Received: 25 June 2019 | Accepted: 8 August 2019
DOI: 10.1111/2041-210X.13287

RESEARCH ARTICLE

Methods in Ecology and Evolution | BRITISH ECOLOGICAL SOCIETY

epower: An R package for power analysis of Before-After-Control-Impact (BACI) designs

Rebecca Fisher[1,2] | Glenn R. Shiell[3,4] | Rohan J. Sadler[5,6] | Karina Inostroza[3] | George Shedrawi[7] | Thomas H. Holmes[2,7] | James M. McGree[8]

[1]Australian Institute of Marine Science, Crawley, WA, Australia; [2]The UWA Oceans Institute and School of Plant Biology, University of Western Australia, Crawley, WA, Australia; [3]BMT, Osborne Park, WA, Australia; [4]School of Molecular and Life Sciences, Curtin University, Bentley, WA, Australia; [5]Pink Lake Analytics Pty Ltd, Guildford, WA, Australia; [6]School of Agricultural and Resource Economics, The University of Western Australia, Crawley, WA, Australia; [7]Department of Biodiversity, Conservation and Attractions, Biodiversity and Conservation Science, Marine Science Program, Kensington, WA, Australia and [8]School of Mathematical Sciences, Queensland University of Technology, Brisbane, Qld, Australia

Correspondence
Rebecca Fisher
Email: r.fisher@aims.gov.au

Funding information
This work was funded by two BMT (http://www.bmt.org) innovations grants.

Handling Editor: Robert Freckleton

Abstract

1. Before-After-Control-Impact (BACI) designs are widespread in environmental science, however their implicitly hierarchical nature complicates the evaluation of statistical power. Here, we describe *epower*, an R package for assessing statistical power of BACI designs.
2. The package uses Bayesian statistical methods via the R-package INLA to fit the appropriate hierarchical model to user supplied pilot survey data. A posterior sample is then used to build a Monte Carlo simulation to test statistical power specifically for the Before/After × Control/Impact interaction term in the BACI model. Power can be assessed for any number of user-specified effect sizes for the existing design, or across a range of levels of replication for any part of the sampling design hierarchy.
3. The package offers a user friendly robust approach for assessing statistical power of BACI designs whilst accounting for uncertainty in parameter values within a fully generalized framework.

KEYWORDS
Bayesian methods, before-after-control-impact, environmental impact detection, hierarchical model, mixed model analysis, power analysis, R, sampling design assessment

1 | INTRODUCTION

Experimental and manipulative sampling designs that incorporate a before and after period (BA) in conjunction with control and impact (CI) treatments are popularly known as Before-After-Control-Impact, or BACI designs (Green, 1979). BACI designs are widely utilized in empirical research, ecology and conservation, but have particular application to the fields of environmental management and environmental impact assessment. In a research context, BACI designs may be employed as an underlying framework for the experimental design (Grear, Luong, & Hudson, 2012; Lunde, Resh, & Johnson, 2012). In conservation BACI designs have been used to examine the impacts of human interactions on wildlife populations (Foroughirad & Mann, 2013), the success of restoration efforts (e.g. Audet, Elsgaard, Kjaergaard, Larsen, & Hoffmann, 2013; Holsman, McDonald, Barreyro, & Armstrong, 2010), environmental impacts (Clark et al., 2018; Claudet, Lenfant, & Schrimm, 2010; Terlizzi et al., 2005), recovery following impact alleviation (Aguado-Giménez et al., 2012) and as a means of validating predictions made during the initial environmental impact assessment (Smith, 1991).

In a model for data from a BACI design there are two fixed effects: a Before-After (BA) and a Control-Impact (CI) (or 'treatment') effect. Individually, these main effect terms are uninformative in impact assessment, as a significant BA effect simply indicates a change from before to after the impact, which may result from natural variation in time. Similarly, a significant CI effect may occur due to natural spatial random variation, reflecting inherent differences between control and impact locations. For a BACI model, it is the interaction term between the BA and CI terms that indicates a significant effect of the treatment.

Natural systems are often highly variable in space and time causing difficulties in the interpretation of BACI experiments due to confounding, such that treatment effects are interpreted as significant when in fact they may not be (Type I error), particularly if the number of control locations and/or after (post- impact) surveys is low (Underwood, 1991, 1993). To avoid high rates of Type I error associated with high levels of variability the use of multiple sampling times before and after an impact (Stewart-Oaten, Murdoch, & Parker, 1986) is recommended, along with multiple control and/or impact sites or locations (Glasby & Underwood, 1998; Terlizzi et al., 2005; Underwood, 1991, 1994). Such designs are referred to as MBACI (Keough & Mapstone, 1997), BACIP (Stewart-Oaten et al., 1986), beyond-BACI (Underwood, 1994), or Impact versus Reference site designs (Stewart-Oaten & Bence, 2001). In most cases, extended BACI designs are typically asymmetric, with uneven numbers of control and impact sites, and/or before and after sampling times.

Spatiotemporal variability can mask real treatment effects, e.g. (Coleman, Palmer-Brodie, & Kelaher, 2013) leading to high Type II error (β, failure to find a significant effect when in fact there is one) and subsequently low statistical power ($1 - \beta$). The two types of inference error: Type I and II (Neyman & Pearson, 1967) each have implications for applied ecological studies, but the severity of the consequences of committing these errors depends on the purpose of the study/experiment, and on the perspective of the relevant stakeholder(s). Assessing statistical power of such designs is essential for interpreting findings, especially where these findings will contribute to future assessments of potential environmental impact (Constable, 1991) and wherever possible should be done 'a-priori' using real pilot survey or 'before' data to understand if existing and/or proposed sampling designs are sufficient to detect relevant changes.

Extended BACI models are hierarchical, representing mixed models with both fixed and random factors, capturing multiple sources of variation across space and time. Analysis of such designs is becoming straightforward through a range of statistical packages in R, e.g. LME4, (Bates, 2010; Pinheiro, Bates, DebRoy, & Sarkar, 2013) and elsewhere, e.g. PERMANOVA+ (Anderson, Gorley, & Clarke, 2008). However, assessment of statistical power for mixed models is challenging, requiring simulation (Benedetti-Cecchi, 2001; Underwood & Chapman, 2003). Here, we describe *epower*, an R package designed to assess statistical power of BACI designs for detecting environmental impacts. While there are some existing R packages available for the assessment of statistical power of mixed models (Green & MacLeod, 2016; Johnson, Barry, Ferguson, & Muller, 2015), *epower* is unique in that it uses Bayesian statistical methods to incorporate uncertainty associated with, for example, hierarchical variance estimates and also allows a broad range of statistical distributions to be specified, ensuring it can handle most common response variables. In addition, the *epower* package has been developed specifically for the purposes of assessing the power of the BACI interaction term, making it simpler to apply in practice, with broad accessibility achieved directly through R as well as an excel workbook interface, allowing greater uptake in an environmental management setting.

2 | THE *EPOWER* PACKAGE

2.1 | Designs for assessing environmental impact

An impact is defined as a change in an indicator of interest (the response variable) at the defined (impact) location, and is measured as the difference in the response before and after the disturbance compared to any changes in the control locations. To demonstrate an impact, there must be statistical evidence for a difference between the control and impacted locations (CI, Figure 1) accounting for any natural before and after (BA, Figure 1) changes following the disturbance (Underwood, 1993). Because ecological systems can be highly variable in space and time, any such interaction must be detectable beyond naturally occurring spatial and temporal variability.

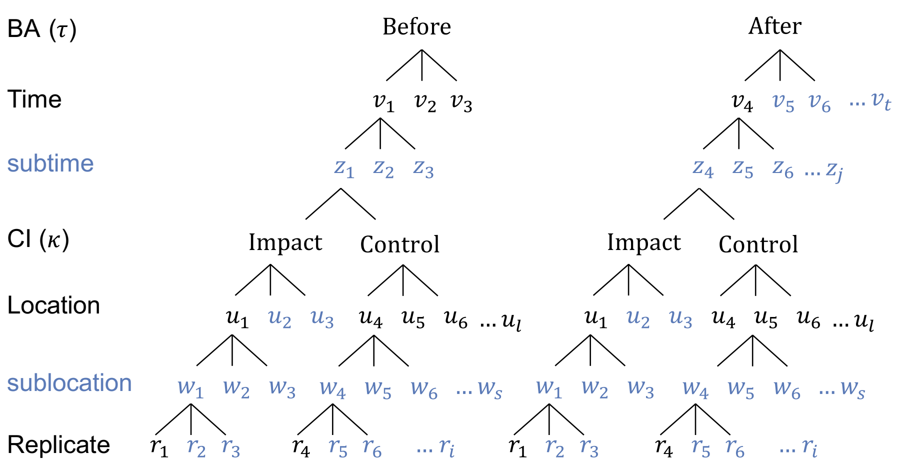

FIGURE 1 Conceptual diagram of a Before-After-Control-Impact (BACI) sampling design. Elements in blue indicate optional parts of the design and those shown in black are essential for a true BACI assessment of impact

As such, sampling designs to detect change should involve replicate control locations nested within the control and impact fixed effect to assess among location variation, as well as repeated sampling periods prior to disturbance to assess temporal variation (Figure 1). In the conceptual BACI design in Figure 1, we have included three control locations (u_1, u_2, u_3) and three before sampling periods (v_1, v_2, v_3). Often sampling designs to assess impact also contain other hierarchical spatial (such as sublocations nested within Locations) or temporal elements (e.g. months nested within years) that allow further partitioning of spatial and temporal variation (Figure 1). These extra hierarchical levels are optional and are not required for a BACI impact assessment (see blue text in Figure 1 and Model 1). Because of their spatial and temporal hierarchies BACI sampling designs are most appropriately analysed using generalized mixed models (GLMMs) that capture spatial and temporal grouping factors as random effects. For the detection of a 'significant' impact, the specific term of interest is the BACI interaction term ($\tau\kappa$, Model 1), after having taken into account location and sampling time random effects (u and v and optionally w and z, Model 1). This BACI term must be deemed 'significant' beyond any inherent random spatial and temporal interactions in the system (uv and optionally vw, uz and wz, Model 1) In this paper (and in *epower*), this is achieved through an assessment of the Bayesian posterior probability of Model 1 compared to an alternative model (Model 2) which is identical in every respect to Model 1, except without the BACI interaction term (details given in the next section).

Net change in the environmental response variable of interest should be assessed using GLMMs, and should include all available reference sites and baseline data available, following a typical BACI approach (Green, 1979). The *epower* package has been developed based on the assumption that the BACI experimental survey design contains at least one random spatial grouping factor (i.e. replicate control locations in addition to at least one impact location, denoted as Location in Figure 1 and $\underline{u}$ in Model 1), and at least one random temporal grouping factor (i.e. replicate before surveys in time, denoted as Time in Figure 1 and v in Model 1), but allows for up to one additional spatial and/or temporal hierarchical (nested) random effect to be included (i.e. sublocation, or subtime, Figure 1, w and z Model 1). Where fixed replicates are used (for example, the same transect is visited repeatedly every time), this level corresponds to 'sublocation' in the spatial hierarchy, with the number of randomly measured replicates set at 1 (as indicated by the fact that any replicates >1 are optional, Figure 1). Including the fixed replicates as a 'sublocation' in the spatial hierarchy ensures that a random variance associated with the fixed transects is properly estimated by the mixed model, accounting for their non-independence (as in the case study below). Non-nested random effects are assumed to be fully crossed, thus interactions are also included where appropriate (for example, there is a Time × Location random effect capturing random within Location temporal variation). For the design in Figure 1, Model 1 can be described as follows when the response is assumed to follow a Normal distribution:

$$y_{iltsj} \sim N\left(\mu_{iltsj}, \sigma_\varepsilon^2\right)$$

$$\mu_{iltsj} = \mu + u_l + v_t + k_{lt} + w_{[l]s} + z_{[t]j} + m_{ts} + p_{lj} + q_{sj}$$

$$\mu = \beta_0 + \tau + \kappa + (\tau\kappa)$$

$$u_l \sim N\left(0, \sigma_u^2\right), \quad v_t \sim N\left(0, \sigma_v^2\right), \quad k_{lt} \sim N\left(0, \sigma_k^2\right),$$
$$w_{[l]s} \sim N\left(0, \sigma_w^2\right), \quad z_{[t]j} \sim N\left(0, \sigma_z^2\right)$$

$$m_{ts} \sim N\left(0, \sigma_m^2\right), \quad p_{lj} \sim N\left(0, \sigma_p^2\right), \quad q_{sj} \sim N\left(0, \sigma_q^2\right)$$

$$\sigma^2 = \sigma_u^2 + \sigma_v^2 + \sigma_k^2 + \sigma_w^2 + \sigma_z^2 + \sigma_m^2 + \sigma_p^2 + \sigma_q^2 + \sigma_\varepsilon^2,$$

where μ_{iltsj} represents the conditional mean response with the Before-After effect (τ), Control-Impact effect (κ) and BACI interaction term ($\tau\kappa$), u_l, v_t, k_{lt}, $w_{[l]s}$, $z_{[t]j}$, m_{ts}, p_{lj} and q_{sj} denote the random effects for location l, time t, the interaction between location and time, sublocation w_s nested within location l, subtime z_j, nested within time t, the interaction between time and sublocation, the interaction between location and subtime, and the interaction between sublocation and subtime (respectively), and ε denotes the residual error, for t = 1, …, T (the number of time points), l = 1, …, L (the number of locations), s = 1, …, S (the number of sublocations), j = 1, …, J (the number of subtimes) and i = 1, …, R (the number of replicate observations). The total variance σ is given by the sum of the residual error variance and the individual random factor and interaction variances.

To complete the description of the above Bayesian model, the prior distributions need to be defined (see below).

For responses that are not continuous, e.g. binary or count data, Model 1 is specified as above but without the residual variance term, instead with whatever required additional hyperparameters (for example, the overdispersion parameter of the negative binomial distribution), and using the default link functions available in INLA (see below). Model 2 is then specified exactly as Model 1 but without the ($\tau\kappa$) term in the mean response.

2.2 | Bayesian mixed models and Integrated nested Laplace approximation

As described above, BACI designs require that statistical testing is performed in the context of mixed modelling, capable of testing for the *fixed* BACI interaction term whilst including any relevant *random* factors and their interactions (Model 1). While there are a range of methods available for fitting mixed models in R (Bates, 2010; Bolker et al., 2009; Pinheiro et al., 2013) and elsewhere (Anderson, 2001), model and parameter uncertainty are most rigorously handled within a Bayesian inference framework, and is thus chosen for use within the *epower* package. An additional advantage of Bayesian modelling here is that the model can be fit even if the pilot data are minimal, which is often difficult using maximum likelihood approaches.

Bayesian inference is a form of statistical inference where prior knowledge is combined with information from the data to form a posterior distribution. It is this distribution with which inference is based. It follows that the posterior distribution can be defined as follows:

$$p(\theta|y,d)=\frac{p(\theta)\,p(y|\theta,d)}{p(y|d)},$$

where $p(\theta)$ denotes the prior distribution/information about the parameter θ, $p(y \mid \theta, d)$ denotes the likelihood, $p(\theta \mid y, d)$ denotes the posterior distribution and $p(y|d)=\int_\theta p(\theta)\,p(y \mid \theta, d)\,d\theta$ is the normalizing constant which ensures the posterior distribution is a proper density function (integrates to one). Throughout this paper, the design d denotes the BACI design, and, as such, contains all information about, for example, the number of observations in the Before-After and Control-Impact locations.

The normalizing constant is sometimes referred as the *marginal likelihood*, and can be used for model choice when more than one model is considered for analysis (MacKay, 2003). The above expression for the posterior distribution can be extended to allow for model uncertainty as follows:

$$p(\theta_m|y,M=m,d)=\frac{p\left(\theta_m|M=m\right)p\left(y|\theta_m,M=m,d\right)}{p\left(y|M=m,d\right)},$$

where m = 1, …, K and M is the random variable associated with the model indicator.

Typically, there will be prior information associated with the K. models, and this can be denoted as $p(M = m)$. Then, the posterior model probability of each model can be found as follows:

$$p\left(M=m|y,d\right)=\frac{p\left(M=m\right)p\left(y|M=m,d\right)}{\sum_{k=1}^{K}p\left(M=k\right)p(y|M=k,d)}.$$

The preference is then for the model with the highest posterior model probability. Thus, uncertainty in the choice of preferred model is captured via a Bayesian approach to statistical inference. In *epower* there are only two models (i.e. K = 2), and therefore only two outcomes, and thus a preference either way is indicated by a probability >.5. In evaluating the posterior model probability, there is an inbuilt penalty for model complexity, thus when there is no impact (and there is no preference for either model a priori), there will be a preference for the simpler model (without the BACI interaction term). This can also be influenced by the chosen priors on the parameters. The priors used in *epower* are based on the default priors supplied by INLA (see below) and are $N(0, 0)$ for fixed effect intercept terms and log-gamma(1, 5e−5) for (shape, inverse-scale) for log-precision for variance terms.

Given the advantages of a Bayesian inference framework, we adopt this framework for both the analysis of these BACI designs (when before and after data are available, Model 1), and for the estimation of their statistical power (if only before data are available). The general approach to estimating statistical power for mixed models is based on simulating a large number of datasets, with corresponding model fits. This poses a substantial computational challenge, particularly in the context of Bayesian inference. To achieve reasonable computational times, we used an integrated nested Laplace approximation (INLA) for approximating posterior distributions (Rue, Martino, & Chopin, 2009). This is a deterministic approach for approximating a posterior distribution which is flexible enough such that it can be used across a wide variety of statistical models. Importantly, one can form an approximation much more efficiently (computationally) when compared to other methods such as Markov Chain Monte Carlo and sequential Monte Carlo. Another advantage of INLA is that one obtains an efficient estimate of the marginal likelihood of a given model. As the marginal likelihood has an inbuilt penalty for model complexity, it can be used for model choice with a preference for the model with the largest of such values (MacKay, 2003). Once the marginal likelihood for each competing model has been approximated, posterior model probabilities can be found straightforwardly (as shown above).

Power evaluations may be misleading if the Monte-Carlo procedure constructed from the posterior distribution generates unrealistic spatiotemporal patterns of means and variances in the simulated variable. Similar issues apply to power analyses using simulations based on variance estimates obtained through other mixed modelling approaches, such as glmer and lmer from the LME4 package in R, as used by the SIMR (Green & MacLeod, 2016). Verifying that the model used to estimate power is appropriate for the observed data is essential. A common approach for assessing goodness-of-fit of Bayesian models is the probability integral transform, or PIT values (Dawid, 1984), defined as

$$P(Y_i \le y_i|y_{-i}),$$

where y_{-i} denotes the full dataset without the ith observation.

Accordingly, the PIT values should follow a standard uniform distribution. The *epower* package will automatically generate a plot of the PIT values for the model fitted to the pilot data, and this plot should be inspected to determine if there is any evidence that the model is not appropriate for the data. The estimate of power given by *epower* is based on the assumption that the model is appropriate for the data (and that the 95% credible intervals include the true value of the parameters). If such assumptions are not valid, then the estimate of power may be unreliable because the Monte-Carlo procedure may generate unrealistic spatiotemporal patterns of means and variances in the simulated datasets. In this case, alternative models should be explored, to ensure adequate model fit to the data.

A range of environmental monitoring designs are often implemented in practice, and as such the specification of the random effects for Models 1 and 2 may change dynamically with respect to the specific design being examined (see annotated blue text in Figure 1 and Model 1). The *epower* package implements designs of up to two spatial hierarchical levels and two temporal hierarchical levels. By default, *epower* will fit a complete (and identical) random structure during the estimation of both Model 1 and Model 2, with this structure dependent on the design of the sampling program as specified by the user, and contained within the pilot data being tested. This includes estimates of random normal deviations (as precision) among repeated Locations (control and impact, after accounting for the fixed effect of control vs. impact), sublocations, sampling Times, and sampling subtimes (whichever of these are present in the design, see Figure 1), as well as their relevant interactions. Currently *epower* assumes that if two spatial or temporal levels exist, one level is nested within the other (ie Sites nested within locations, months nested within Years). A row level random effect is also included where a binomial model is specified and the number of trials per row is greater than 1 (i.e. multiple trials occur for a single row of data). Random effect terms are included through the use of the `f(z, model="iid")` syntax available in inla() which implements an independent random noise model for random term z (Blangiardo, Cameletti, Baio, & Rue, 2013).

3 | IMPLEMENTATION AND USAGE

The *epower* package has been designed such that it can be used by ecologists and environmental managers with a sound understanding of the statistical principles around BACI designs, a working knowledge of Microsoft Excel, but with limited practical programming skills in R. Our desire was that the *epower* package would be accessible to a wide range of applied ecological consultants and research scientists. The *epower* package uses an excel workbook as an interface for supplying relevant pilot data, specifying elements of the sampling design as well as the scenarios to be explored (green boxes, Figure 2). For those proficient in the use of R, an alternative R only workflow is also provided.

In addition to the excel workbook interface, the *epower* package is composed of a series of R functions (blue boxes, Figure 2), that import and unpack the data contained in the excel interface, perform an `inla` generalized mixed model statistical analysis of the pilot data, extracts and collates the relevant parameter estimates, builds and runs the Monte Carlo simulation based on a matrix of scenarios (all possible combinations of specified numbers of Locations, sublocations (if present), Times, subtimes (if present), and replicates), and finally summarizes the simulation results. There are only two functions that are called by the user (if using the excel interface), including: `fitData`, and `assessPower` (Figure 2). While `fitData` must be called at the outset to run the *epower* package, the function `assessPower` is optional. The outputs generated by the *epower* package and written to file include the model fit statistics and a PIT histogram generated by `fitData`; and optionally a table of scenario details and associated proportions of detected impacts (across the number of simulation iterations) generated by `assessPower`. Detailed usage instructions and a description of all the design elements available in *epower* can be found in Appendix S1.

4 | CASE STUDY

We used *epower* to assess the statistical power of the sampling design implemented by Western Australia's Department of Biodiversity, Conservation and Attractions to detect the impacts of natural and anthropogenic disturbances on the cover of hard corals within the Ningaloo marine reserve. This program is used to assess the effectiveness of ongoing management strategies and facilitate the adaptive management of ecological values within the reserves for conservation purposes.

As disturbances are spatially and temporally variable, the design encompasses sites spread across the entire length of the marine reserve, incorporating three distinct oceanographic sectors (Bundegi, North and South) and areas of high and low human use (see Appendix S2, Figure S2.1). Our analysis examined power of the monitoring program to detect change in total hard coral cover, a metric often used to assess the relative condition of these coral dominated communities. As *epower* has been developed specifically for the purposes of evaluating power within a BACI context, here we focus on assessing power to detect a press disturbance (natural or otherwise) occurring at the scale of a single monitoring site (small scale local impact), using the other nearby monitoring sites as controls. Because coral cover can change markedly through time, we used only the three most recent sampling periods from the supplied dataset (2013, 2014 and 2015) for each of the 22 sites as our 'before' data in the power analysis (Figure S2.2).

Power for a range of multiplicative effect sizes was examined for each site within each sector for the existing design (three fixed transects at each site, with 300 points scored across the collected images on each transect), using the three with the most similar standard deviation and mean of coral cover to the 'impact' location as 'controls' and a single hypothetical 'after' survey, as well as for two and three post impact surveys. Here, we use the three most similar sites as control locations to represent the hypothetical optimal design for detecting an impact at a given location. In a BACI analysis reference sites should represent broad scale 'natural' spatial variability. However, reference sites should also be selected such that they are as much as possible representative of the impact location. The long-term monitoring program was not developed specifically with a BACI analysis in mind, and there are sometimes large differences in the coral cover and communities among sites within each sector (see Figure S2.2). Using the sites that are most similar satisfies the requirement that the controls are the most representative of the impact location of those available.

For the analysis of the case study data we included random effects of Location (u_l from Model 1, here termed 'site'), sublocation ($w_{[l]s}$ from Model 1, the fixed transects) and Time (v_t from Model 1) in a binomial model, with 'successes' the number of points scored as hard coral, and 'trials' the total number of scored points. For this design, there is only one replicate, because the transects are fixed and their repeated sampling through time must be modelled using transect as a 'sub-location' random effect. We examined power of the existing design to detect changes of −0.2 (20% loss) through to −0.8 (80% loss), given a single 'after' survey, as well as for two and three post-impact surveys. We also explored how changing the number of fixed transects (sublocations) and scored points ('trials' in the binomial call) altered the statistical power to detect a 30% decline for three sites spaning a range of statistical power for the existing design.

Prior to examining power, we assessed the goodness-of-fit of the models, using posterior predictive checks and inspected the distribution of the PIT values. Overall, the simulated data (transparent red

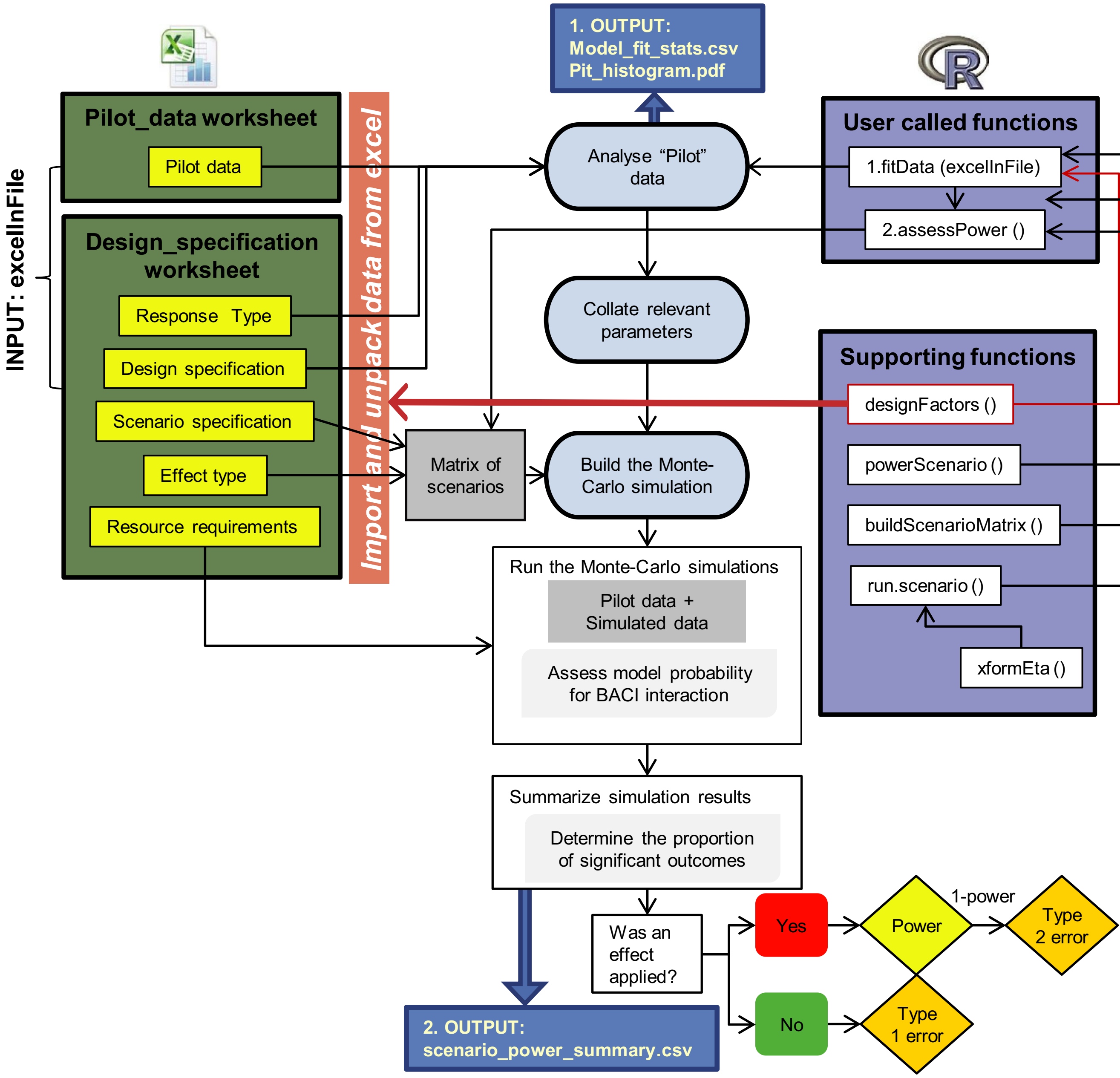

FIGURE 2 Conceptual flow diagram of *epower*

circles) appear consistent with the observed data (solid black circles) with no obvious systematic bias (Figure 3). A common measure used to further assess goodness-of-fit based on these posterior predictive checks is to evaluate the proportion of data points that lie outside the 95% credible intervals of the simulated data. For the three examples, this proportion was 0.09, 0.03 and 0.03 respectively, which are reasonable given only ~33 observations in each example, and thus do not indicate any significant lack of fit.

The distribution of PIT values for each example appear to be relatively evenly distributed across the interval 0 to 1 for Tantabiddi and Coral Bay, given the relatively small sample size (Figure 3). However, for Bundegi, the PIT values appear to be higher than expected indicating a tendency to under predict cross-validated data (Figure 3). Such a tendency indicates a lack of fit of the model, and may explain the higher than expected number of points lying outside the 95% credible intervals (9%). However, we note that the range of response values for Bundegi is relatively narrow as coral cover is extremely low at this site, so this under prediction does not translate into large absolute differences between the observed and predicted response. Nonetheless, the estimated power for this example should be interpreted with caution.

Power varied substantially among the sites examined across the Ningaloo marine reserves (Appendix S2, Figure S2.3) and was consistently lower when only a single ‘after’ survey was carried out, and increased with both two and three ‘after’ surveys (Figure 4). For most sites, there was at least 80% power (dashed horizontal line) to detect

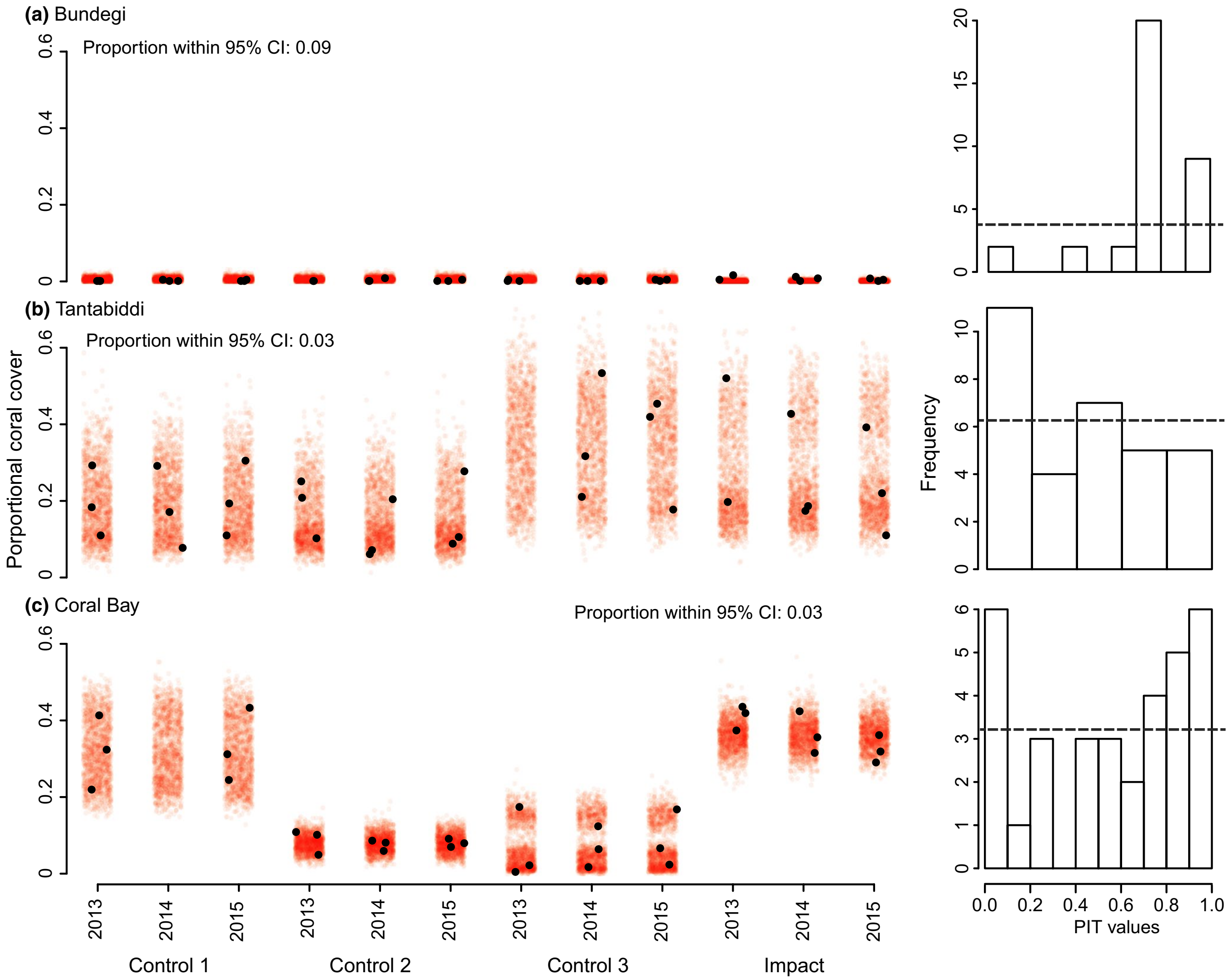

FIGURE 3 Posterior predictive checks of model fits for the occurrence of hard coral at three representative sites, based on the three years of sampling (2013, 2014 and 2015). Left-hand plots show the raw data for the three fixed transects (solid black circles) overlayed on top of their 500 simulated data points (transparent red circles). Right hand plots show histograms of the probability integral transform (PIT) values provided by INLA. The horizontal dashed line indicates the bar height expected if the histogram were a uniform distribution

a 60% decline in coral cover, with the exception of the extremely low coral cover sites in the Bundegi sector (Bundegi, Figure 4; Appendix S2, BUN, Figure S2.3), along with Murat (also in the Bundegi sector) which showed extremely high within site variance (Appendix S2, Figure S2.2). Coral Bay in the Southern sector showed the highest power, with ~80% power to detect a 30% decline, at least with two or three 'after' surveys (Figure 4).

We also explored how power changed with the number of fixed transects (1–6) as well as the number of scored points (100, 300 and 500) on a given transect, for an impact effect size of 30% (of the mean before state) decline for three selected sites representing the range of observed outcomes in terms of statistical power (high power, medium power and low power, according to the existing design, Appendix S2, Figure S2.3). Power of 80% to detect a 30% decline in live coral cover could not be achieved at the nominal 'low' powered site, even when 6 (up from 3) fixed transects and 500 points (up from 300) were used, suggesting that at some locations high power to detect change in the context of a press disturbance will be difficult to achieve (Figure 4). In contrast, for the Coral Bay 'high' power site, 80% power for a 30% decline could potentially be achieved with only two fixed transects, providing at least 300 points are analysed on each transect (Figure 4). Between five and six transects would be required if only 100 points per transect were analysed. Clearly, two transects with more points analysed will be more cost-effective, as points are analysed in the lab and are therefore cheaper to increase than transects, which add substantially more work in the field. Beyond three replicate transects, power to detect a 30% decline increased relatively systematically at site Tantabiddi, with an increase in both the number of replicate transects, as well as with the number of points, although the gain in power from 300 to 500 points was relatively diminished compare to that between 100 and 300

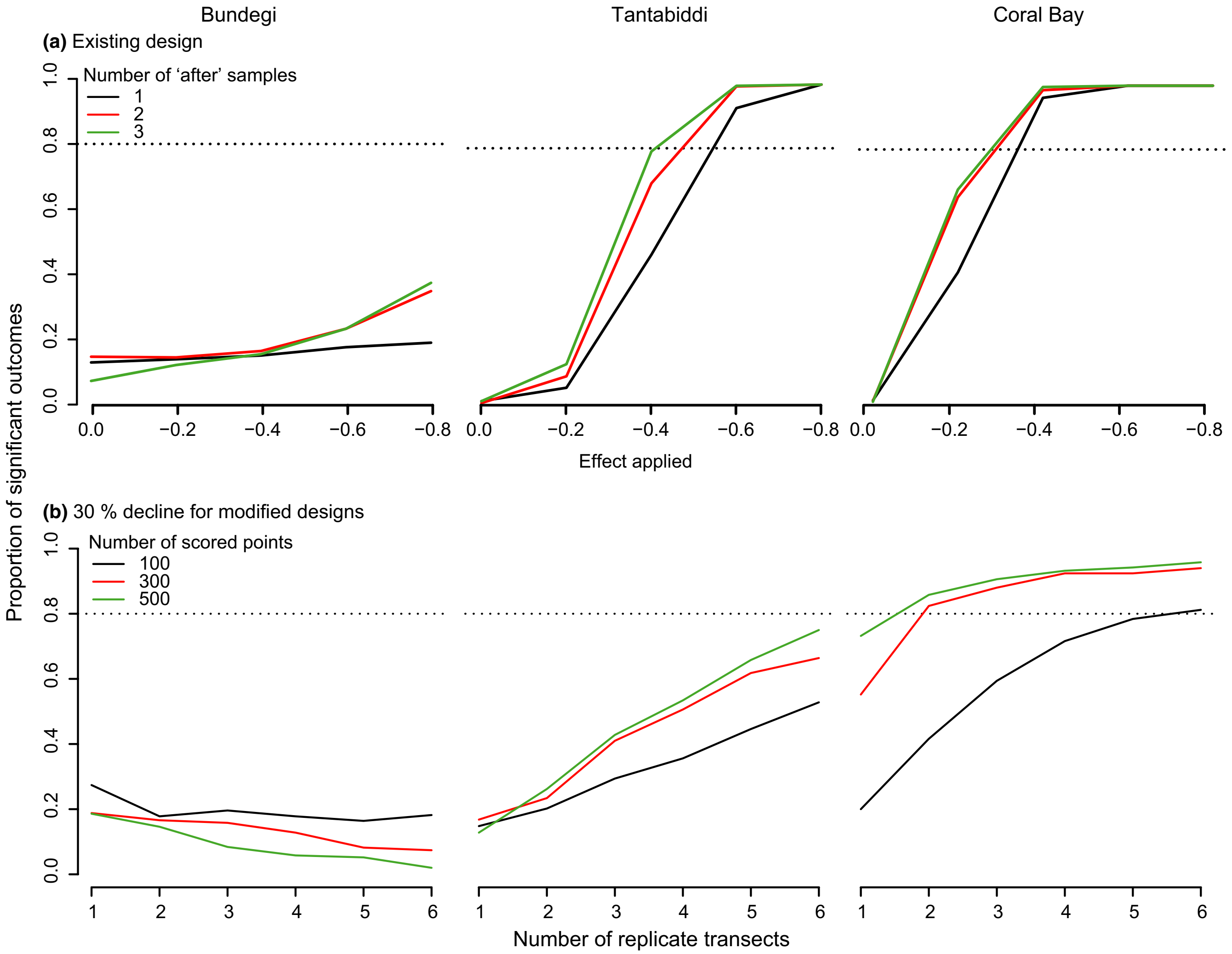

FIGURE 4 Statistical power to detect declines in the occurrence of hard coral at three representative sites across the Ningaloo marine reserves, based on the three years of sampling (2013, 2014 and 2015) for the existing design (a) with 1, 2 and 3 hypothetical 'after' surveys; and across a range of designs (b) for two 'after' surveys and a 30% decline. Power at each site was examined as individual potential 'impact' sites, with the three most similar (in terms of standard deviation of coral cover) used as 'controls'. The dashed horizontal line indicates 80% power

(Figure 4). Tantabiddi represents a site with 'moderate' power, and is perhaps representative of the bulk of the sites examined within the Ningaloo marine reserves (Appendix S2, Figure S2.3). For such sites it appears that high power (e.g. 80%) to detect a 30% decline in live coral cover can probably be achieved, but with a substantial increase in sampling effort (i.e. >6 replicate transects and 300–500 points; Figure 4).

5 | DISCUSSION

The *epower* package represents an easy to use, robust solution for assessing statistical power for existing hierarchical BACI sampling designs, and for exploring theoretical power by increasing or decreasing the level of replication at any point in the design. The package allows users to effectively allocate resources while balancing the risk of Type I and II errors. The *epower* package is of practical value to regulators wishing to ensure environmental impact assessments have adequate power to detect meaningful change, and to proponents wishing to achieve acceptable statistical power, while minimising the risk of committing a Type I error.

Our case study demonstrated that the power to detect meaningful change in live coral cover at sites across the Ningaloo Marine Park varied substantially across local scales. Consistent with previously published expectations for beyond BACI designs (Underwood, 1994), power was exceptionally low for sites with very low coral cover, and for sites with high temporal variability. Additionally, the case study demonstrated (in a practical sense) that power could be improved by increasing the level of replication at the transect and point-count levels, without adjusting the effect size. In this example at site Tantabiddi (a 'typical' site), power to detect a 30% decline in coral cover increased from ~40% under the existing design (three fixed transects with 300 analysed points each) to nearly 80% for a design with 6 fixed transects and 500 analysed points each; demonstrating the package's utility as a decision-making tool and to determine trade-offs between the cost of a program versus its ability to detect meaningful change (and therefore add value).

Power to detect a change in the context of a BACI design will differ to the power to detect a trend in time across all the sites for a given study region. For the long-term monitoring program associated with the Ningaloo marine reserves in Western Australia, further analysis would be required to assess power within a broader ecosystem context (Kang, McGree, Drovandi, Caley, & Mengersen, 2016) such that the overall information gathered by the design is maximized, e.g. (Falk, McGree, & Pettitt, 2014) before decisions around the optimal sampling design for this program could be made. However, in other settings, such as in the case of a pilot study for an environmental impact assessment survey, low power for a BACI interaction term may warrant a dramatic increase in levels of replication from that used in the existing pilot or 'before' data because committing a Type II error may result in irreversible environmental harm.

The *epower* toolbox adopts a Bayesian framework for inference, where two models are fit (one with, and one without the BACI interaction term) and evaluated based on the posterior model probability of one model over another. Using this approach, no significance level is set and there is no a priori defined '*alpha*'. For our case study, the rate of Type I error (indicated by the number of significant outcomes for zero effect) was generally very low suggesting that hierarchical BACI designs are relatively robust to issues of Type I error, an outcome that is expected given this was the motivation for their design (Glasby & Underwood, 1998; Stewart-Oaten et al., 1986; Terlizzi et al., 2005; Underwood, 1994).

All estimates of statistical power based on pilot data assume the pilot data adequately captures the real spatio-temporal variance in the system. For the Monte-Carlo simulation approach used here (and elsewhere, see Green & MacLeod, 2016) where variance components are estimated from mixed model fits of this pilot data, we make the additional assumption that the model used represents an adequate fit. In developing *epower* considerable effort was made to ensure models were correctly fitted. The *epower* package uses Bayesian statistical methods to incorporate uncertainty associated with hierarchical variance estimates, and these are appropriately captured in the simulations used to assess statistical power. While this comes at additional computational cost, the cost is outweighed by the ability of *epower* to capture this uncertainty, which cannot be accounted for using traditional methods. A broad range of statistical distributions can be specified, ensuring *epower* appropriately handles a range of common response variables and that simulation outputs are meaningful. However, care must be taken to ensure the fitted Bayesian mixed model is appropriate to the data and experimental design. We provide 'PIT' diagrams (Dawid, 1984) to aid in model validation, with a relatively even distribution across the range being desired. We encourage users to utilize this diagnostic tool, and where models clearly do not provide an adequate fit to the existing data, alternative modelling approaches should be explored.

The *epower* package assumes that temporal, sublocation and residual variance (when relevant) are similar across the replicate locations, and do not change between 'Before' and 'After' the impact; if they do change, then we advise caution. This may occur, for example, when control locations are not representative of impact locations, because they (a) have higher or lower initial abundances of target sensitive receptors, or (b) there are large differences in the variability of receptors. We mitigated this issue by using the three most recent years as our before data (to avoid large deviations due to temporal changes), and the three most similar sites (in terms of temporal and spatial standard deviations) for each 'impact' site to be assessed.

While *epower* models spatial and temporal autocorrelation associated with repeated sampling at the same locations (via random effects) this does not account for spatial or temporal serial correlation, such as when there are underlying trends through time or across the spatial domain of the study. Serial correlation may bias estimates of temporal variance and alternative modelling approaches should be used when this is evident.

Potential future updates to the package include the addition of other distributions as these are made available within INLA; the capacity to update the number of simulations sequentially, which would greatly improve the use of the package for large datasets; the capacity to allow testing of a 'slope' effect (i.e. a temporal trajectory or distance of impact effect) instead of just the BACI interaction term, and a cost-optimization algorithm. However, even without such improvements, *epower* currently represents a useful addition to the ecological and environmental practitioner's toolkit where analysis of a BACI design is intended.

ACKNOWLEDGEMENTS

C Schwarz provided valuable advice on the development of statistical testing procedures for BACI designs and Shaun Wilson provided valuable guidance in the analysis and interpretation of the case study. All data collection for the case study was carried out by staff from the Department of Biodiversity, Conservation and Attractions' Marine Science Program and Exmouth Region. Patricia Menendez provided valuable comments on the draft manuscript.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

AUTHORS' CONTRIBUTIONS

R.F. developed the formulation of the BACI design, encoded the toolbox, analysed the test case data and led the writing of the manuscript. G.S. conceived the project, obtained funding, contributed to the interface design. J.M.M. provided advice on implementation of the toolbox using INLA, the use of Bayes factors for statistical hypothesis testing (model selection) and the use of a posterior sample as the basis for Monte-Carlo simulation. R.J.S. restructured the code as the '*epower* package', and added parallelization.

K.I. and G.S. aided with testing of the toolbox. T.H.H. and G.R.S. collected the data, provided context for the case studies and aided the case study analysis. All authors contributed to draft revisions.

DATA AVAILABILITY STATEMENT

The *epower* package is freely available to users under licence, subject to the users agreement with BMT's Terms of use. Code and example datasets are available on Github at https://github.com/bmtglobal/epower or Zenodo at https://doi.org/10.5281/zenodo.3368039 (Fisher et al., 2019).

ORCID

Rebecca Fisher https://orcid.org/0000-0001-5148-6731

REFERENCES

Aguado-Giménez, F., Piedecausa, M. A., Gutiérrez, J. M., García-Charton, J. A., Belmonte, A., & García-García, B. (2012). Benthic recovery after fish farming cessation: A 'beyond-BACI' approach. *Marine Pollution Bulletin*, *64*, 729–738. https://doi.org/10.1016/j.marpolbul.2012.02.012

Anderson, M. J. (2001). Permutation tests for univariate or multivariate analysis of variance and regression. *Canadian Journal of Fisheries and Aquatic Sciences*, *58*, 626–639. https://doi.org/10.1139/f01-004

Anderson, M. J., Gorley, R. N., & Clarke, K. R. (2008). *Permanova+ for PRIMER*. Plymouth, UK: Primer-E.

Audet, J., Elsgaard, L., Kjaergaard, C., Larsen, S. E., & Hoffmann, C. C. (2013). Greenhouse gas emissions from a Danish riparian wetland before and after restoration. *Ecological Engineering*, *57*, 170–182. https://doi.org/10.1016/j.ecoleng.2013.04.021

Bates, D. M. (2010). LME4*: Mixed-effects modeling with* R*. New York: Springer.*

Benedetti-Cecchi, L. (2001). Beyond BACI: optimization of environmental sampling designs through monitoring and simulation. *Ecological Applications*, *11*, 783–799. https://doi.org/10.1890/1051-0761(2001)011[0783:BBOOES]2.0.CO;2

Blangiardo, M., Cameletti, M., Baio, G., & Rue, H. (2013). Spatial and spatio-temporal models with R-INLA. *Spatial and Spatio-temporal Epidemiology*, *4*, 33–49. https://doi.org/10.1016/j.sste.2012.12.001

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H., & White, J. S. (2009). Generalized linear mixed models: A practical guide for ecology and evolution. *Trends in Ecology & Evolution*, *24*, 127–135. https://doi.org/10.1016/j.tree.2008.10.008

Clark, G. F., Knott, N. A., Miller, B. M., Kelaher, B. P., Coleman, M. A., Ushiama, S., & Johnston, E. L. (2018). First large-scale ecological impact study of desalination outfall reveals trade-offs in effects of hypersalinity and hydrodynamics. *Water Research*, *145*, 757–768. https://doi.org/10.1016/j.watres.2018.08.071

Claudet, J., Lenfant, P., & Schrimm, M. (2010). Snorkelers impact on fish communities and algae in a temperate marine protected area. *Biodiversity and Conservation*, *19*, 1649–1658. https://doi.org/10.1007/s10531-010-9794-0

Coleman, M. A., Palmer-Brodie, A., & Kelaher, B. P. (2013). Conservation benefits of a network of marine reserves and partially protected areas. *Biological Conservation*, *167*, 257–264. https://doi.org/10.1016/j.biocon.2013.08.033

Constable, A. J. (1991). The role of science in environmental-protection. *Australian Journal of Marine and Freshwater Research*, *42*, 527–538. https://doi.org/10.1071/MF9910527

Dawid, A. P. (1984). Present position and potential developments: Some personal views: Statistical theory: The prequential approach. *Journal of the Royal Statistical Society. Series A (General)*, *147*, 278–292. https://doi.org/10.2307/2981683

Falk, M. G., McGree, J. M., & Pettitt, A. N. (2014). Sampling designs on stream networks using the pseudo-Bayesian approach. *Environmental and Ecological Statistics*, *21*, 751–773. https://doi.org/10.1007/s10651-014-0279-2

Fisher, R., Glenn, S. R., Sadler, R. J., Inostroza, K., Shedrawi, G., Holmes, T. H., & McGree, J. M. (2019). R code and example datasets showing use of *epower* - software for the analysis of BACI designs (Version v1.3). *Methods in ecology and evolution. Zenodo*. https://doi.org/10.5281/zenodo.3368039

Foroughirad, V., & Mann, J. (2013). Long-term impacts of fish provisioning on the behavior and survival of wild bottlenose dolphins. *Biological Conservation*, *160*, 242–249. https://doi.org/10.1016/j.biocon.2013.01.001

Glasby, T. M., & Underwood, A. J. (1998). Determining positions for control locations in environmental studies of estuarine marinas. *Marine Ecology Progress Series*, *171*, 1–14. https://doi.org/10.3354/meps171001

Grear, D. A., Luong, L. T., & Hudson, P. J. (2012). Sex-biased transmission of a complex life-cycle parasite: Why males matter. *Oikos*, *121*, 1446–1453. https://doi.org/10.1111/j.1600-0706.2012.20358.x

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*, 493–498. https://doi.org/10.1111/2041-210x.12504

Green, R. H. (1979). *Sampling design and statistical methods for environmental biologists*. New York, NY: Wiley.

Holsman, K. K., McDonald, P. S., Barreyro, P. A., & Armstrong, D. A. (2010). Restoration through eradication? Removal of an invasive bioengineer restores some habitat function for a native predator. *Ecological Applications*, *20*, 2249–2262. https://doi.org/10.1890/09-1168.1

Johnson, P. C., Barry, S. J., Ferguson, H. M., & Muller, P. (2015). Power analysis for generalized linear mixed models in ecology and evolution. *Methods in Ecology and Evolution*, *6*, 133–142. https://doi.org/10.1111/2041-210X.12306

Kang, S. Y., McGree, J. M., Drovandi, C. C., Caley, M. J., & Mengersen, K. L. (2016). Bayesian adaptive design: Improving the effectiveness of monitoring of the Great Barrier Reef. *Ecological Applications*, *26*, 2635–2646. https://doi.org/10.1002/eap.1409

Keough, M. J., & Mapstone, B. D. (1997). Designing environmental monitoring for pulp mills in Australia. *Water Science and Technology*, *35*, 397–404. https://doi.org/10.2166/wst.1997.0567

Lunde, K. B., Resh, V. H., & Johnson, P. T. J. (2012). Using an ecosystem-level manipulation to understand host-parasite interactions and how they vary with study venue. *Ecosphere*, *3*, art84. https://doi.org/10.1890/ES12-00001.1

MacKay, D. J. C. (2003). *Information theory, inference and learning algorithms*. Cambridge: Cambridge University Press.

Neyman, J., & Pearson, E. S. (1967). *Joint statistical papers*. Berkeley, CA: University of California Press.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Development Core Team. (2013). NLME: Linear and nonlinear mixed effects models. R package version 3.1-109.

Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society Series B-Statistical Methodology*, *71*, 319–392. https://doi.org/10.1111/j.1467-9868.2008.00700.x

Smith, M. P. L. (1991). Environmental-impact assessment – The roles of predicting and monitoring the extent of impacts. *Australian Journal of Marine and Freshwater Research*, *42*, 603–614. https://doi.org/10.1071/MF9910603

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

305–339. https://doi.org/10.1890/0012-9615(2001)071[0305:TASVIE]2.0.CO;2
Stewart-Oaten, A. W., Murdoch, W. W., & Parker, K. R. (1986). Environmental Impact Assessment: 'psuedoreplication' in time? *Ecology and Society*, *67*, 929–940. https://doi.org/10.2307/1939815
Terlizzi, A., Benedetti-Cecchi, L., Bevilacqua, S., Fraschetti, S., Guidetti, P., & Anderson, M. J. (2005). Multivariate and univariate asymmetrical analyses in environmental impact assessment: A case study of Mediterranean subtidal sessile assemblages. *Marine Ecology Progress Series*, *289*, 27–42. https://doi.org/10.3354/meps289027
Underwood, A. J. (1991). Beyond BACI – Experimental-designs for detecting human environmental impacts on temporal variations in natural-populations. *Australian Journal of Marine and Freshwater Research*, *42*, 569–587. https://doi.org/10.1071/MF9910569
Underwood, A. J. (1993). The mechanics of spatially replicated sampling programs to detect environmental impacts in a variable world. *Australian Journal of Ecology*, *18*, 99–116. https://doi.org/10.1111/j.1442-9993.1993.tb00437.x
Underwood, A. J. (1994). On beyond BACI – Sampling designs that might reliably detect environmental disturbances. *Ecological Applications*, *4*, 3–15. https://doi.org/10.2307/1942110
Underwood, A. J., & Chapman, M. G. (2003). Power, precaution, Type II error and sampling design in assessment of environmental impacts. *Journal of Experimental Marine Biology and Ecology*, *296*, 49–70. https://doi.org/10.1016/S0022-0981(03)00304-6

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Fisher R, Shiell GR, Sadler RJ, et al. *epower*: An R package for power analysis of Before-After-Control-Impact (BACI) designs. *Methods Ecol Evol*. 2019;10:1843–1853. https://doi.org/10.1111/2041-210X.13287