CrossMark

# Evaluating impacts using a BACI design, ratios, and a Bayesian approach with a focus on restoration

**Mary M. Conner · W. Carl Saunders ·
Nicolaas Bouwes · Chris Jordan**

**Abstract** Before-after-control-impact (BACI) designs
are an effective method to evaluate natural and human-
induced perturbations on ecological variables when
treatment sites cannot be randomly chosen. While effect
sizes of interest can be tested with frequentist methods,
using Bayesian Markov chain Monte Carlo (MCMC)
sampling methods, probabilities of effect sizes, such as a
≥20 % increase in density after restoration, can be
directly estimated. Although BACI and Bayesian
methods are used widely for assessing natural and
human-induced impacts for field experiments, the ap-
plication of hierarchal Bayesian modeling with MCMC
sampling to BACI designs is less common. Here, we
combine these approaches and extend the typical
presentation of results with an easy to interpret ratio,
which provides an answer to the main study question—
"How much impact did a management action or natural
perturbation have?" As an example of this approach, we
evaluate the impact of a restoration project, which im-
plemented beaver dam analogs, on survival and density
of juvenile steelhead. Results indicated the probabilities
of a ≥30 % increase were high for survival and density
after the dams were installed, 0.88 and 0.99, respective-
ly, while probabilities for a higher increase of ≥50 %
were variable, 0.17 and 0.82, respectively. This ap-
proach demonstrates a useful extension of Bayesian
methods that can easily be generalized to other study
designs from simple (e.g., single factor ANOVA, paired
$t$ test) to more complicated block designs (e.g., cross-
over, split-plot). This approach is valuable for estimat-
ing the probabilities of restoration impacts or other
management actions.

M. M. Conner (✉)
Department of Wildland Resources, Utah State University, 5230
Old Main Hill, Logan, UT 84322, USA
e-mail: mary.conner@usu.edu

W. C. Saunders · N. Bouwes
Department of Watershed Sciences, Utah State University, 5210
Old Main Hill, Logan, UT 84322, USA

W. C. Saunders · N. Bouwes
Eco Logical Research, Inc., Box 706, Providence, UT 84332,
USA

C. Jordan
NOAA Fisheries, Northwest Fisheries Science Center,
Mathematical Ecology and Systems Monitoring Program, 2725
Montlake Blvd E, Seattle, WA 98112, USA

## Introduction

A common approach to evaluate the impacts of natural
or human-induced perturbations on ecosystems where
the allocation of treatment and control sites cannot be
assigned randomly is a before-after-control-impact/
treatment (BACI) design (Eberhardt 1976; Green
1979). A variety of BACI designs have been proposed

to draw inferences about impacts (e.g., BACIPS, MBACI, and beyond-BACI, following the nomenclature of Downes et al. 2002). A primary example and impetus for development of the method was evaluation of the impacts of a nuclear power plant on many ecological response variables, from zooplankton abundance (Mathur et al. 1980; Bence et al. 1996) to communities of macroinvertebrates and related physical variables (Schroeter et al. 1993). BACI designs continue to be used to evaluate impacts from natural perturbations (Russell et al. 2015) and management actions (Desrosiers et al. 2006; Louhi et al. 2010; Hanisch et al. 2013), as well as for a wide variety of smaller-scale field experiments, including evaluating restoration actions (Rumbold et al. 2001; Muotka and Syrjänen 2007; Bousquin and Colee 2014). Similar to studies of larger-scale impacts, efficiently evaluating restoration activities is complicated because, in many cases, restoration actions cannot be implemented in randomly selected locations owing to factors such as access requirements and land ownership, and replication is often restricted due to limited numbers of potential restoration sites, cost of restoration, and logistical constraints.

Analysis of BACI designs has conventionally involved the use of general linear models (e.g., analysis of variance, see Downes et al. 2002) or the use of intervention analyses (Carpenter et al. 1989; Stewart-Oaten and Bence 2001). A particularly useful modification is where impacted and control sites are treated as fixed effects and sampling is conducted at simultaneous (paired) time periods in treatment and control sites before and after perturbation (BACIPS; Stewart-Oaten et al. 1986; Underwood 1994). Treatment effects for BACIPS designs are often estimated as the mean difference between treatment and control sites after the treatment minus the mean difference between treatment and control sites before the treatment $(d_{\text{treat-control after}} - d_{\text{treat-control before}})$ (Stewart-Oaten et al. 1986; Bence et al. 1996), or via a treatment (control–treatment) × time (before–after) interaction term (Russell et al. 2009; Popescu et al. 2012). This design allows treatment impacts to be distinguished from background time effects shared by all sites, as well as from background differences between treatment and control sites (Popescu et al. 2012). In essence, this design controls for spatial differences between treatment and control sites such that they do not have to be identical. Because of its applicability to restoration field experiments, here we focus on the BACIPS design.

Although the design works well for testing for perturbation effects in field experiments, the results (i.e., treatment minus control difference or significant interaction term) from BACIPS designs analyzed using frequentist statistical approaches typically lack meaningful probabilistic interpretation and are thus not easily understood by nonscientific audiences (Eberhardt 1976; Crome et al. 1996). There is a continuum of interpretability of frequentist results, with $P$ values being perhaps the least understandable to a lay audience and effect sizes and their confidence intervals being more accessible. However, the interpretation of confidence interval is also not intuitive: the interval the unknown true mean change would fall between at the frequency of the confidence level if the experiment were repeated. Bayesian approaches have advantages for interpretation. Because the Bayesian approach is explicitly conditioned on the observed data, Bayesian inference provides direct probability assessments of the response parameter that are more straightforward to interpret (e.g., probability of a % increase or decrease in population size) (Crome et al. 1996; Wade 2000). Moreover, the Bayesian approach has the flexibility to use posterior distributions to estimate a variety of comparisons (Wade 2000) and to report the probability of observing a range of effects sizes (Gelman et al. 2004; Kery 2010; King et al. 2010). In addition, by conditioning on the data, not a specific hypothesis, and providing inference about a range of effect sizes, a Bayesian approach reduces the potential for type I and type II errors, a long-standing criticism of the analysis of BACI data (Mapstone 1995; Murtaugh 2002). If study results, particularly contentious ones, can be conveyed in a manner that is accessible, yet accurate, to both scientific and lay audiences alike, they are far more likely to be embraced by resource managers (Crome et al. 1996).

Here, we present a method with Bayesian interpretability to evaluate responses of treatment sites to natural perturbations or management actions via an adaptable proportional response variable combined with a Bayesian hierarchical model and Markov chain Monte Carlo (MCMC) sampling to estimate the probability of observing different effect sizes. To demonstrate these techniques and highlight their usefulness for evaluating restoration actions, we use a dataset from a BACIPS field experiment combined with a Bayesian MCMC approach to evaluate the effectiveness of a river restoration project to increase juvenile steelhead (*Oncorhynchus mykiss*) survival and density. While we use a

particular study design here (BACIPS), this approach can be readily adapted for a wide variety of statistical study designs, from simple (e.g., single factor ANOVA, paired $t$ test) to more complicated block designs (e.g., crossover, split-plot).

## Materials and methods

Study area and field sampling

The data used to demonstrate this analysis method were collected in Bridge Creek and Murderers Creek, tributaries to the John Day River and part of the larger Columbia River Basin (Fig. 1). The John Day River is occupied by federally threatened steelhead that spawn in both Bridge and Murderers creeks. After emergence, these tributaries provide rearing habitat for juvenile steelhead (anadromous life history of *O. mykiss*) as well as rainbow trout (resident life history of *O. mykiss*). Owing to historical land use practices, extensive portions of Bridge Creek have undergone substantial down-cutting, resulting in a narrow incised straightened channel that lacks habitat complexity necessary to support robust *O. mykiss* populations. In an attempt to aggrade the channel by capturing fine sediments and ultimately increase channel complexity, beaver dam analogs (BDAs) spanning tributary channels were constructed within four treatment reaches on Bridge Creek (Pollock et al. 2014; Bouwes et al. 2016). This restoration strategy assumed that BDAs and subsequent colonization by resident beaver would increase both the total surface area available to juvenile *O. mykiss* as well as increase habitat complexity available for juvenile fish (Bouwes et al. 2016). The BDAs were installed during December 2009. The control watershed, Murderers Creek, was chosen because it is a stream of similar size, discharge, and gradient and resides in the same biome as the treatment watershed (Bridge Creek).

Mark-reencounter sampling was conducted from January 2007 through September 2012 to estimate seasonal survival and density of juvenile steelhead. More complete descriptions of the study area and sampling methods are provided in Tattam et al. (2013) and
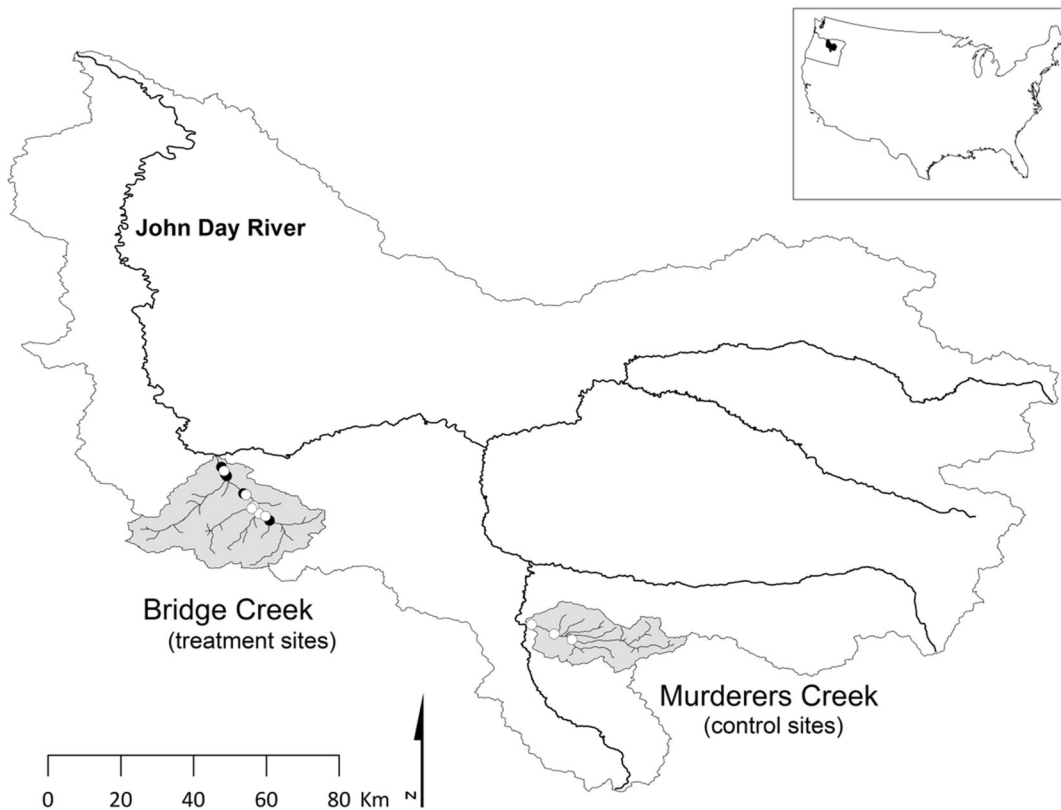


**Fig. 1** Study sites in Bridge and Murderers creeks, located in the John Day River Basin, OR, USA. *Open circles* indicate sites used for both analysis of abundance and survival data, while *filled circles* indicate additional sites used only for analysis of survival data

Pollock et al. (2012). Juvenile steelhead were captured by electroshocking at permanent sites that ranged from 500 to 1000 m long. All steelhead greater than 60 mm were tagged with passive integrated transponders (PITs) and released at the site of capture. In the Bridge Creek, the treatment watershed, 13 sites (four in the treatment reaches and nine in nontreated reaches) were sampled throughout watershed over the study period. Because fish can move between sites, we considered all sites in the watershed to be treatment sites. In the lower portion of Bridge Creek, in four of the sites, an insufficient number of fish could be tagged to obtain accurate density estimates, and thus were not included in the density analyses. In Murderers Creek, which served as the control watershed, three sites were sampled in its lower portion (Fig. 1). Sites were sampled on two consecutive days (closed-capture sessions), and each site was revisited during three seasons, generally representing summer (June), fall (September), and winter (December–January). The entire sampling period within a season was relatively short, averaging 1 to 2 weeks (in order to sample all sites), with each site having approximately the same period of time between closed-capture sessions (although period length varied from season to season). This yielded three biologically relevant seasons for survival rates—summer (June–September), fall

(October–December), and winter/spring (January–May; Table 1)—and three estimates of population abundance for each year (Table 1).

## Statistical methods

The overall goal of this study is to demonstrate the use of a Bayesian approach to estimate the probability of observing different restoration treatment effect sizes for a BACIPS study for parameters of different scales (i.e., one constrained 0–1 and the other not). To do this, we generated the best estimates of juvenile steelhead survival and density before and after the restoration action was implemented on treatment and control watersheds, and then used these to estimate probabilities of increases or decreases in response to the restoration.

## Survival

We generated encounter histories for each individual PIT-tagged fish from active tagging, mobile antenna surveys, and continuous detections from passive instream antenna (PIA) arrays, located in four locations in Bridge Creek and one location in Murderers Creek. Separate encounter histories were generated for treatment and control watersheds. Because continuously collected detections by PIAs

**Table 1** Seasonal periods over which juvenile steelhead survival was estimated on Bridge and Murderers creeks before and after construction of BDAs on Bridge Creek, OR

| Year | Season | Impact period | Start | End | Total days |
|------|--------|---------------|-------|-----|------------|
| 2007 | Spring | Before | June 4, 2007 | September 1, 2007 | 89 |
| 2007 | Fall | Before | September 1, 2007 | November 28, 2007 | 88 |
| 2008 | Winter | Before | November 28, 2007 | June 1, 2008 | 186 |
| 2008 | Spring | Before | June 1, 2008 | September 1, 2008 | 92 |
| 2008 | Fall | Before | September 1, 2008 | December 12, 2008 | 102 |
| 2009 | Winter | Before | December 12, 2008 | June 10, 2009 | 180 |
| 2009 | Spring | Before | June 10, 2009 | September 14, 2009 | 96 |
| 2009 | Fall | Before | September 14, 2009 | January 22, 2010 | 130 |
| 2010 | Winter | After | January 22, 2010 | June 9, 2010 | 138 |
| 2010 | Spring | After | June 9, 2010 | September 13, 2010 | 96 |
| 2010 | Fall | After | September 13, 2010 | January 24, 2011 | 133 |
| 2011 | Winter | After | January 24, 2011 | June 24, 2011 | 151 |
| 2011 | Spring | After | June 24, 2011 | September 16, 2011 | 84 |
| 2011 | Fall | After | September 16, 2011 | January 10, 2012 | 116 |
| 2012 | Winter | After | January 10, 2012 | June 21, 2012 | 163 |
| 2012 | Spring | After | June 21, 2012 | September 11, 2012 | 82 |

were an important method for reencountering PIT-tagged fish, we used the Barker model (Barker 1997) rather than a Cormack-Jolly-Seber (CJS) model to estimate survival. We censored encounter histories for fish detected leaving tributaries (resighted at terminal antenna arrays) to reduce bias in survival estimates owing to permanent emigration (Horton and Letcher 2008; Conner et al. 2014). We used Program MARK (White and Burnham 1999; White et al. 2001) to analyze these data.

Because seasonal periods ($t$) were of slightly unequal length (Table 1), we standardized survival estimates ($S_t$) to a 3-month period (e.g., $S_t = 0.6$ is probability animal survived for 3 months) using unequal time intervals in Program MARK (White et al. 2001). There were eight seasons pre-restoration and seven seasons post-restoration over which survival was estimated (the last survival estimate was not used because it was confounded with resight probabilities). Because the study was designed as a BACIPS study, we analyzed data from treatment and control watersheds separately and left estimates of $S_t$ unconstrained in all models. That is, $S_t$ was estimated for each season before and after implementation of the BDAs for control ($S_{t\ control}$) and treatment watersheds ($S_{t\ treat}$). Before proceeding with a hierarchical model for $S_t$ using a Bayesian approach, we wanted to find the best model for the other parameters in the Barker model (e.g., p, R, F, etc.). To this end, we constructed a series of more parsimonious models for all other model parameters in the Barker model (see Supplemental Information for the details of model construction and model sets) and used the top model structure (i.e., model with the lowest AICc; Lebreton et al. 1992; Burnham and Anderson 2002) from which to estimate posterior distributions of $S_t$. Note that we did not use model averaging as part of the analysis because it would be a much more complex approach. That is, we would need to do MCMC simulations for each model in the set, and then apply the model weight and average across the 5000 simulations for each model, and then do the averaging across time periods (and sampling sites for abundance); this was beyond the scope of what we wanted to highlight for this paper.

We used a Bayesian hierarchical model with hyperdistributions to estimate mean survival and get "shrinkage" estimates for $S_t$ for treatment and control groups by before and after periods (e.g., $\tilde{S}_{control,before}, \tilde{S}_{treat,after}$). That is, we specified four hyperparameters. For these hyperdistributions, we used MCMC sampling implemented in Program MARK to generate posterior distributions of $S_{t\ control}$ and $S_{t\ treat}$, which were shrinkage estimates that we used for estimating ratios to evaluate the treatment effect as described below. Because this was the first time, we analyzed the data using a BACI model and because we used different subsets of the data for previous analyses, we used uninformative "flat" priors for the hyperpriors of the four estimates of mean survival ($S$):

$$\tilde{S} \sim N(\mu, \sigma)$$

$$\mu \sim N(0, 100)$$

$$^1/_{\sigma^2} \sim \gamma(0.001, 0.001)$$

where $\gamma$ represents a gamma distribution. In addition to the parameters included in hyperdistributions, there were additional "nuisance" parameters ($\theta$) in the Barker model (e.g., recapture probability, resighting probability, etc; see Supplemental information for Barker model parameter specification). These parameters also require a prior distribution. All additional model parameters were logit transformed to constrain the real estimates to be between 0 and 1. For these, we used a normal prior on the logit scale:

$$logit(\theta) \sim N(0, 1.75),$$

which is a relatively flat prior when back transformed to the real 0–1 scale (2.5th and 97.5th percentiles of approximately 0.02 and 0.98, with a uniform distribution between those percentiles when back transformed). We assessed convergence of the Markov chains by visual inspection of the trace of MCMC chains of the posterior samples of the parameters and by using the Gelman-Rubin statistic, R-hat (Gelman et al. 2004). For each parameter, we used ten chains of 1000 each and used a threshold of R-hat <1.1 to indicate adequate sampling of the posterior distribution. Based on diagnostics in Program MARK's MCMC routine (Cooch and White 2016), we determined posterior distributions needed to be thinned and accordingly saved every sixth sample to achieve first-order Markovian independence. We used 1000 burn in samples and kept 5000 samples after thinning.

To estimate treatment effects for this BACIPS study, we used the posterior distributions of $S_t$ to estimate the posterior distribution of the ratio of treatment to control

watersheds ($R_{t\ t|c}$) as $R_{t\ t|c} = S_{t\ \text{treat}} / S_{t\ \text{control}}$) for each time period. We then estimated the posterior distribution of the treatment effect for survival ($R_{S\ \text{BACI}}$) as $R_{S\ \text{BACI}} = \overline{R}_{t|c\ \text{after}} / \overline{R}_{t|c\ \text{before}}$. That is, for each MCMC sample, we calculated the mean ratio from the seven seasons after the BDAs were installed and the mean ratio from the eight seasons before the BDAs were installed, and then divided them. Note that because the ratios were log-normally distributed, we did all calculations on the log scale, and then back transformed the final $R_{S\ \text{BACI}}$ for each MCMC sample. We estimated the median and 2.5 and 97.5 percentiles for the distribution of $R_{S\ \text{BACI}}$.

Density

Abundance for each site was estimated from the two closed-capture sessions, which occurred at the start of each of the seasonal time periods described above for survival (Table 1), except for one additional capture session that occurred pre-treatment in January (winter) 2007. Thus, for each closed-capture session, a fish could have a 10 (captured the first session but not captured the second session), 11 (captured the first session and captured the second session), or 01 (not captured the first session but captured the second session) encounter history. We summarized these encounter histories for each of the two closed-capture sessions across sites for each time period (season). There were 18 abundance estimates, 10 before and 8 after BDAs were installed for each site. There were three additional abundance estimates relative to survival because there was an additional closed-capture session at the start of the study, survival could not be estimated for the last seasonal period, as discussed above, and survival is an interval estimate (i.e., there is one survival estimate between two closed-capture sessions/estimates).

We used a Bayesian MCMC approach to generate a posterior distribution of abundance ($N$) for each site and time period based on number of unique individuals captured ($n$) and capture probability. We used closed-capture model $M_0$ (Otis et al. 1978) and a data augmentation procedure (Royle and Dorazio 2012) for closed-capture models following Royle et al. (2007). We augmented each sample ($n$) by 500 ($z$) because this was more than twice any empirical abundance estimate for any of the study sites. This augmentation provided, in essence, a relatively uninformative prior (i.e., $M = z + n$,

and $N \sim DU(0, M)$ where DU = discrete uniform distribution; for details, see Royle and Dorazio 2012). To obtain a posterior distribution of site abundances, we used WinBUGS (Lunn et al. 2000), called from matbugs (available from http://code.google.com/p/matbugs/) in MATLAB (v. R2012b; MATLAB_8.0 2012). We ran model $M_0$ for each site and time period using 20,000 MCMC samples after discarding the first 1000 samples as burn in for each of three chains. We thinned by saving every third sample to reduce autocorrelations between samples; thus, we retained 5000 samples. We determined if the Markov chains converged using the Gelman-Rubin statistic (called Brooks-Gelman-Rubin statistics in WinBUGS), R-hat (Gelman et al. 2004). For each site and period, we used three chains of 5000 each and used a threshold of R-hat <1.1 for $N$ to indicate adequate sampling of the posterior distribution.

From the posterior distributions of abundance, we generated posterior distributions of density ($D$) for each site as fish/100 m by dividing each abundance estimate by the site length and then standardizing to 100 m. To generate one estimate per time period for treatment and control sites, we averaged the log of the density estimates across treatment and control sites for each time period. We did this for each MCMC sample to generate a posterior distribution of average density for treatment and control watersheds for each period. Then, similar to survival, we used the ratio of these estimates to estimate treatment effects for this BACIPS study. That is, we calculated the ratio of the treatment to control watersheds ($R_{t\ t|c}$) as $R_{t\ t|c} = \overline{D}_{t\ \text{treat}} / \overline{D}_{t\ \text{control}}$) for each time period. We then estimated the posterior distribution of the treatment effect for density ($R_{D\ \text{BACI}}$) as $R_{D\ \text{BACI}} = \overline{R}_{t|c\ \text{after}} / \overline{R}_{t|c\ \text{before}}$. That is, for each MCMC sample, we calculated the mean ratio from the eight seasonal periods after BDAs were installed and the mean ratio from the ten seasonal periods before the BDAs were installed, and then divided them. Note that because the ratios were log-normally distributed, we did all calculations on the log scale, and then back transformed the final $R_{D\ \text{BACI}}$ for each MCMC sample. We estimated the median and 2.5 and 97.5 percentiles for the distribution of $R_{D\ \text{BACI}}$.

Results

We used 5728 and 2410 marked juvenile steelhead on treatment and control watersheds before beaver dam

analogs were installed, and 7892 and 2227 after, for the analysis of survival. The Barker global model of survival fit adequately (i.e., there was not significant overdispersion or underdispersion); $c = 1.15$ for the control watershed data set and $c = 1.21$ for the treatment watershed. Because $c > 1$, we corrected and used QAICc for subsequent survival analyses. Survival estimates from the top-ranked model showed seasonal temporal variation for treatment and control watersheds, with the control watershed showing a consistent pattern of lower winter and higher spring and fall survival (Fig. 2a). Despite the temporal variation, the average survival on

the treatment watershed increased after installation of the BDAs, relative to the control watershed (Fig. 2b); $\overline{R}_{t|c \text{ before}} = 0.83$ and $\overline{R}_{t|c \text{ after}} = 1.13$, which resulted in an overall treatment effect $R_{S \text{ BACI}} = 1.36$. This indicates that survival on the treatment watershed increased, on average, 36 % after the beaver dam analogs were installed, relative to survival on the control watershed.

We used 4441 and 2440 marked juvenile steelhead on treatment and control sites before BDAs were installed, and 4955 and 1636 after for the analysis of density. Recapture rates were very similar for treatment and control sites both before (0.12 for both treatment and control) and after the

**Fig. 2** The **a** 3-month juvenile survival probability of steelhead and **b** the ratio of survival on treatment sites to control sites for each period (*gray line*), and the geometric mean of the ratios (*black line*) before and after BDAs were installed in Bridge Creek, OR, 2007–2012. For **b**, the *shaded area* represents the 95 % credible interval and the *dashed line* represents no treatment effect between control and treatment sites during the before or after periods; *values above the dashed line* indicate survival was higher on treatment sites relative to control sites, and *values below* indicate survival was lower on treatment sites relative to control sites
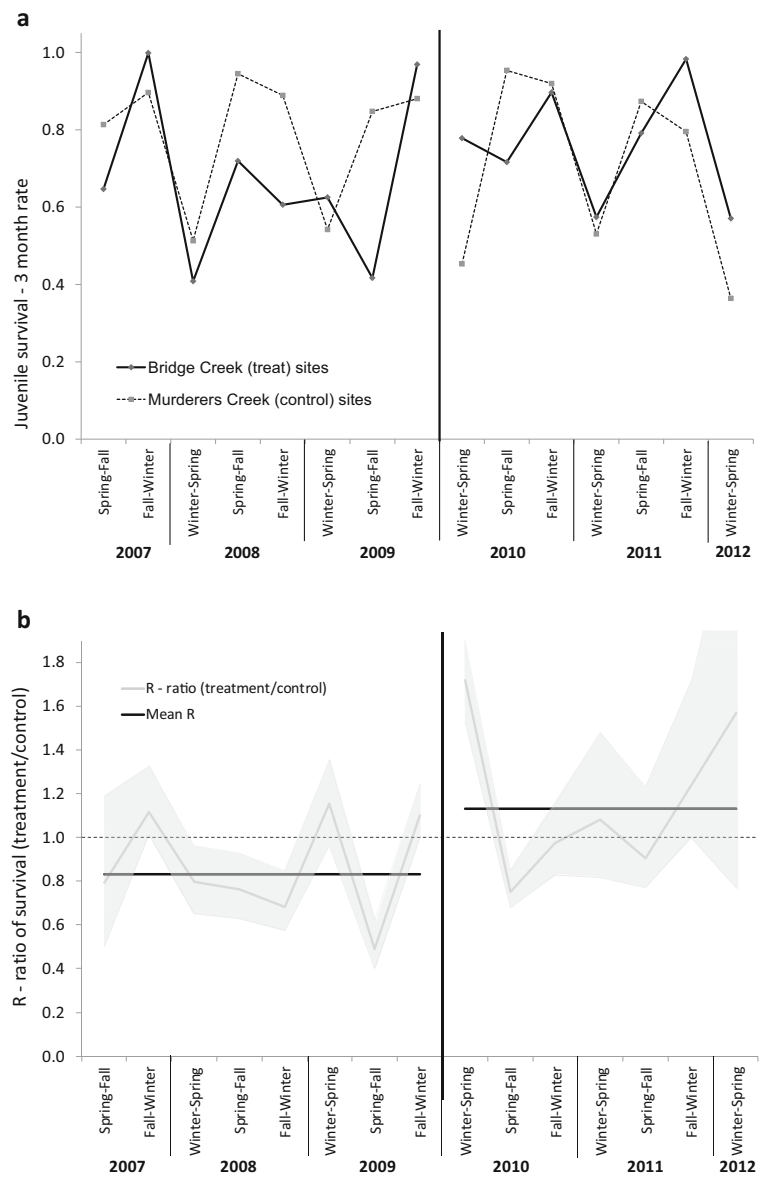
installation of BDAs (0.07 treatment and 0.10 control). Similar to survival, density estimates showed seasonal variation for treatment and control watersheds, with the control watershed showing a consistent pattern of lower winter and higher spring and fall density (Fig. 3a). The average density on the treatment watershed also increased after installation of the BDAs, relative to the control watershed (Fig. 3b); $\overline{R}_{t|c \text{ before}} = 0.60$ and $\overline{R}_{t|c \text{ after}} = 0.95$, which resulted in an overall treatment effect $R_{D \text{ BACI}} = 1.58$. This indicates that density on treatment watershed increased, on average, 58 % after the BDAs were installed, relative to density on the control watershed.

The posterior distributions of $R_{S \text{ BACI}}$ and $R_{D \text{ BACI}}$ indicate a zero probability that survival or density decreased after the BDAs were installed (Fig. 4). Note that a decrease would have been indicated by $R_{\text{BACI}} < 1$. After BDAs were installed, the probability of an increase of ≥30 % on the treatment watershed relative to the control watershed was high for both survival (0.88) and abundance (0.99; Table 2). The largest difference in the impact of the BDAs was for the probability of a ≥50 % increase; for survival it was only 0.17, while for abundance it was 0.82 (Table 2 and shown by shaded areas, Fig. 4). The posterior distribution of $R_{D \text{ BACI}}$ was

**Fig. 3** The **a** density of steelhead and **b** the ratio of density on treatment sites to the control sites for each period (*gray line*), and the geometric mean of the ratios (*black line*) before and after BDAs were installed in Bridge Creek, OR, 2007–2012. For **b**, the shaded area represents the 95 % credible interval and the *dashed line* represents no treatment effect during the before or after periods; *values above the dashed line* indicate density was higher on treatment sites relative to control sites, and *values below* indicate density was lower on treatment sites relative to control sites
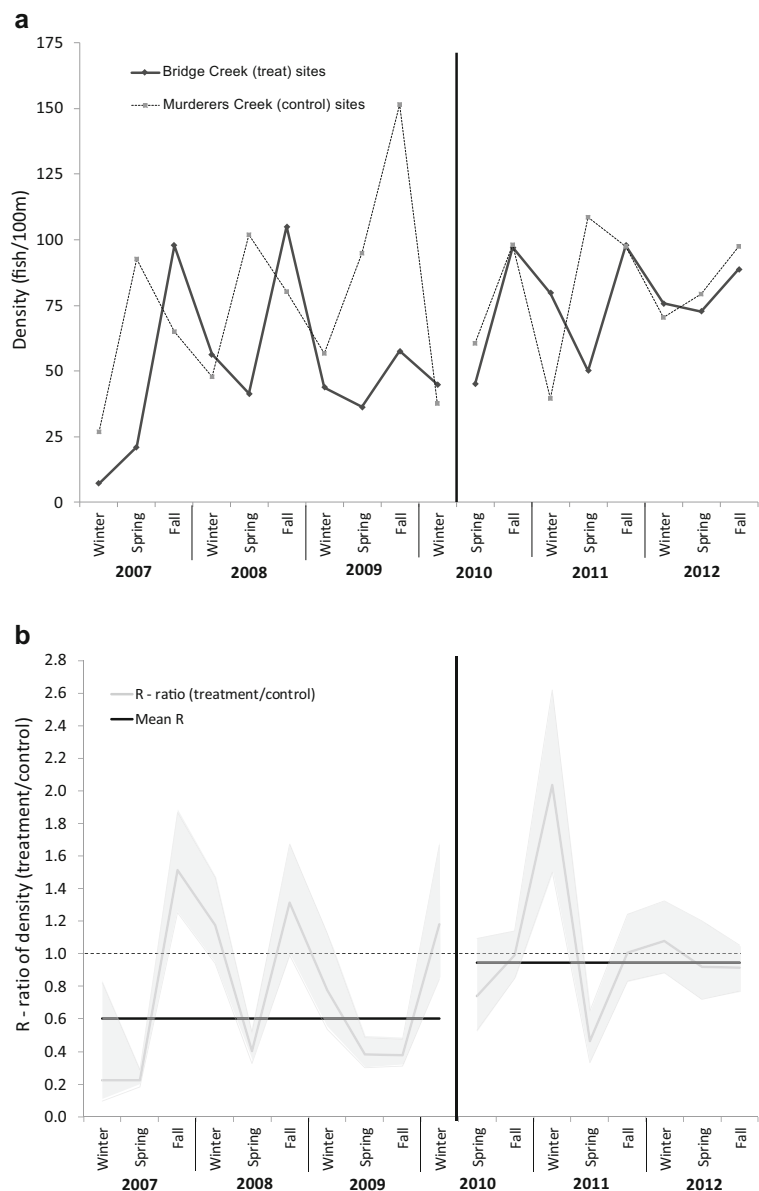
**Fig. 4** Distribution of the relative change ($R_{BACI}$) in **a** juvenile survival and **b** density of steelhead in Bridge Creek, OR, 2007–2012. The relative change is the ratio of the geometric mean of the ratios (for each period) of survival and density on treatment sites (Bridge Creek) relative to control sites (Murderers Creek), with the ratio after divided by the ratio before BDAs were installed. Relative change >1 indicates an increase on treatment sites relative to control sites after the BDAs were installed relative to before the dams were installed, while values of <1 indicate a decrease on treatment sites relative to control sites after the BDAs were installed relative to before they were installed. The *shaded columns* are an example showing the probability that survival or density increased by ≥50 %
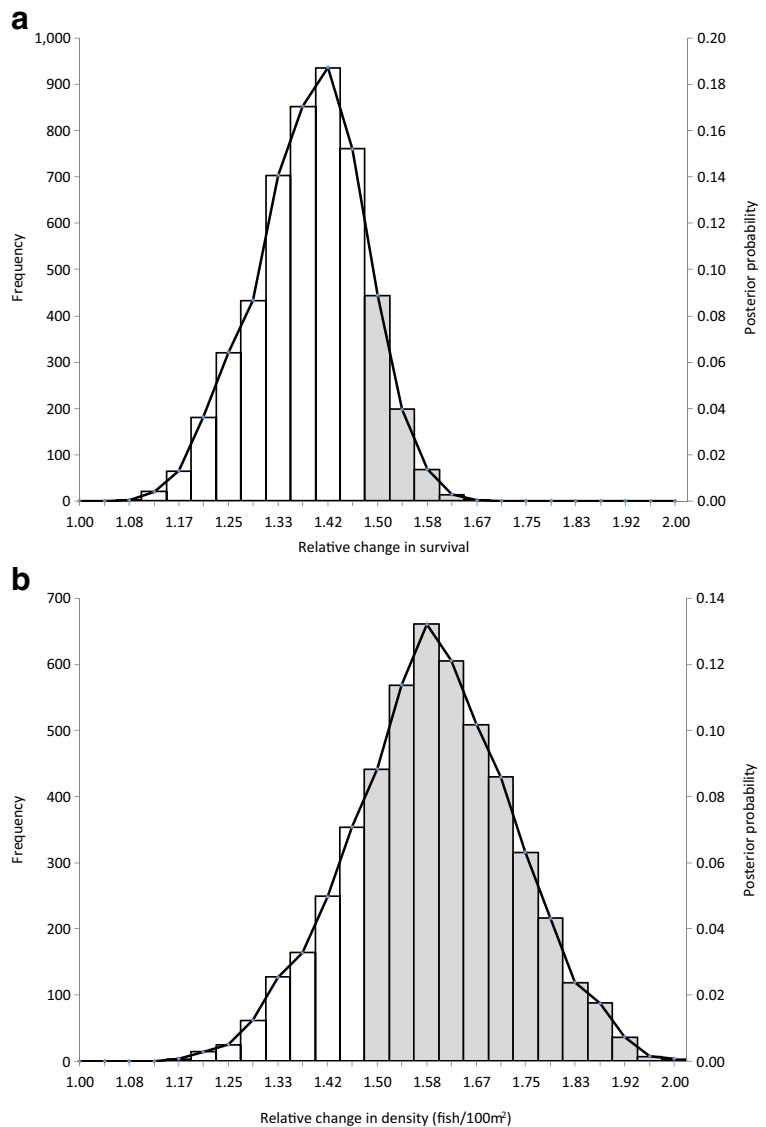


shifted to the right relative to $R_{S\ BACI}$, and so density showed higher probabilities of greater potential increases after the installation of BDAs compared to survival (Fig. 4 and Table 2). The variation in relative change was also greater for density than survival; the posterior distribution CI width was 43 % wider for density compared to survival (Fig. 4).

## Discussion

Our results demonstrate a useful extension of Bayesian methods to estimate probabilities of different effect sizes for BACI style study designs. Here, for two different

population parameters that had output that differed in distribution and magnitude, we quantified the probability that restoration had a negative or positive impact. In addition, we can readily evaluate different levels of impact. For example, the probability that BDAs increased both survival and density of juvenile steelhead by ≥50 % was 0.17 and 0.82, respectively (Table 2), and we can compare this to the probability of a more moderate increase of ≥30 % (0.88 and 0.99, respectively; Table 2). Indeed, the output metrics from a posterior distribution are flexible, and metrics such as presented here are intuitive to restoration and other management concerns and well adapted for decision making (Wade 2000).

**Table 2** Estimates of the probability juvenile survival and density increased or decreased a given percentage after BDAs were installed on study sites in Bridge Creek, OR, 2007–2012

| Parameter | ≥0 % | ≥20 % | ≥30 % | ≥50 % | ≥100 % |
|---|---|---|---|---|---|
| Increase | | | | | |
| Survival | 1.00[a] | 0.99 | 0.88 | 0.17 | 0.00 |
| Density | 1.00 | 1.00 | 0.99 | 0.82 | 0.00 |
| Decrease | | | | | |
| Survival | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Density | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |

[a] Probabilities are based on a posterior distribution of relative change ($R_{\mathrm{BACI}}$), which is the geometric mean of the ratios (for each period) of survival and density on the treatment watershed (Bridge Creek) relative to the control watershed (Murderers Creek), with the ratio after divided by the ratio before BDAs were installed

The combination of a ratio test statistic and the Bayesian approach yields results that are directly applicable to restoration and management questions (as well as for evaluating natural perturbation impacts). First, using a Bayesian MCMC approach to estimate this test statistic is particularly useful because the posterior probability distribution of the treatment effect ($R_{\mathrm{BACI}}$) can be used to directly draw inferences about the probability that there was a change in the response variable, given the observed data (Crome et al. 1996). Secondly, using a Bayesian approach provides accurate estimates of variation for the ratios (or any contrast, including nonlinear contrasts, of interest), whereas approaches to estimate variance from combined or transformed variables, such as the Delta method, can yield poor estimates where the function is nonlinear (Cooch and White 2016) or when the variance in the measured response is relatively large (e.g., $CV > 20$–50 %; Zhou 2002).

Additionally, using a test statistic that is a ratio of treatment to control observations provides directly interpretable effects in terms of the percent response of treatment sites, relative to control sites, after a restoration action was implemented relative to before period ($R_{\mathrm{BACI}}$). Thus, if $R_{\mathrm{BACI}} = 1.28$, there was a 28 % increase in the response variable in the treatment watershed after manipulation. As ratios provide an interpretation based on proportional responses, effect sizes are directly comparable across multiple response variables, relative to management targets or biologically reasonable responses, which can vary in both magnitude (daily growth versus animal abundance) and domain (e.g., survival [0–1] versus density [positive numbers]). Thus, while the mechanisms for changes in density (reproduction, mortality, immigration, emigration) and survival (mortality) following manipulation differ significantly, a ratio test statistic can be used to draw inference about the probability a manipulation would achieve management goals with different effect sizes across response variables in a consistent and easily comparable manner. For example, here we can easily compare the probabilities of restoration targets such as a 50 % increase in density (0.82) and a 20 % increase in survival (0.99). However, while a ratio test statistic provides a useful metric to quantify changes in a response variable following manipulations and facilitates comparison of observed effect sizes from multiple response variables or potential study targets, it does not directly imply biological significance of that response to a population of interest. For example, the impacts of a 20 % increase in juvenile steelhead survival following BDA installation on the population as a whole is dependent on the survival rate prior to manipulation and would need to be evaluated using a population projection model, to put it within the context of other demographic constraints.

Since the initial proposal of before-after (Box and Tiao 1965) and BACI (Green 1979) designs, the development of more sophisticated study designs, including the paired BACIPS (Stewart-Oaten et al. 1986), beyond-BACI (Underwood 1994), and multiple BACI (MBACI; Keough and Quinn 2000), has spawned an unresolved debate about the most appropriate study design to draw inferences from field studies involving nonrandom assignment of unreplicated treatments (Reckhow 1990; Underwood and Chapman 2003; Webb et al. 2010). However, studies are often constrained by resources, the existence of suitable reference sites, and the ability to collect data at reference and impact sites both before and after a perturbation

occurs for a long enough time series to have power to detect a change at impact sites. These constraints can result in high rates of rejecting the null hypothesis when in fact there was no impact (type I error; Murtaugh 2002), or sometimes accepting a null hypothesis when in fact there was an impact (type II error; Benedetti-Cecchi 2001). While it does not mitigate the importance of good study design, the Bayesian approach we describe partially alleviates the concern over type I and II errors by directly estimating the probability of observing an effect size (or range of effect sizes), conditional on the observed data, as opposed to probability of observing the data (or more extreme data), conditional on a specific hypothesis and assumptions that may not be satisfied by the study design and data.

We concur with recent assertions that estimation of effect size is more important, and more informative, than significance testing for management applications (Stewart-Oaten et al. 1992; Mapstone 1995; Crome et al. 1996). Manipulation of Bayesian posterior distributions allows analysts to determine the probability of observing any effect size of interest, or contrast the probability of effect sizes that differ in magnitude. For example, Bayesian approaches have been used to determine the probability that mean pH in Adirondack lakes increased by ≥10 % during a 7-year study period (Reckhow 1990), California spotted owl populations increased or decreased by ≥0, 30, and 50 % during a 20-year study period (Conner et al. 2013), bird community composition changed by greater than or equal to −25, 0, and 25 % owing to logging practices (Crome et al. 1996), and that there was a ≥75 % reduction in occupancy across sites after a hurricane (Russell et al. 2015). Thus, analyses can be readily framed to report the probability that a change of a magnitude deemed to be important to managers/policymakers has occurred. In contrast, the question asked by a classical hypothesis test is whether the test statistic calculated from the sample mean was unusual in comparison to what we would expect to calculate if there was no change. Inferences drawn about significant effect sizes from hypothesis-driven approaches can be subject to questions of biological significance and are often difficult to interpret with regard to management goals or conservation targets.

The combination of a ratio test statistic and Bayesian approach can easily be generalized to wide variety of study designs and provide an answer to the main study question—"How much impact (positive or negative) did the restoration action (or natural perturbation) have?"

While determining restoration management effects in a field setting is the main focus of this paper, the ratio and Bayesian approach could be applied to controlled experiments or treatment contrasts of other response variables as well. Primary to adapting this approach to other applications is defining the set of models that capture the study design and processes determining the response variables; in addition, this approach has the additional advantage that priors can be incorporated, if the data are available, for Bayesian analysis (see Wade 2000; Hobbs and Hooten 2015). Indeed, manipulation of posterior distributions can facilitate inferences drawn using a ratio test statistic. For example, Kimball et al. (2014) describe a split-plot designed experiment to evaluate the impacts of water and nitrogen input on percent cover of native shrubs. They provide estimates of percent native cover for different input levels, but could recast the results to describe the probability that water reduction (emulating drought conditions) decreased the percent native cover by 50 %, or some relevant ecological or management threshold. For other non-BACI study designs in less controlled field experiments, Bayesian methods have been used to describe widely ranging response variables of interest, including growth of individuals (Tanentzap et al. 2014; Tang et al. 2014), occupancy of species across a landscape (Russell et al. 2009), structure of physical habitat (Wallis et al. 2008), and biochemical makeup of terrestrial and aquatic systems (Qian et al. 2005; Larssen et al. 2006; Tanentzap et al. 2014). Such models can easily be adapted to provide posterior distributions that facilitate ratio contrasts between treatment and control experimental units for either controlled experiments, as well as field studies based on non-BACI study designs.

Management applications

BACI designs have been used to evaluate a variety of field experiments where randomization of treatment and/or control sites is not possible (Skilleter et al. 2006; Conner et al. 2007; Pitcher et al. 2009; Russell et al. 2015). While any management action can be evaluated with this approach, we believe it has particular relevance to restoration activities. Ecologists and managers tasked with conserving wildlife species, especially those showing declining population trends, often employ habitat restoration to enhance population vital rates and increase abundance. However, the majority of restoration activities go unevaluated (Bernhardt et al.

2005), while those that have been evaluated show varying degrees of success (Thompson 2006; Roni et al. 2008; Stewart et al. 2009; Whiteway et al. 2010). As a result, information is sparse as to which restoration activities recover declining populations as well as the extent to which restoration actions affect population responses. The BACI design can yield inference about impacts of restoration across broad scales (Underwood 1994; Keough and Mapstone 1995; Stewart-Oaten and Bence 2001), but to date has yet to be incorporated in many evaluations of restoration effectiveness (Miao et al. 2009). This is particularly unfortunate because, in many cases, the planning and permitting process involved with restoration activities provide an opportunity to initiate carefully designed BACI type studies, providing a time series of data both before and after restoration activities occur. In addition, the analysis of BACI data using a Bayesian approach is particularly well suited for the evaluation of restoration effectiveness as the inferences drawn about restoration impacts can be easily understood by many stakeholders.

The combination of a ratio test statistic and Bayesian approach we outline here, in conjunction with carefully a designed BACIPS study, provides ecologists and managers with an elegant means to quantify the probability of various effect sizes of interest, which can be a useful for managers trying to balance trade-offs between costly management actions and conservation of wildlife populations. Moreover, this approach provides results that are easily understandable to ecologists, managers, and stakeholders with a nonscientific background alike. We hope this approach will be useful for field ecologists and managers involved in restoration studies, but will also have wider application for any field study that suffers from a lack of adequate randomization and replication.

## References

Barker, R. J. (1997). Joint modeling of live-recapture, tag-resight, and tag-recovery data. *Biometrics, 53,* 666–677.

Bence, J. R., A. Stewart-Oaten, and S. C. Schroeter. 1996. Estimating the size of an effect from a before-after-control-impact paired series design.*in* R. J. Schmitt and C. W. Osenberg, editors. Detecting ecological impacts: concepts and applications in coastal habitats. Academic Press, San Diego, California.

Benedetti-Cecchi, L. (2001). Beyond BACI: optimization of environmental sampling designs through monitoring and simulation. *Ecological Applications, 11,* 783–799.

Bernhardt, E. S., Palmer, M., Allan, J., Alexander, G., Barnas, K., Brooks, S., Carr, J., Clayton, S., Dahm, C., & Follstad-Shah, J. (2005). Synthesizing U. *S. river restoration efforts. Science, 308,* 636–637.

Bousquin, S. G., & Colee, J. (2014). Interim responses of littoral river channel vegetation to reestablished flow after Phase I of the Kissimmee River Restoration Project. *Restoration Ecology, 22,* 388–396.

Bouwes, N., Weber, N., Jordan, C. E., Saunders, W. C., Tattam, I. A., Volk, C., Wheaton, J. M., and Pollock, M. M. (2016). Ecosystem experiment reveals benefits of natural and simulated beaver dams to a threatened population of steelhead (Oncorhynchus mykiss). *Scientific Reports, 6,* 28581.

Box, G. E. P., & Tiao, G. C. (1965). A change in level of a nonstationary time series. *Biometrika, 52,* 181–192.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: second edition*. New York, New York, USA: Springer-Verlag.

Carpenter, S. R., Frost, T. M., Heisey, D., & Kratz, T. K. (1989). Randomized intervention analysis and the interpretation of whole-ecosystem experiments. *Ecology, 70,* 1142–1152.

Conner, M. M., Bennett, S. N., Saunders, W. C., & Bouwes, N. (2014). Comparison of tributary survival estimates of steelhead using Cormac–Jolly–Seber and Barker models: implications for sampling efforts and designs. *Tranactions of American Fisheries Society, 143,* 320–333.

Conner, M. M., Keane, J. J., Gallagher, C. V., Jehle, G., Munton, T. E., Shaklee, P. A., & Gerrard, R. A. (2013). Realized population change for long-term monitoring: California spotted owl case study. *The Journal of Wildlife Management, 77,* 1449–1458.

Conner, M. M., Miller, M. W., Ebinger, M. R., & Burnham, K. P. (2007). A meta-BACI approach for evaluating focal management intervention on chronic wasting disease in free-ranging mule deer. *Ecological Applications, 17,* 143–150.

Cooch, E. G. and G. C. White. 2016. Program MARK: "a gentle introduction", 14th Edition. Available at http://www.phidot.org/software/mark/docs/book/.

Crome, F. H. J., Thomas, M. R., & Moore, L. A. (1996). A novel Bayesian approach to assessing impacts of rain forest logging. *Ecological Applications, 6*, 1104–1123.

Desrosiers, M., Planas, D., & Mucci, A. (2006). Short-term responses to watershed logging on biomass mercury and methylmercury accumulation by periphyton in boreal lakes. *Canadian Journal of Fisheries & Aquatic Sciences, 63*, 1734–1745.

Downes, B. J., Barmuta, L. A., Fairweather, P. G., Faith, D. P., Keough, J., Lake, P. S., Mapstone, B. D., & Quinn, G. P. (2002). *Monitoring ecological impacts: concepts and practice in flowing waters*. Cambridge, England: Cambridge University Press.

Eberhardt, L. L. (1976). Quantitative ecology and impact assessment. *Journal of Environmental Management, 4*, 27–70.

Gelman, A., Carlin, J. A., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis* (Second ed.). Boca Raton, Florida, USA: Chapman & Hall/CRC.

Green, R. H. 1979. Sampling design and statistical methods for environmental biologists. Wiley Interscience, Chichester, England.

Hanisch, J. R., Tonn, W. M., Paszkoswki, C. A., & Scrimgeour, G. J. (2013). Stocked trout have minimal effects on littoral invertebrate assemblages of productive fish-bearing lakes: a whole-lake BACI study. *Freshwater Biology, 58*, 895–907.

Hobbs, N. T., & Hooten, M. B. (2015). *Bayesian models: a statistical primer for ecologists*. Princeton University Press.

Horton, G. E., & Letcher, B. H. (2008). Movement patterns and study area boundaries: influences on survival estimation in capture-mark-recapture studies. *Oikos, 117*, 1131–1142.

Keough, M. J. and B. D. Mapstone. 1995. Protocols for designing marine ecological monitoring programs associated with BEK mills. National Pulp Mills Research Program Technical Report No. 11, CSIRO, Canberra.

Keough, M. J., & Quinn, G. (2000). Legislative vs. practical protection of an intertidal shoreline in southeastern Australia. *Ecological Applications, 10*, 871–881.

Kery, M. (2010). *Introduction to WinBUGS for ecologists: a Bayesian approach to regression, ANOVA, mixed models and related analyses*. San Diego, CA: Academic Press.

Kimball, S., Goulden, M. L., Suding, K. N., & Parker, S. (2014). Altered water and nitrogen input shifts succession in a southern California coastal sage community. *Ecological Applications, 24*, 1390–1404.

King, R., Morgan, B. J. T., Gimenex, O., & Brooks, S. P. (2010). *Bayesian analysis for population ecology*. Boca Raton, FL: CRC Press.

Larssen, T., Huseby, R. B., Cosby, B. J., Høst, G., Høgåsen, T., & Aldrin, M. (2006). Forecasting acidification effects using a Bayesian calibration and uncertainty propagation approach. *Environmental Science & Technology, 40*, 7841–7847.

Lebreton, J.-D., Burnham, K. P., Clobert, J., & Anderson, D. R. (1992). Modeling survival and testing biological hypotheses using marked animals: a unified approach with case studies. *Ecological Monographs, 62*, 67–118.

Louhi, P., Mäki-Petäys, A., Erkinaro, J., Paasivaara, A., & Muotka, T. (2010). Impacts of forest drainage improvement on stream biota: a multisite BACI-experiment. *Forest Ecology and Management, 260*, 1315–1323.

Lunn, D. J. A., Thomas, B. N., & Spiegelhalter, D. (2000). WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing, 10*, 325–337.

Mapstone, B. D. (1995). Scalable decision rules for environmental impact studies: effect size, type I, and type II errors. *Ecological Applications, 5*, 401–410.

Mathur, D., Robbins, T. W., & Purdy, E. L. (1980). Assessment of thermal discharges on zooplankton in Conowingo Pond, Pennsylvania. *Canadian Journal of Fisheries and Aquatic Sciences, 37*, 937–944.

MATLAB_8.0. 2012. The MathWorks, Inc., Natick, Massachusetts, USA.

Miao, S., Carstenn, S., Thomas, C., Edelstein, C., Sindhoj, E., & Gu, B. (2009). Integrating multiple spatial controls and temporal sampling schemes to explore short- and long-term ecosystem response to fire in an everglades wetland. In S. Miao, S. Carstenn, & M. Nungesser (Eds.), *Real world ecology* (pp. 73–109). New York: Springer Science + Business Media.

Muotka, T., & Syrjänen, J. (2007). Changes in habitat structure, benthic invertebrate diversity, trout populations and ecosystem processes in restored forest streams: a boreal perspective. *Freshwater Biology, 52*, 724–737.

Murtaugh, P. A. (2002). On rejection rates of paired intervention analysis. *Ecology, 83*, 1752–1761.

Otis, D. L., K. P. Burnham, G. C. White, and D. R. Anderson. 1978. Statistical Inference from Capture Data on Closed Animal Populations. Wildlife Monographs:3–135.

Pitcher, C. R., Burridge, C. Y., Wassenberg, T. J., Hill, B. J., & Poiner, I. R. (2009). A large scale BACI experiment to test the effects of prawn trawling on seabed biota in a closed area of the Great Barrier Reef Marine Park, Australia. *Fisheries Research, 99*, 168–183.

Pollock, M., J. M. Wheaton, N. Bouwes, C. Volk, N. Weber, and C. E. Jordan. 2012. Working with beaver to restore salmon habitat in the Bridge Creek intensively monitored watershed: Design rationale and hypotheses. U.S. Department of Commerce, NOAA, Seattle, WA.

Pollock, M. M., Beechie, T. J., Wheaton, J. M., Jordan, C. E., Bouwes, N., Weber, N., & Volk, C. (2014). Using beaver dams to restore incised stream ecosystems. *Bioscience, 64*, 279–290.

Popescu, V. D., de Valpine, P., Tempel, D., & Peery, M. Z. (2012). Estimating population impacts via dynamic occupancy analysis of before–after control–impact studies. *Ecological Applications, 22*, 1389–1404.

Qian, S. S., Reckhow, K. H., Zhai, J., & McMahon, G. (2005). Nonlinear regression modeling of nutrient loads in streams: a Bayesian approach. *Water Resources Research, 41*.

Reckhow, K. H. (1990). Bayesian inference in non-replicated ecological studies. *Ecology, 71*, 2053–2059.

Roni, P., Hanson, K., & Beechie, T. (2008). Global review of the physical and biological effectiveness of stream habitat rehabilitation techniques. *North American Journal of Fisheries Management, 28*, 856–890.

Royle, J. A., & Dorazio, R. M. (2012). Parameter-expanded data augmentation for Bayesian analysis of capture-recapture models. *Journal of Ornithology, 152*, 21–37.

Royle, J. A., Dorazio, R. M., & Link, W. A. (2007). Analysis of multinomial models with unknown index using data augmentation. *Journal of Computational and Graphical Statistics, 16*, 67–85.

Rumbold, D. G., Davis, P. W., & Perretta, C. (2001). Estimating the effect of beach nourishment on *Caretta caretta* (loggerhead sea turtle) nesting. *Restoration Ecology, 9*, 304–310.

Russell, J. C., Stjernman, M., LindstrÖM, Å., & Smith, H. G. (2015). Community occupancy before-after-control-impact (CO-BACI) analysis of Hurricane Gudrun on Swedish forest birds. *Ecological Applications, 25*, 685–694.

Russell, R. E., Royle, J. A., Saab, V. A., Lehmkuhl, J. F., Block, W. M., & Sauer, J. R. (2009). Modeling the effects of environmental disturbance on wildlife communities: avian responses to prescribed fire. *Ecological Applications, 19*, 1253–1263.

Schroeter, S. C., Dixon, J. D., Jon, K., Smith, R. O., & Bence, J. R. (1993). Detecting the ecological effects of environmental impacts: a case study of kelp forest invertebrates. *Ecological Applications, 3*, 331–350.

Skilleter, G. A., Pryor, A., Miller, S., & Cameron, B. (2006). Detecting the effects of physical disturbance on benthic assemblages in a subtropical estuary: a beyond BACI approach. *Journal of Experimental Marine Biology and Ecology, 338*, 271–287.

Stewart-Oaten, A., & Bence, J. R. (2001). Temporal and spatial variation in environmental impact assessment. *Ecological Monographs, 71*, 305–339.

Stewart-Oaten, A., Bence, J. R., & Osenberg, C. W. (1992). Assessing effects of unreplicated perturbations - no simple solutions. *Ecology, 73*, 1396–1404.

Stewart-Oaten, A., Murdoch, W. W., & Parker, K. R. (1986). Environmental impact assessment: "pseudoreplication" in time? *Ecology, 67*, 929–940.

Stewart, G. B., Bayliss, H. R., Showler, D. A., Sutherland, W. J., & Pullin, A. S. (2009). Effectiveness of engineered in-stream structure mitigation measures to increase salmonid abundance: a systematic review. *Ecological Applications, 19*, 931–941.

Tanentzap, A. J., Szkokan-Emilson, E. J., Kielstra, B. W., Arts, M. T., Yan, N. D., & Gunn, J. M. (2014). Forests fuel fish growth in freshwater deltas. *Nature Communications, 5*.

Tang, M., Jiao, Y., & Jones, J. W. (2014). A hierarchical Bayesian approach for estimating freshwater mussel growth based on tag-recapture data. *Fisheries Research, 149*, 24–32.

Tattam, I. A., Ruzycki, J. R., Li, H. W., & Giannico, G. R. (2013). Body size and growth rate influence emigration timing of *Oncorhynchus mykiss*. *Transactions of the American Fisheries Society, 142*, 1406–1414.

Thompson, D. M. (2006). Did the pre-1980 use of in-stream structures improve streams? A reanalysis of historical data. *Ecological Applications, 16*, 784–796.

Underwood, A., & Chapman, M. (2003). Power, precaution, type II error and sampling design in assessment of environmental impacts. *Journal of Experimental Marine Biology and Ecology, 296*, 49–70.

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

Wade, P. R. (2000). Bayesian methods in conservation biology. *Conservation Biology, 14*, 1308–1316.

Wallis, E., R. M. Nally, and P. Lake. 2008. A Bayesian analysis of physical habitat changes at tributary confluences in cobble-bed upland streams of the Acheron River basin, Australia. Water Resources Research 44.

Webb, J. A., Stewardson, M. J., & Koster, W. M. (2010). Detecting ecological responses to flow variation using Bayesian hierarchical models. *Freshwater Biology, 55*, 108–126.

White, G. C., & Burnham, K. P. (1999). Program MARK: survival estimation from populations of marked animals. *Bird Study, 46*(Supplement), 120–139.

White, G. C., Burnham, K. P., & Anderson, D. R. (2001). Advanced features of program MARK. In R. Field, R. J. Warren, H. Okarma, & P. R. Sievert (Eds.), *Land and people: priorities for the 21st century: Proceedings from The 2nd International Wildlife Management Congress* (pp. 368–377). Bethesda, Maryland, USA: The Wildlife Society.

Whiteway, S. L., Biron, P. M., Zimmermann, A., Venter, O., & Grant, J. W. A. (2010). Do in-stream restoration structures enhance salmonid abundance? A meta-analysis. *Canadian Journal of Fisheries and Aquatic Sciences, 67*, 831–841.

Zhou, S. (2002). Estimating parameters of derived random variables: comparison of the delta and parametric bootstrap methods. *Transactions of the American Fisheries Society, 131*, 667–675.