control sites may confuse true effects; selection of appropriate, independent control sites can be challenging; more resource-intensive than BA or BACI without paired sampling
MBACI	Minimum three monitoring sites—one impact and at least two control reaches	Two years of monitoring, pre- and post-impact	Allows differentiation between changes related to dam removal and changes caused by other factors; allow detection of divergence between impact and control sites	More sites may increase parameter variability; selection of appropriate, independent control sites can be challenging; more resource intensive than BA or BACI
Synchronous similarity analysis	Two monitoring sites—impact and one control reach	One year of monitoring	Pre-impact data not required	Differences observed cannot be attributed to dam removal; selection of appropriate, independent control sites can be challenging

DOI: 10.1002/rra

Diachronous study designs

Before-After (BA) study designs are one of the simplest diachronous approaches for dam removal studies and have been widely used to describe channel and ecological changes following restoration. The BA methodology is potentially one of the least resource-intensive study designs. However, while investigators may use this approach to evaluate hypotheses regarding whether a site changes after an impact, one cannot differentiate environmental change related to the dam removal from change caused by external factors (Underwood, 1991; Underwood, 1994b; Roni *et al*., 2005).

By adding a single reference or control site to a BA scenario, investigators may implement a *Before-After-Control-Impact* (BACI) study design to increase confidence in evaluating causality associated with a dam removal. Unlike the BA scenario, BACI designs control for factors that may cause changes to the impacted site but are unrelated to the dam removal. This is accomplished by evaluating the interaction between time and location effects and using variability between within-site sampling as the error term (Stewart-Oaten *et al*., 1986).

A number of authors (Campbell and Stanley, 1966; Eberhardt, 1976; Skalski and McKenzie, 1982; Stewart-Oaten, 1996) have advocated for a BACI design that is based on a time-series of differences, known as the *Before-After-Control- Intervention-Paired Sampling* (BACIPS) design. By adding pre-removal differences between the impacted and control site to post-removal observations, the expected condition absent the impact is estimated. Ecological data gathered through time tends to be serially correlated, where observations collected close to one another in time tend to be similar, which may lead to underestimation of parameter variance (Stewart-Oaten, 1996). BACIPS methodology allows for evaluation of serial correlation, making it applicable to analysis of time-series data following dam removal. However, this approach is data intensive, requiring a substantial number of data years to compare Before-removal differences with After-removal differences. This caveat is a particular challenge to dam removal studies, which often have very little, if any, pre-removal monitoring.

Evaluation of impact through a BACI or BACIPS design is valid only when control and impact sites track consistently through time; Type I errors (an effect is detected when none has occurred) or Type II errors (failure to detect an effect when it does exist) may occur when other sources of variability unrelated to the impact are present (Underwood, 1992; Osenberg and Schmitt, 1996). The propensity for natural temporal variability, or change in the relationship between the impacted and control sites over time, to conceal or augment perceived effect of the impact is a criticism of BACI designs that use only one control site (Underwood, 1992). When a study design includes only one control site, researchers must assume that any differences between the impacted and control sites observed over time are due to dam removal, which may or may not be true. By adding replication in space, the *Multiple-Before-After-Control-Impact* (MBACI) design addresses this issue by sampling multiple control sites (Underwood, 1994a, 1994b). An obvious drawback to the MBACI design is the additional cost and time required to sample multiple control sites. A less apparent disadvantage is that MBACI designs may introduce additional parameter variability. As river conditions are influenced by a range of processes that vary considerably through space, variability of response metrics may increase as number of sites increases, complicating the detection of small effects.

Selection of appropriate control sites may also prove challenging to BACI, BACIPS and MBACI designs applied to river monitoring. Upstream control sites are not independent of the study reaches, or in the case of multiple upstream controls, one another (Hurlbert, 1984; Norris and Hawkins, 2000). For example, downstream reaches are influenced by upstream hydraulic and sediment conditions, while migration of fish or drifting of invertebrates confounds biological independence. Violating the independence, assumption of BACI and MBACI design may lead to invalid conclusions (Osenberg *et al*., 1996; Stewart-Oaten, 1996). Furthermore, as dams are sometimes sited at locations of geologic transition, upstream conditions may not serve as an appropriate control for downstream reaches. One solution to the independence dilemma of river research is to choose control sites within a different river system. However, it may be difficult to find potential monitoring sites on nearby rivers that functionally resemble the experimental reach to the extent that they may serve as adequate controls. In past dam removal studies that employed a BACI or MBACI design, investigators have variously selected one control reach upstream of the reservoir (Stanley *et al*., 2002; Doyle *et al*., 2003a; Cheng and Granata, 2007; Kibler *et al*., in review), multiple upstream control reaches (Casper *et al*., 2006), an entire second intact dam and reservoir system upstream of the removed dam (Ahearn and Dahlgren, 2005) or have monitored reaches on a different river within the same basin (Kanehl *et al*., 1997).

In addition to challenges of establishing control reaches, there is often a great deal of uncertainty regarding the timing of a dam removal, which clashes with the nature of intensively-planned and highly-coordinated scientific study. Frequently, dams are removed before research hypotheses are formulated and before adequate baseline data have been collected. Even the most rigorous studies typically have only a single year of baseline data (Kanehl *et al*., 1997; Stanley *et al*., 2002; Doyle *et al*., 2003a; Casper *et al*., 2006). As conditions within river systems naturally fluctuate over time

DOI: 10.1002/rra

in response to climate, hydrology and stochastic events, one year of data collection represents only one out of many possible conditions. Lacking sufficient baseline data, dam removal studies have little power with which to statistically detect change within the context of natural background variability, and may rely on descriptive or qualitative methods to convey outcomes.

Synchronous similarity analysis

Although a single year of baseline data may fail to sufficiently capture interannual parameter variability, one year of pre-impact data does allow investigators the flexibility to implement a robust experimental design, and thus may be invaluable. Without pre-implementation data, BA, BACI, MBACI and BACIPS designs are not applicable to dam removal studies. Monitoring questions shift from a comparison of Before/After conditions at a site to a synchronous comparison across space to evaluate differences in conditions between impacted and control sites. The synchronous comparison of disturbed and control conditions when pre-impact data are lacking is commonly referred to as a space-for-time substitution (Kondolf, 1997; Piégay and Schumm, 2003) and is limited by a few fundamental caveats. For instance, when using a synchronous design, researchers cannot explicitly attribute differences between impacted and control sites to the dam removal, as opposed to other undocumented factors. As a key objective of dam removal monitoring is to attribute observed changes to a specific action, the dam removal, synchronous studies often will not satisfy the goals of dam removal research.

Another assumption of the synchronous study design is that the control sites do indeed represent conditions that would be typical in impacted sites if not for the presence of the dam and subsequent removal. In contrast to diachronous designs, a synchronous study compares conditions at a disturbed site to those at a control site *directly*. Thus, if the control site is an insufficient analogue for pre-impact conditions at the dam site, the differences detected are questionable. Piégay and Schumm (2003) emphasize the error in using a space-for-time substitution to compare reaches that are influenced by different geomorphic processes. For this reason, studies lacking pre-impact data (synchronous comparisons of control and impacted sites) provide a limited scope of inference for evaluating effects of dam removal.

Implications of study design

Selection of study design and associated analytical approaches for monitoring dam removals determines the types of hypotheses that a study may address. For example, the BA design addresses the question ‘Did this site change over time?’ The investigator may speculate that any observed change occurred in response to the dam removal, but uncertainty is high because any number of factors may have contributed to the change. The BACI design addresses the same question, but enhances the certainty with which the investigator may conclude that change occurred in response to the dam removal; thus the question posited may become ‘Did the dam removal induce change at this site?’ Directly establishing causality in dam removal studies is obviously desirable, but often not possible due to the fact that dam removals are nonreplicated experiments with nonrandomly assigned treatments (Hurlbert, 1984). Thus, as is the case with much environmental research, dam removal studies often file into the category of observational studies (Hacking, 1965; Michener, 1997), where possible inferences do not necessarily include causality.

EVALUATING THE SIGNIFICANCE OF DAM REMOVAL OBSERVATIONS

Statistical significance

Evaluating the significance of observed changes in ecological research is often a challenging task. Significance of environmental change can be defined statistically, through statistical hypothesis testing. However, the standard designation of statistical significance (e.g. setting an α level of 0.05 as the rejection threshold for the null hypothesis) is still subject to investigator’s choice/bias and may be inappropriate or misleading in the context of environmental research (Quinn and Dunham, 1983; McBride *et al*., 1993; Johnson, 1999; Parkhurst, 2001). Sample size has considerable influence on whether or not the null hypothesis is rejected (Green, 1989) and particular vigilance must be exercised when conducting statistical hypothesis testing with low sample sizes. Models lacking statistical power due to low sample size are prone to Type II errors and may fail to reject the null hypothesis even when an effect has occurred (Johnson, 1999). In a similar manner, high parameter variability may also lead to difficulty in creating statistical models with satisfactory statistical power to detect changes (Green, 1989; Di Stefano, 2001). Parameters measured within the dynamic river environment inherently are often characterized by relatively high spatial and temporal variability (Whiting and Dietrich, 1991; Whiting, 1997). From an experimental design perspective, it may be desirable to implement a robust MBACI design by monitoring multiple control sites, or to select control sites in a different river system as to avoid flouting the independence assumptions of BACI designs. However, investigators who wish to apply statistical hypothesis testing to dam removal effects may find themselves conflicted when weighing the benefits of monitoring multiple, diverse sites against the parameter variability that is introduced by doing so.

DOI: 10.1002/rra

Increasing variability of monitored parameters may initiate need for greater sample sizes to attain sufficient statistical power, which may be difficult given the constraints of a dam removal study.

As discussed above, dam removal studies may be plagued by barriers that potentially preclude conscionable application of statistical hypothesis testing including high variability within monitored parameters, and sample sizes that are smaller than optimal. To illustrate the challenges of establishing statistical significance of parameters observed in dam removal, we present two examples from dam removals in Oregon that indicate the length of data record and sample size required to apply statistical hypothesis testing to dam removal monitoring with reasonably powerful models. Using statistical power analysis, we evaluate (1) the number of monitoring years at a given level of sampling required to detect statistically significant changes to mean particle size in riffles downstream of the Brownsville Dam removal (Calapooia River) and (2) the number of adult coho salmon that must be observed following the Savage Rapids Dam removal (Rogue River) for confidence of a population-level effect.

In the first example, data to determine change in downstream riffle grain sizes were collected prior to removal of the Brownsville Dam (Kibler *et al.*, in review). Effect size is presented as the magnitude of mean difference between two samples (mean pre-removal parameter less mean post-removal parameter), normalized by the pooled sample standard deviation, and presented as change. According to this analysis, detection of effect sizes less than 100% change require no less than 5 years of pre-removal data, and small effects (<50% change) are virtually undetectable at this level of sampling (Figure 1).

In an alternate example, we evaluate the minimum effect size needed to demonstrate a statistically significant change in numbers of adult coho salmon following the Savage Rapids dam removal, given population estimates from Rogue River ambient monitoring data. Background variability was established with 25 years of population estimates obtained from seining at Gold Ray Beach near the mouth of the Rogue River (Oregon Department of Fish and Wildlife, 2008). Baseline population estimates were found to vary over an order of magnitude between years, with an average of 10 100 adults counted per year when hatchery and wild stocks were combined. Power analysis to estimate the detectable effect size on population estimates indicates that approximately 18 330 more coho salmon must be observed at the mouth of the Rogue in order to detect a beneficial effect of the dam removal. As this estimate represents 130 to 270% change in the average population of coho salmon, at this level of sampling only large changes in population will be detected as statistically significant, although much smaller effects may be ecologically significant. In light of the high variability of population estimates (coefficient of variation of 0.85 for these estimates), other approaches to monitoring fish responses, such as tracking compositional shifts, migration patterns and DNA (Brenkman *et al.*, 2008; McHenry and Pess, 2008), may be better suited to understanding fisheries impacts associated with dam removals.

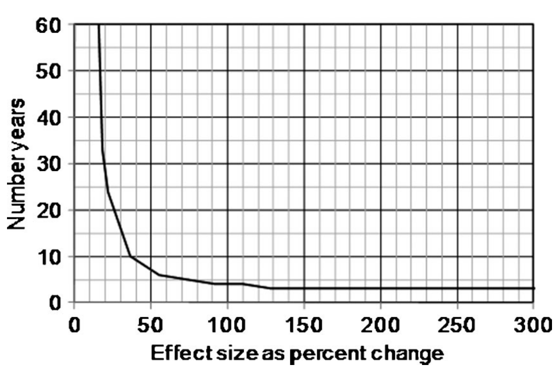

Figure 1. Years of data required to detect statistically significant change of mean particle size (D50) in riffles downstream of the Brownsville Dam removal with power $(1-\beta)$ of at least 0.80. Based on particle size distributions from bulk samples $(n=4)$ collected as part of baseline field reconnaissance

These preliminary analyses illustrate the importance of baseline sampling and power analysis to inform study design (Fairweather, 1991), and demonstrate that considerable resources may be required to achieve reasonable statistical power, particularly when parameters are highly variable. Baseline studies also inform approaches for analysing data, ensuring that appropriate analyses are applied for a given statistical power and that Type II errors are avoided. When employing power analysis to evaluate whether statistical hypothesis testing is suitable for a given data set, researchers should comprehend the implications of thresholds associated with power analysis. The power of a model, denoted by $(1-\beta)$, is the probability that a Type II error will not be made, while the alpha (α) parameter specifies the probability that a Type I error will be made (Green, 1989). Using a model with the standard threshold of sufficient statistical power of $(1-\beta)=0.80$ indicates that the researcher acknowledges a 20% chance of making a Type II error. When compared with the accepted probability of making a Type I error, which is controlled by the α parameter at a standard 5%, some researchers question whether or not $(1-\beta)=0.80$ sets a high enough standard, particularly considering the potential consequences of Type II errors in the context of environmental research (Fairweather, 1991; McBride *et al.*, 1993; Dayton, 1998; Parkhurst, 2001). As environmental research often focuses to detection of dangerous or deleterious effects, concluding that no effect has occurred when, in actuality, an effect has occurred, is likely to be more catastrophic than making a Type I error, as

DOI: 10.1002/rra

the potentially hazardous action will continue, having been confirmed statistically benign (Dayton, 1998; Di Stefano, 2001). In this way, Type II errors in environmental analysis may propagate misinformed management decisions and incorrect conceptual models (Fairweather, 1991).

In addition to potential deficiencies in statistical power, dam removal studies also often violate basic assumptions inherent to robust experimental designs, such as BACI, which may invalidate the option of statistical hypothesis testing. As discussed above, independence of samples through space and time is a key assumption of BACI designs that may be violated by the longitudinal nature of the river continuum (Hurlbert, 1984; Norris and Hawkins, 2000). If control and impact sites are located within the same river system, they may not be independent of one another. Violation of the independence assumption confounds many standard parametric tests (Reckhow *et al*., 1990), making statistical hypothesis testing a questionable practice in the context of river research.

Ecological significance

One promising alternative to statistical hypothesis testing is for environmental researchers to define significance of results according to system thresholds and ecologically relevant change (Perry, 1986; Yoccoz, 1991). Defining significance as the potential for ecological change may be more applicable to ecosystem management than a definition based on statistical significance (Jones and Mattloff, 1986; Richter *et al*., 1996). Often ecosystem response to perturbation displays nonlinear behaviour, responding to thresholds such that very small increments of change in a given parameter may initiate ecosystem transformation to an alternative stable state (Scheffer *et al*., 2001). In this sense, the concept of the environmental threshold, or nonlinear response to perturbation, is of interest to ecosystem managers, though triggers of nonlinear response are not necessarily defined by statistical significance. Thus, ideally, researchers may compare environmental change after dam removal to known system thresholds to determine the ecological significance of observed change.

In response to the argument that statistical significance provides no indication of ecological significance (Perry, 1986; Yoccoz, 1991), some authors have advocated for the use of equivalence tests over the traditional hypothesis test of significant difference (Patal and Gupta, 1984; McBride *et al*., 1993). The equivalence test requires the investigator to select a threshold value for an ecologically significant difference before testing the null hypothesis that the observed change exceeds this threshold value. This differs from traditional hypothesis testing in two ways. First, equivalence testing requires that the investigator define a threshold for ecologically significant change prior to analysis, basing determination of significance upon how the effect size compares to an environmentally-relevant value. Secondly, the null hypothesis of the equivalence test posits that change has occurred and exceeds the chosen threshold, while the null hypothesis of traditional hypothesis testing states that no significant change has occurred (Parkhurst, 2001). As the burden of proof lies in demonstrating that no significant change has occurred, the null hypothesis given by the equivalence test provides a more conservative analysis for environmental research. As Type II errors are now controlled by the α parameter (typically set at $\alpha = 0.05$), equivalence testing underscores the precautionary principle of avoiding environmental impact (McBride, 1999; Parkhurst, 2001). At the same time, statistically significant change given by statistical hypothesis testing could be classified as nonsignificant by equivalence tests if change falls below the threshold of ecological significance (McBride *et al*., 1993). Thus, the equivalence test may be more valid for targeted environmental management than the traditional hypothesis test.

Sensible as the concept of ecological significance may be, implementation is not without challenges. The greatest challenge to the approach of equivalence testing lies in selection of ecologically significant threshold values (Groffman *et al*., 2006). Ecological thresholds may or may not exist for a given parameter, and thresholds may be ecosystem-specific and highly variable through space (Scheffer *et al*., 2001). Thus, identifying thresholds that can be used in the context of decision-making may be a time- and resource-intensive endeavour. However, some parameters of interest to dam removal studies have been thoroughly researched and some thresholds may be inferred from scientific literature. For example, in the assessment of impacts associated with downstream sediment deposition following dam removal in the Pacific Northwest, one might evaluate how observed changes in grain size relate to those of preference for spawning salmon, for which a well-developed literature exists (Sowden and Power, 1985; Reiser and White, 1988; Groot and Margolis, 1991; Kondolf and Wolman, 1993 and Hard *et al*., 1996). Similarly, threshold values have been established in some areas relevant to dam removal, including stream temperature (see Brett, 1952; Brett, 1956; Sullivan *et al*., 2000 for salmonid temperature thresholds, Sweeney and Schnack, 1977; Sweeney, 1978; Nordlie and Arthur, 1981; Quinn *et al*., 1994 for temperature thresholds of macroinvertebrates), and nutrient triggers for algae blooms (Dodds *et al*., 1998; Roelke, 2000 and Gunderson and Holling, 2002). Although ecological significance may be difficult to infer quantitatively in some situations, when it is possible to frame observed change within the context of ecologically significant change, this may be a superior method of assigning significance as compared to statistical significance.

Practical significance

Although the commonly accepted definition of statistical significance may be inappropriate for some studies (McBride *et al*., 1993), and even when ecologically significant thresholds of some parameters and locations can be uncertain, well-designed monitoring studies may still draw defensible conclusions regarding dam removal outcomes. In lieu of reporting *p*-values and other common metrics associated with traditional statistical hypothesis testing, investigators may infer the practical significance of observed changes. We define the concept of practical significance to include estimates of observed changes that exceed both parameter estimate uncertainty and extremes of natural background variability of that parameter. Reporting statistics, such as estimates of effect and confidence intervals, placed within perspective of uncertainty and parameter variability, may be a justifiable means for defining significance when traditional hypothesis testing is unsuitable and the ecological context for observed changes is unknown.

Determination of practical significance requires that all sources of uncertainty in parameter definition be acknowledged and quantified. Methods of field data collection can often introduce several sources of uncertainty that can obscure the assessment of significance. Each should be considered and estimated independently before being aggregated to a single uncertainty measurement for the method. Standard practices exist for the mathematical treatment of measurement error (Taylor, 1997) and may require collection of additional field data such as repeat sampling. For example, Walter and Tullos (2009) estimated uncertainty associated with river bathymetry surveys by gathering repeat samples and quantifying the error between the replicates. Alternatively, estimates of error associated with particular methods may be published. For example, the Oregon Department of Fish and Wildlife (ODFW) builds an 11% repeat survey into its Aquatic Habitat Inventory and reports measurement error in the form of a signal to noise ratio (Anlauf and Jones, 2007). The Environmental Monitoring and Assessment Program (EMAP), performed by the US Environmental Protection Agency (EPA), also reports measurement error associated with their monitoring techniques (Kaufmann *et al*., 1999).

The second component to establishing the practical significance of observed change is to report observed effects within the context parameter variability. This is distinct from reporting measurement uncertainty in that it requires that the researcher also provides information regarding the extremes of natural expression of the parameter in question. This may be accomplished by reporting data gathered at one or multiple control sites over the same time period. Alternatively, comparisons could include observed changes versus the range of variability observed at the impacted site for a period of time prior to the disturbance event. Practical significance may be attributed to observed changes that, after considering measurement error, exceed change that occurred in control sites or prior to removal.

SALIENCE OF DAM REMOVAL RESEARCH

Dam removal has tremendous potential for restoring the connectivity of natural flows, hydraulic characteristics and isolated or impaired fisheries of fragmented river ecosystems. The true number of dams impounding rivers in the United States is largely unknown. The National Inventory of Dams lists approximately 83 000 dams over 1.8 m in height (FEMA, 2009), but the number of dams unaccounted for by federal or state inventories may be much greater (Aspen Institute, 2002). Many of these dams no longer serve their intended purpose, or have exceeded their working life, and some are considered public safety hazards (Doyle *et al*., 2003b). Dams that no longer serve a viable function continue to degrade aquatic ecosystems by blocking migration routes and interrupting flows of water and sediment through the watershed. Dam removal is increasingly the preferred option when faced with expensive retrofitting to meet regulations (such as the Endangered Species Act) or to satisfy dam safety standards (Doyle *et al*., 2003b). Although approximately 680 dams have been removed in the last 100 years (American Rivers, Friends of the Earth, Trout Unlimited, 1999; Gleick *et al*., 2009), outcomes from less than 5% of these removals have been published in scientific literature (Hart *et al*., 2002). Thus, uncertainty regarding consequences of dam removal is high (Aspen Institute, 2002), particularly regarding the magnitude, timing and spatial extent of physical and ecological outcomes (Hart *et al*., 2002; Heinz Center, 2002; Stanley and Doyle, 2003). To this end, research into the outcomes of dam removal is a particularly salient research topic with vast learning potential regarding ecosystem disturbance and recovery, restoration science and geomorphic processes. Given the potentially high impact of dam removal research, it is vital that investigators are able to confidently report results from rigorously implemented monitoring studies.

As the tradeoffs between the disturbance and benefits of dam removal must be considered (Stanley and Doyle, 2003), it is also essential that practitioners and decision makers have tools at their disposal for predicting potential positive and negative outcomes. To inform future work, researchers ideally will utilize results from past studies to develop and refine current conceptual and predictive models of dam removal outcomes, creating an adaptive management mechanism for dam removal practice and research. Developing educational opportunities from past and ongoing dam removals will ultimately propagate better

DOI: 10.1002/rra

decision-making with regard to future dam removals. However, careful consideration of the experimental design of dam removal studies and reporting of significance of results is needed to ensure that conceptual and predictive decision-making tools are based on reliable information.

ACKNOWLEDGEMENTS

This research was supported by grants from the Oregon Water Enhancement Board (OWEB), project number 207-091, and from the National Oceanic and Atmospheric Association (NOAA), National Marine Fisheries Service, award number NA07NMF4630201. We would like to thank Dan VanDyke from the ODFW for supplying Rogue River coho salmon population data.

REFERENCES

Ahearn DS, Dahlgren RA. 2005. Sediment and nutrient dynamics following a low-head dam removal at Murphy Creek, California. *Limnology and Oceanography* **50**(6): 1752–1762.

American Rivers, Friends of the Earth, Trout Unlimited. 1999. *Dam Removal Success Stories: Restoring Rivers Through Selective Removal of Dams that Don't Make Sense*. American Rivers, Friends of the Earth, and Trout Unlimited: Washington DC.

Anlauf KJ, Jones KK. 2007. Stream Habitat Conditions in Western Oregon, 2005. *Monitoring Program Report Number OPSW-ODFW- 2007-05*. Oregon Department of Fish and Wildlife, Salem.

Aspen Institute. 2002. *Dam Removal: A New Option for a New Century.* Aspen Institute: Queenstown.

Bernhardt ES, Palmer MA, Allan JD, Alexander G, Barnas K, Brooks S. 2005. Synthesizing US river restoration efforts. *Science* **308**: 636–637.

Bernhardt ES, Sudduth EB, Palmer MA, Allan JD, Meyer JL, Alexander G, Follastad-Shah J, Hassett B, Jenkinson R, Lave R, Rumps J, Pagano L. 2007. Restoring rivers one reach at a time: results from a survey of U.S. river restoration practitioners. *Restoration Ecology* **15**(3): 482–493.

Brenkman SJ, Pess GR, Torgersen CE, Kloehn KK, Duda JJ, Corbett SC. 2008. Predicting recolonization patterns and interactions between potamodromous and anadromous salmonids in response to dam removal in the Elwha River, Washington State, USA. *Northwest Science* **82** (Special Issue): 91–106.

Brett JR. 1952. Temperature tolerance in young Pacific salmon, genus *Oncorhynchus*. *Journal of the Fisheries Resources Board of Canada* **9**: 265–323.

Brett JR. 1956. Some principles in the thermal requirements of fishes. *The Quarterly Review of Biology* **31**(2): 75–87.

Campbell DT, Stanley JC. 1966. *Experimental and Quasi-experimental Designs for Research.* Rand McNally: Chicago.

Casper AF, Thorp JH, Davies SP, Courtemanche DL. 2006. Ecological responses of zoobenthos on the Kennebec River, Maine, USA. *Archiv Fur Hydrobiologie (Large Rivers Supplement 115)* **16**(4): 541–555.

Cheng F, Granata T. 2007. Sediment transport and channel adjustments associated with dam removal: field observations. *Water Resources Research* **43**: 1–14.

Dayton PK. 1998. Reversal of the burden of proof in fisheries management. *Science* **279**(5352): 821–822.

Di Stefano J. 2001. Power analysis and sustainable forest management. *Forest Ecology and Management* **154**: 141–153.

Dodds WK, Jones JR, Welch EB. 1998. Suggested classification of stream trophic state: distributions of temperate stream types by chlorophyll, total nitrogen, and phosphorus. *Water Resources* **32**(5): 1455–1462.

Doyle MW, Stanley EH, Harbor JM. 2003b. Channel adjustments following two dam removals in Wisconsin. *Water Resources Research* **39**(1): 1–15.

Doyle MW, Stanley EH, Harbor JM, Grant GS. 2003a. Dam removal in the United States: emerging needs for science and policy. *Eos* **84**(29): 32–33.

Eberhardt LL. 1976. Quantitative ecology and impact assessment. *Journal of Environmental Management* **4**: 27–70.

Fairweather PG. 1991. Statistical power and design requirements for environmental monitoring. *Australian Journal of Marine Freshwater Research* **42**: 555–567.

Federal Emergency Management Agency. 2009. Dam safety in the United States: a progress report on the National Dam Safety Program. Fiscal Years 2006 and 2007. FEMA P-759.

Gleick PH, Cooley H, Cohen MJ, Marikawa M, Morrison J, Palanappn M. 2009. Dams removed or decommissioned in the United States, 1912 to present. In *The World's Water 2008–2009*, Pacific Institute for Studies in Development, Environment, and Security. Island Press: Washington DC.

Green RH. 1989. Power analysis and practical strategies for environmental monitoring. *Environmental Research* **50**: 195–205.

Groffman PM, Baron JS, Blett T, Gold AJ, Goodman I, Gunderson LH, Levinson BM, Palmer MA, Paerl HW, Peterson GD, Poff NL, Rejeski DW, Reynolds JF, Turner MG, Weathers KC, Wiens J. 2006. Ecological thresholds: the key to successful environmental management or an important concept with no practical application? *Ecosystems* **9**: 1–13. DOI: 10.1007/s10021-003-0142-z

Groot C, Margolis L. 1991. *Pacific Salmon Life Histories*. University of British Columbia Press: Vancouver.

Gunderson L, Holling C. 2002. *Panarchy: Understanding Transformations in Human and Natural Systems*. Island Press: Washington, DC.

Hacking I. 1965. *The logic of statistical inference*. Cambridge University Press: Cambridge.

Hard JJ, Kope RG, Grant WS, Waknitz WF, Parker LT, Waples RS. 1996. Status review of pink salmon from Washington. *NOAA Technical Memorandum NMFS-NWFSC-25*, US Department of Commerce, Seattle.

Hart DD, Johnson TE, Bushaw-Newton KL, Horwitz RJ, Bednarek AT, Charles DF, Kreeger DA, Velinsky DA. 2002. Dam removal: challenges and opportunities forecological research and river restoration. *Bioscience* **52**(8): 669–681.

Heinz Center. 2002. *Dam Removal Research: Status and Prospects*. Heinz Center for Science, Economics, and the Environment: Washington DC.

Hurlbert SH. 1984. Pseudoreplication and the design of ecological field experiments. *Ecological Monographs* **54**(2): 187–211.

Johnson DH. 1999. The insignificance of statistical significance testing. *Journal of Wildlife Management* **63**(3): 763–772.

Jones D, Mattloff N. 1986. Statistical hypothesis testing in biology: a contradiction in terms. *Journal of Economic Entomology* **79**(5): 1156–1160.

Kanehl PD, Lyons J, Nelson JE. 1997. Change in the habitat and fish community of the Milwaukee River, Wiconsin, following removal of the Woolen Mills Dam. *North American Journal of Fisheries Management* **17**: 387–400.

Kaufmann PR, Levine P, Robison EG, Seeliger C, Peck DV. 1999. *Quantifying Physical Habitat in Wadeable Streams*, EPA/620/R-99/003. US Environmental Protection Agency: Washington, DC.

Kibler KM, Tullos DD, Kondolf GM. in review. Evolving conceptual models of dam removal: systematic monitoring of the removal of a small, gravel-filled dam. *Journal of the American Water Resources Association.*

Kondolf GM. 1997. Application of the pebble count: notes on purpose, method, and variants. *Journal of the American Water Resources Association* **33**(1): 79–87.

Kondolf GM, Micheli EL. 1995. Evaluating stream restoration projects. *Environmental Management* **19**(1): 1–15.

Kondolf GM, Wolman MG. 1993. The sizes of salmonid spawning gravels. *Water Resources Research* **29**: 2275–2285.

McBride GB. 1999. Equivalence tests can enhance environmental science and management. *Australia and New Zealand Journal of Statistics* **41**(1): 19–29.

McBride GB, Loftis JC, Adkins NC. 1993. What do significance tests really tell us about the environment? *Environmental Management* **17**(4): 423–432.

McHenry ML, Pess GR. 2008. An overview of monitoring options for assessing the response of salmonids and their aquatic ecosystems in the Elwha River following dam removal. *Northwest Science* **82** (Special Issue): 29–47.

Michener WA. 1997. Quantitatively evaluating restoration experiments: research design, statistical analysis, and data management considerations. *Restoration Ecology* **5**(4): 324–337.

Nordlie KJ, Arthur JW. 1981. Effect of elevated water temperature on insect emergence in outdoor experimental channels. *Environmental Pollution (Series A)* **25**: 53–65.

Norris RH, Hawkins CP. 2000. Monitoring river health. *Hydrobiologia* **435**(3): 5–17.

Oregon Department of Fish and Wildlife. (ODFW). 2008. Rogue River Population Estimates: Gold Ray Beach. Data provided by Dan Van Dyke, Rogue District Fish Biologist, Central Point, Oregon.

Osenberg CW, Schmitt RJ. 1996. Detecting ecological impacts caused by human activities. In *Detecting Ecological Impacts: Concepts and Applications in Coastal Habitats*, Schmitt RJ, Osenberg CW (eds). Academic Press: London.

Osenberg CW, Schmitt RJ, Holbrook SJ, Abu-Saba KE, Flegal AR. 1996. Detection of environmental impacts: natural variability, effect size, and power analysis. In *Detecting Ecological Impacts: Concepts and Applications in Coastal Habitats*, Schmitt RJ, Osenberg CW (eds). Academic Press: London.

Parkhurst DE. 2001. Statistical significance tests: equivalence and reverse tests should reduce misinterpretation. *Bioscience* **51**(12): 1051–1057.

Patal HI, Gupta GD. 1984. A problem of equivalence in clinical trials. *Biomedical Journal* **26**(5): 471–474.

Perry JN. 1986. Multiple-comparison procedures: a dissenting view. *Journal of Economic Entomology* **79**: 1149–1155.

Piégay H, Schumm SA. 2003. System approaches in fluvial geomorphology. In *Tools in Fluvial Geomorphology*, Kondolf GM, Piégay H (eds). John Wiley & Sons: West Sussex.

Pohl MM. 2002. Bringing down our dams: trends in American dam removal rationales. *Journal of the American Water Resources Association* **38**(6): 1511–1519.

Quinn JF, Dunham AE. 1983. On hypothesis testing in ecology and evolution. *The American Naturalist* **122**(5): 602–617.

Quinn JM, Steele GL, Hickey CW, Vickers ML. 1994. Upper thermal tolerances of twelve New Zealand stream invertebrate species. *New Zealand Journal of Marine and Freshwater Research* **28**: 391–397.

Reckhow KH, Clements JH, Dodd RC. 1990. Statistical evaluation of mechanistic water-quality models. *Journal of Environmental Engineering* **116**(2): 250–268.

Reiser DW, White RG. 1988. Effects of two sediment size-classes on survival of steelhead and chinook salmon eggs. *North American Journal of Fisheries Management* **8**: 432–437.

Richter BD, Baumgartner PV, Powell J, Braun DP. 1996. A method for assessing hydrologic alteration in ecosystems. *Conservation Biology* **10**(4): 1163–1174.

Roelke D. 2000. Copepod food-quality threshold as a mechanism influencing phytoplankton succession and accumulation of biomass, and secondary productivity: a modeling study with management implications. *Ecological Modelling* **134**: 245–274.

Roni P, Liermann MC, Jordan C, Steel EA. 2005. Steps for designing a monitoring and evaluation program for aquatic restoration. In *Monitoring Stream and Watershed Restoration*, Roni P (ed.). American Fisheries Society: Bethesda.

Scheffer M, Carpenter S, Foley JA, Folkes C, Walker B. 2001. Catastrophic shifts in ecosystems. *Nature* **413**: 591–596.

Skalski JR, McKenzie DH. 1982. A design for aquatic monitoring systems. *Journal of Environmental Management* **14**: 237–251.

Sowden TK, Power G. 1985. Prediction of rainbow trout embryo survival in relation to groundwater seepage and particle size of spawning substrates. *Transactions of the American Fisheries Society* **114**: 804–812.

Stanley EH, Doyle MW. 2003. Trading off: the ecological effects of dam removal. *Frontiers in Ecology and the Environment* **1**(1): 15–22.

Stanley EH, Leubke MA, Doyle MW, Marshall DW. 2002. Short-term changes in channel form and macroinvertebrate communities following low-head dam removal. *Journal of the North American Benthological Society* **21**(1): 172–187.

Stewart-Oaten A. 1996. Problems in the analysis of environmental monitoring data. In *Detecting Ecological Impacts: Concepts and Applications in Coastal Habitats*, Schmitt RJ, Osenberg CW (eds). Academic Press: London.

Stewart-Oaten A, Murdoch WW, Parker KR. 1986. Environmental impact assessment: 'Psuedoreplicaton' in time? *Ecology* **67**: 929–940.

Sullivan K, Martin DJ, Cardwell RD, Toll JE, Duke S. 2000. *An Analysis of the Effects of Temperature on Salmonids of the Pacific Northwest with Implications for Selecting Temperature Criteria*. Sustainable Ecosystems Institute: Portland.

Sweeney BW. 1978. Bioenergentic and developmental response of a mayfly to thermal variation. *Limnology and Oceanography* **23**(3): 461–477.

Sweeney BW, Schnack JA. 1977. Egg development, growth and metabolism of *Sigara alternata* (Say) (Hemiptera:Corixidae) in fluctuating thermal environments. *Ecology* **58**: 265–277.

Taylor JR. 1997. *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*. University Science Books: Sausalito.

Underwood AJ. 1991. Beyond BACI: experimental design for detecting human environmental impacts on temporal variations in natural populations. *Australian Journal of Marine and Freshwater Research* **42**: 569–587.

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.

Underwood AJ. 1994a. Spatial and temporal problems with monitoring. In *Rivers Handbook* (vol. 2), Calow P, Petts GE (eds). Blackwell: Oxford.

Underwood AJ. 1994b. On beyond BACI: sampling designs that might reliably detect environmental disturbances. *Ecological Applications* **4**(1): 3–15.

Walter C, Tullos DD. 2009. Downstream channel changes after a small dam removal: using aerial photos and measurement error for context; Calapooia River, Oregon. *River Research and Applications* DOI: 10.1002/rra.1323

Whiting PJ. 1997. The effect of stage on flow and components of the local force balance. *Earth Surface Processes and Landforms* **22**: 517–530.

Whiting PJ, Dietrich WE. 1991. Convective accelerations and boundary shear stress over a channel bar. *Water Resources Research* **27**: 783–796.

Yoccoz NG. 1991. Use, overuse, and misuse of significance tests in evolutionary biology and ecology. *Bulletin of the Ecological Society of America* **72**(2): 106–111.