

REVIEW

Effectively integrating experiments into conservation practice

Nancy Ockendon¹  | Tatsuya Amano^{2,3} | Marc Cadotte⁴  | Harriet Downey⁵  |
Mark H. Hancock^{6,7} | Ann Thornton⁵ | Paul Tinsley-Marshall⁸  |
William J. Sutherland^{5,9} 

¹ Endangered Landscape Programme, Cambridge Conservation Initiative, Cambridge, UK

² School of Biological Sciences, University of Queensland, Brisbane, Queensland, Australia

³ Centre for Biodiversity and Conservation Science, University of Queensland, Brisbane, Queensland, Australia

⁴ Department of Biological Sciences, University of Toronto Scarborough, Toronto, Ontario, Canada

⁵ Conservation Science Group, Department of Zoology, University of Cambridge, Cambridge, UK

⁶ Cairngorms Connect, Achantoul, Aviemore, UK

⁷ RSPB Centre for Conservation Science, RSPB, Inverness, UK

⁸ Kent Wildlife Trust, Tyland Barn, Sandling, Maidstone, Kent, UK

⁹ Biosecurity Research Initiative at St Catharine's (BioRISC), St Catharine's College, University of Cambridge, Cambridge, UK

Correspondence

Nancy Ockendon, Endangered Landscape Programme, Cambridge Conservation Initiative, Cambridge, CB2 3QZ, UK.

Email: n.ockendon@jbs.cam.ac.uk

[Correction added on 22 June 2021, after first online publication: Conflict of Interest statement has been added.]

Funding information

Arcadia; Australian Research Council, Grant/Award Number: FT180100354; MAVA; The David and Claudia Harding Foundation

Handling Editor: Max Lambert

Abstract

1. Making effective decisions in conservation requires a broad and robust evidence base describing the likely outcomes of potential actions to draw on. Such evidence is typically generated from experiments or trials that evaluate the effectiveness of actions, but for many actions evidence is missing or incomplete.
2. We discuss how evidence can be generated by incorporating experiments into conservation practice. This is likely to be most efficient if opportunities for carrying out informative, well-designed experiments are identified at an early stage during conservation management planning.
3. We consider how to navigate a way between the stringent requirements of statistical textbooks and the complexities of carrying out ecological experiments in the real world by considering practical approaches to the key issues of replication, controls and randomization.
4. We suggest that routinely sharing the results of experiments could increase both the value for money and effectiveness of conservation practice.
5. We argue that with early planning and a small additional input of effort, important new learning can be gained during the implementation of many conservation actions.

KEYWORDS

controls, experiment, evidence, experimental design, management trial, manipulation, nature conservation, replication

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Ecological Solutions and Evidence* published by John Wiley & Sons Ltd on behalf of British Ecological Society

1 | INTRODUCTION

Despite an increasing appreciation of the importance of using evidence in conservation planning and policy, many actions lack a sufficient evidence base (Christie et al., 2021). Indeed, many routinely implemented interventions have no evidence for their effectiveness (Sutherland et al., 2020). Decisions taken on the basis of incomplete or inaccurate information can lead to inefficient use of the limited resources available and poorer outcomes. Routinely testing management interventions, by integrating well-designed experiments into conservation practice, could significantly increase the evidence available for decision makers, improving the effectiveness of many conservation actions and increasing value for money (Feinsinger, 2001; Cadotte et al., 2020).

However, designing experiments or management trials that yield useful results and are feasible to implement as part of conservation practice can be challenging. Although most ecology undergraduate students are taught experimental design, these courses frequently focus on ideal scenarios, with large numbers of uniform experimental units (such as petri dishes or field plots) where treatments and controls can be replicated, randomized or stratified with ease. While important, such principles are difficult to apply in complex natural ecological systems. This leads to a divergence in terms of experimental design and data analysis between the expectations of academic researchers and the realities for practitioners working on the ground (Feinsinger, 2001; Cadotte et al., 2017). The minimum standards often recommended as necessary for robust statistical inference can appear unachievable within conservation projects. Our experience, including editing and reviewing articles from conservation practice submitted to the *Conservation Evidence Journal* (AT, NO, TA and WJS) and *Ecological Solutions and Evidence* (MC) and as conservationists faced with the challenge of learning from practice (MHH and PT-M), is that there are many opportunities to produce useful evidence in conservation practice which could be realized with small changes to intervention design (e.g. Douglas et al., 2019).

In this article, we consider some of the issues associated with generating new knowledge in conservation practice and discuss the neglected question of when 'non-ideal' experiments (e.g. with small numbers of replicates or without untreated controls) are worth implementing. We propose a series of steps to identify opportunities to include trials that will yield useful results and discuss practical approaches to issues such as replication, randomization and the need for controls. Our intention is to be inclusive, and we have therefore not assumed too much prior statistical knowledge, in the hope of reaching a wide audience, particularly those working in conservation on the ground. We believe that there are significant opportunities to carry out simple manipulative experiments that, for a modest additional input of effort, will yield results that can both inform ongoing adaptive management and improve practice in the wider conservation community.

2 | PLANNING EXPERIMENTS IN THE REAL WORLD: IDENTIFYING OPPORTUNITIES

In order to generate evidence in conservation practice, the first challenge is to identify opportunities where an experimental component can be informatively and efficiently integrated into management (Figure 1). Such opportunities have three key requirements: (i) an action is being undertaken, and a better understanding of the effectiveness of this action would make a difference to conservation practice (i.e. the effectiveness of the action is uncertain and the costs of this uncertainty are significant); (ii) the skills necessary to design, carry out and analyse the results of the experiment are available; and (iii) a well-designed experiment can be included in existing workplans relatively easily (e.g. an action is being repeated many times allowing different treatments to be compared; staff capacity is available to monitor the outcome). These components may vary, both through time and across actions, so opportunities for including experimental tests should be regularly reviewed.

3 | PLANNING EXPERIMENTS IN THE REAL WORLD: STAGES IN THE PROCESS

Once an opportunity to answer a question has been identified, the next challenge is to design an effective experiment to address it, given the resources available. The questions presented in Table 1, along with the worked example in Box 1, describe the different stages of this process to be considered by, for example, a site manager, conservation officer or reserve team. Although the stages are presented as a list, the planning process may not be linear and some stages may require iterative feedback and adaptation (e.g. if a selected treatment becomes unfeasible or site conditions change). Considering all these stages before work gets underway on a project is likely to produce results that are more informative than if existing monitoring is retrospectively reframed in an experimental format.

4 | CHALLENGES OF EXPERIMENTAL DESIGN

There are a range of issues to consider when designing an experiment to test the effectiveness of a conservation action (stage 6 in Table 1). Many of these come down to distinguishing the effects caused by the action from the natural variation in conditions that exists, both through time and within and between sites and individuals. Suppose in the example in Box 1, the different sign treatments are rotated on a weekly basis. It is possible that one of the weeks coincides with a period of especially pleasant weather, with effects on visitor groups, the number of dogs being walked and their tendency to be let off the lead. Any difference observed between treatments could then be due to differences in behaviour between groups in response to the weather, rather than a response to the different sign treatments. To use another example, if, following a change to the grazing intensity in a field, the density

FIGURE 1 Diagram to identify when an experiment can be usefully and efficiently included in conservation practice. The optimal conditions for carrying out an experiment arise when the need for results, the availability of opportunities and the necessary skills all coincide.

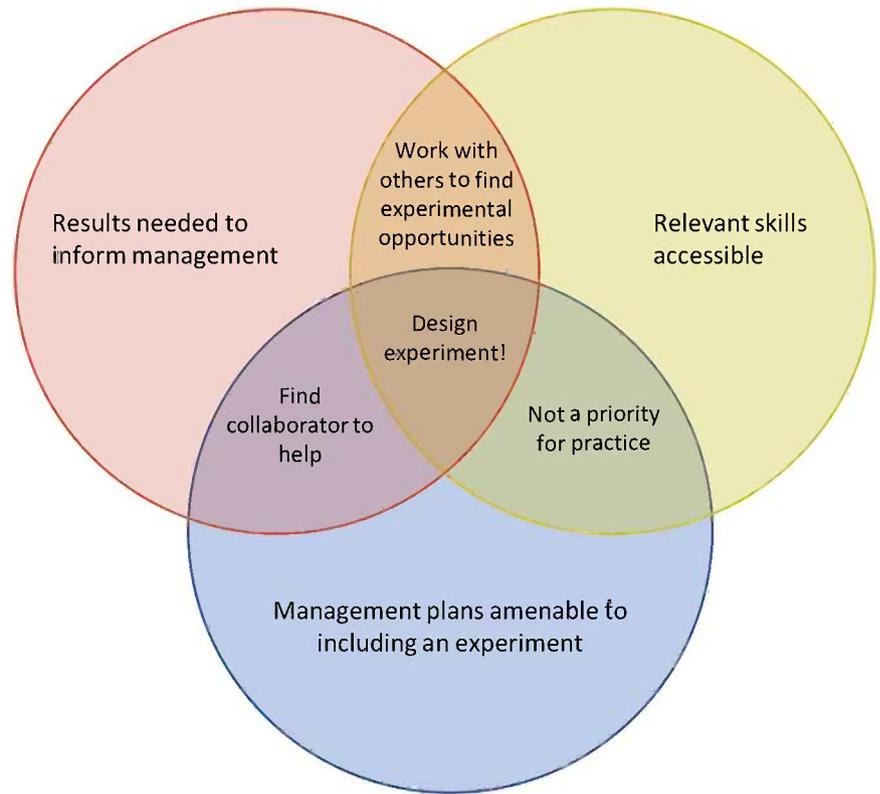


TABLE 1 Ten questions to consider in the process of identifying a conservation management question and designing an experiment to answer it

| Stage | Description |
|--|---|
| 1. What is the main target of the action? | What is the conservation importance of the areas where an experiment is being considered (e.g. species or habitat)? |
| 2. What are the main problems? | Identify the issue(s) of concern. |
| 3. What new knowledge is most needed? | What knowledge is needed but currently missing? What evidence would be most useful? |
| 4. What are the broad options? | What are the possible responses to dealing with the problem? Which of these options is most important, and also feasible, to investigate? |
| 5. What are the different variants? | Are there different ways of carrying out the chosen option, for example at different times of year or using different techniques? |
| 6. What are the sensible experimental design options? | Considering the various constraints, what experimental designs can be implemented? |
| 7. What is feasible to measure? | What can be monitored and is sufficiently important, abundant and quantifiable to generate useful results? What skills and capacity are available amongst staff and volunteers? |
| 8. What are the options for analysing the results? | What statistical approaches are available and appropriate for analysing the data collected? Who analyses the data? |
| 9. Is it worth carrying out the chosen experimental design? | What is the balance between the need to answer the question and the resources it will take to carry out the experiment? |
| 10. How will the results be documented and shared? Where will this plan be reported? | What is a realistic plan for publishing or otherwise sharing the results? Where will this plan be stated? |

of orchids increases, this could well be a result of the altered grazing regime (e.g. Hutchings, 2010). However, it could also be due to another factor – perhaps the rabbit population has decreased or, unbeknown to the researcher, the farmer has changed fertilizer regime. The three cor-

nerstones of experimental design – replication, control and randomization – are used to reduce the chance that these ‘background’ variables are overwhelmingly influencing the outcome and increase the certainty that any observed changes can be ascribed to the action taken.

Box 1. A worked example to identify a question and plan an experiment (Table 1): reducing disturbance of ground-nesting birds by dogs. Numbers in parentheses refer to the stages given in Table 1

A site is important for its population of ground-nesting birds (1). However, a long-distance footpath cuts across one part of the site and dogs off the lead belonging to walkers on this path are thought to be disturbing the nesting birds (2). The reserve team would like to know how to change the behaviour of walkers to reduce disturbance by their dogs (3).

Although other options, such as closing the footpath or prohibiting dogs, were considered the team decides the most practical option is to put up a sign near the path (4). The team are not sure whether a sign emphasising the emotional or the conservation motive for keeping dogs on leads would be most effective (5). Therefore, they decide to include three different treatments in the trial: no sign, a sign asking walkers to keep dogs on leads to reduce disturbance to birds, and a sign with a photo of a dog with a dead wader chick and an emotional appeal. The signs can be changed by the livestock manager on the first visit to the site each day. It would take a day to make the signs and create a schedule and a few minutes each day to switch signs (6).

In terms of monitoring, staff carrying out livestock management tasks at the same time each day can observe people with dogs who have passed the signs and document whether their dogs are on the lead. This would take about 10 minutes each day. Results can be recorded on data cards (7). The assistant warden, who is looking for new challenges and helped plan the experiment, will oversee the analysis (she suggests a chi-squared test). Time will be allocated for this (8).

Once the experimental set-up has been decided, the team considers whether it is worth carrying out the experiment. The fact that the footpath is long distance provides an opportunity, as most walkers only pass the reserve once (the handful of known local regular dog walkers will not be counted more than once). Therefore, the data from each person can be considered independent, facilitating statistical analysis; if the same walkers used the path each day, this experiment might not be worth implementing. This experiment is estimated to take about three person-days in total, carried out in short periods that will not interfere with normal management. The results are expected to take two days to analyse and a week to write up, but this can be done on wet winter days. This seems worthwhile to understand whether a sign can reduce disturbance by dogs (9).

Finally, the team discuss how the results will be shared. The reserve manager would like to write up their findings and the assistant warden published a paper after her MSc and is keen to collaborate. Together they will be given time to write up the experiment and submit it to the open access journal *Ecological Solutions and Evidence*, even if the study does not detect a change in behaviour. This is stated in the organisation's annual plan (10).

4.1 | Replication

If a result is based on observations from a single site or population, there is a serious concern that the result could be due to peculiarities of that site or covariance between the variable of interest and other unknown variables (such as the rabbit population or fertilizer regime in the above example). Implementing an action at more than one independent site, individual or population is therefore usually recommended in order to be confident that the observed results are caused by the action, if replicates respond consistently.

The basic principle is 'the more replicates the better', because applying a treatment to a larger number of independent units increases the accuracy of the results and hence the chance of detecting any effects of the treatment, particularly if these are small (Christie et al., 2019). However, in conservation there is often a practical limit to the number of times a treatment can be independently replicated (How many wetlands can be restored? How many islands can an invasive species be eradicated from?). There are also costs, in terms of time, money and effort, associated with increasing the number of replicates. Therefore, although more replicates will provide more accurate information, this must be balanced against the practicalities of carrying out the experiment.

4.1.1 | Deciding on a realistic number of replicates

The number of replicates needed to detect a specified effect size can be calculated using a power analysis (Lehr, 1992; Crawley, 2015). This can be useful in cases where the system is well understood, as an estimate of the natural variation present is needed to calculate the number of replicates required to detect a specified effect size. In essence, the smaller the effect size we would like to detect, or the greater the variability that exists between replicate units, the more replication is needed. However, frequently this information is not accurately known. In addition, it can be difficult to specify a target effect size – do we wish to increase a population by a specific amount, at a specific rate, or simply increase it? In such cases, the number of experimental replicates can be guided by the type of action and what is practically possible, the likely magnitude of the effect (compared to the natural variation present), and the body of evidence that already exists (Table 2).

It is important to remember that the experimental replicates chosen will affect how widely applicable the inferences from a study may be. If all replicates come from the same field or forest (e.g. grazers are excluded from three different plots within a meadow or deadwood is left standing in five plots in a woodland) then the findings can only confidently be applied to that one area. However, if a large number of

TABLE 2 A guide to an appropriate number of experimental replicates, depending on the effect size we need to detect (relative to the natural variation), practical constraints (e.g. costs, number of sites available, scale of intervention) and the existing body of knowledge about the action. Cell colours indicate where the number of replicates is likely (green, yellow) or unlikely (purple, blue) to yield informative results. Note that single replicate studies should be avoided whenever possible (but see below) and we advise clearly stating the justifications and limitations if single or few replicates are adopted in a study

| Effect size of action relative to the natural variation in the untreated condition | Number of replicates | | |
|--|---|---|--------------------------------------|
| | Single replicate | A few replicates ^a | Many replicates ^b |
| Small e.g. effect of supplementary food on the survival of chicks, effect of removing cuttings to deplete soil nutrients in grasslands. | Unlikely to be informative alone, unless combined with other studies | May provide some information as the basis for further studies | Necessary in order to be informative |
| Moderate e.g. effect of introducing livestock grazing on plant diversity in a meadow, effect of excluding predators on chick survival of ground-nesting birds. | May provide some information as the basis for further studies | Acceptable, if practical constraints limit number of replicates | Informative, adopt if practical |
| Large e.g. effect of ring barking an invasive tree, effect of herbicide on an invasive plant. | Acceptable, particularly if action is novel and practical constraints limit more replicates | Acceptable, any additional replicates will increase the power | Very informative, adopt if practical |

^aApproximately two to five independent replicate units.

^bA minimum of six independent replicate units, but ideally more.

independent replicates of a treatment are carried out across a landscape or region, the findings will have a much wider scope of inference across that landscape or region, and perhaps, with appropriate caveats, more widely.

4.1.2 | Is it ever useful to carry out an experiment with a single replicate?

Although an experiment with $n = 1$ (as described in the left hand column of Table 2) is considered heresy by many scientists, in reality conservationists often discuss the lessons learned from management at a single site, whether it be change in water quality after the introduction of European beavers *Castor fiber* (e.g. Puttock et al., 2017), the effect of reintroducing wolves *Canis lupus* on the vegetation structure in Yellowstone Park, USA (Mao et al., 2005) or unplanned ‘natural experiments’ that provide opportunities to generate new results that would be impossible otherwise, such as the ecological responses to the eruption of Mount St Helens in Washington, USA (Swanson & Crisafulli, 2005). These are all examples of before-and-after comparisons at a single site (see below), where the conservation community accepts the results as useful despite the lack of replication. Although the results might have higher levels of uncertainty and be less generalizable than those from replicated studies, they elucidate processes that would be impossible to test with more formal experiments. Their findings are often discussed in the context of long-term study of the system, so interpretation is aided by an understanding of the underlying mechanism and the previous range of variation.

In addition, even though a single unreplicated experiment may have weak statistical power and low certainty about the direction or magnitude of any effect, the results may be usefully combined with other similar studies, for example by meta-analysis. An analysis of

the effects of eradicating invasive mammals on seabirds across 61 oceanic islands (where each individual study had $n = 1$) found that most seabird populations had increased, with a mean annual recovery rate of 1.12 (Brooke et al., 2018). Here, the results of an intervention that would be very difficult for a single author or group to replicate were combined across a large number of individual studies to produce an estimate of the average impact of the action. This was only possible because the various single-island studies had been properly documented and shared in the published literature. Large-scale habitat restoration is another intervention that is difficult to replicate (Davies & Gray, 2015), but meta-analysis of such projects has been able to quantify the average increase in biodiversity and ecosystem services provided (Rey Benayas et al., 2009). Similarly, the impact of protected areas can normally only be studied at one or a few sites, but meta-analysis can generalize from these smaller studies to measure overall protected area effectiveness much more accurately (Coetzee et al., 2014).

In general, larger organizations are likely to have more opportunities and capacity to carry out replicated large-scale experiments. For example, the RSPB, a large UK conservation organization that owns nature reserves and has its own science staff, has developed novel conservation interventions via long-term trials on its own reserves, co-designed by scientists and practitioners (e.g. Hancock et al., 2009; Malpas et al., 2013). However, if a large-scale intervention is being implemented by a small organization that manages only one site, one option may be to collaborate with other organizations to collect comparable data across multiple sites. This reflects a wider need to think creatively to adapt experimental approaches to organizational and project constraints (illustrated by the top overlap in Figure 1). Governments and major funders also have an important role in facilitating the coordination of intervention testing across organizations, especially for large-scale actions that they support financially.

4.1.3 | Reducing sample variance

The more variable the natural conditions between experimental units are, the more replicates are needed (Table 2). Conversely, reducing the sample variance will reduce the number of replicates required. Sample variance is affected by both the natural variation that exists between replicate units and also by measurement errors. There are several ways by which variance can be reduced, allowing any effects of an action to be more easily detected. Firstly, using replicate units that are as similar to each other as possible in every aspect will reduce variability. After that, improving the rigour of sampling will reduce measurement error. This can be achieved by ensuring that the methodology is consistent across replicates, appropriate equipment is used (e.g. measuring length using calipers rather than measuring tape) and training is provided to those collecting data. Increasing the sampling within replicates, either by taking more samples or larger, more extensive, samples can also increase the accuracy of estimates within each replicate. For example, if, in an experiment to look at the effects of excluding deer on forest structure, the diameter of three trees is measured at each experimental and control site, the accuracy of the estimate of the true average diameter at each site would be low, because the between-individual variance in diameter is likely to be substantial. Therefore, the ability to detect any effect of the deer fencing is also low. Taking measurements from 20 trees at each site would be a relatively straightforward way to obtain a more accurate estimate of the true average diameter. The standard error of the estimated mean diameter would decline as the number of trees increases – this can be investigated (e.g. by plotting standard error against sample size) to find a compromise level of sampling, which reduces the error in the estimate of the mean to an acceptable level.

Another way to address the effects of sample variance is to measure any covariates that might be expected to confound the results and account for these statistically. In the deer fencing example above, if the distance to the forest edge varies across the experimental sites and is likely to affect the trees' diameter, this could be recorded for each plot and accounted for before testing for any effect of fencing. Note that as more variables are added to the analysis, more replicates are needed to allow any relationships to be investigated with the same power.

4.1.4 | Sampling within replicate units: Avoiding pseudoreplication

As discussed above, increasing the number of samples within each experimental unit reduces the sample variance. However, it is important to avoid pseudoreplication: treating multiple samples taken from the same replicate unit as independent samples (Hurlbert, 1984). In the dog-walking example (Table 1), if several dogs being walked by the same person were treated as independent data points this would be pseudoreplication, as the decision to let each of these dogs off the lead would not be taken independently. Such mistakes incorrectly inflate the number of replicates in the study, invalidating statistical conclusions. To avoid this, repeated measurements from the same replicate must either be averaged and treated as a single sample or analysed

using appropriate statistical tests (such as generalized linear mixed models with sampling unit as a random factor). Therefore, although increasing the number of samples within a replicate and increasing the number of treatment replicates can both improve the reliability of the results of an experiment, it is vital that the true number of independent replicates of both treatments and controls is stated clearly and the data are analysed and interpreted appropriately (Davies & Gray, 2015).

4.1.5 | Selecting response variables with large expected effect size

Choosing to measure a variable that is likely to show a bigger change in response to the action could also make an experiment more informative for a given sample size (Table 2). Although the magnitude of the effect of a specific action on a particular variable cannot be altered, variables with a shorter chain of links to the action being taken are likely to show stronger responses. For example, if an education programme aims to reduce the number of herbivores killed by snares, the effect on people's behaviour (e.g. the frequency of trips to set snares) is likely to be greater than any resultant impact in mammal numbers (e.g. abundance of deer species), which are influenced by multiple factors and may take years to respond noticeably. However, this would need to be balanced with the risk that changes in surrogate variables may not respond in the same way as the ultimate target variable. For example, if snare setting declined only in low deer density areas, then the number of animals killed might not decline correspondingly.

4.2 | Controls and comparisons

Including a comparison for an experimental treatment allows any observed changes in the target of an action to be attributed to that action, rather than other (known or unknown) variables. The conventional approach in ecology is to compare treated and untreated control units (a control-impact design). For example, the abundance of fish in a marine protected area may be compared with a similar unprotected area and the protected area is found to contain more fish (e.g. Rakitin & Kramer, 1996). However, without replicates or other comparisons it is unclear whether this result was due to the designated protection, the fact that the site with more fish was selected for protection in the first place, or other factors that differ between the two sites.

Another commonly used option is a before-and-after treatment comparison at a site (before-after design), for example comparing fish abundance in the years before and after the designation of a protected area. However, these results could also be confounded, for example if the climatic conditions or disturbance by tourists changed through time. Therefore, the favoured design is to combine both approaches in a before-after-control-impact (BACI) design, where both treatment and control sites are monitored before and after the treatment (see Examples 2 and 3 in Box 2). Such experiments have been shown to be more accurate than other study designs in detecting the true effect of an intervention (Christie et al., 2019).

Box 2: Examples of real-world experiments in conservation

1. Does modifying the internal design of swift nest boxes increase occupancy? One example of a simple replicated controlled trial was carried out by Action for Swifts (Newell 2019). The study investigated whether the addition of an artificial, molded 'form' into nest boxes for swifts *Apus apus* affected the occupancy rate. It had a very simple set up, consisting of 142 nest boxes across four sites where nest boxes for swifts were already present. Nest forms were allocated to alternate nest boxes at each of the sites before the breeding season began and occupancy was checked at the end of subsequent breeding seasons. Across all four sites there was a significant association between the presence of molded forms and nest box occupancy. This study demonstrated how replication, stratification and controls can be applied to generate new evidence which can inform future design of nest boxes.

2. Does prescribed burning increase native tree regeneration? A large-scale management trial was carried out to test the effect of burning on the rate of regeneration of Scots pine *Pinus sylvestris* in Scotland (Hancock et al. 2009). A randomised, controlled experimental set up was repeated at 10 sites across a nature reserve; at each site two 100 m² plots were burned and two plots left unburned. The number of new tree seedlings was monitored for the next five years and found to be ten times higher in burned areas than unburned areas. This result led to a new programme of prescribed burning management at the site and ultimately an increase in pine regeneration. The experiment also included a manipulation of deer browsing effects (using fencing to exclude deer from half of plots at each site: one burnt and one unburnt) as a second treatment factor. Deer exclusion had only a relatively minor influence on tree establishment, supporting the use of burning without a need to markedly change the deer management regime.

3. Does removal of predatory snails reduce predation pressure on threatened corals and how much effort is required? A replicated, before-after-control-impact study in Florida, USA, tested the effectiveness of two different approaches for removing coral-eating snails *Coralliophila abbreviata* from threatened coral *Acropora palmata* where they were causing significant tissue loss (Williams et al. 2014). Twelve long-term monitoring plots (each 150 m²) across six reefs were assigned to one of three treatments: (1) snails removed by hand from *A. palmata* only, (2) snails removed by hand from all coral species, or (3) no snails removed (control). The baseline snail abundance at all plots was established before the experiment began. Divers took approximately 30 minutes to remove snails from a plot in treatment 1 and 51 minutes in treatment 2. The abundance of snails and the number of feeding scars were reduced in both removal treatments compared to the control but there was no difference between the two treatments. Given there was no difference in outcome between the two removal treatments, this experiment revealed that resources could be more effectively targeted by removing snails from *A. palmata* only.

In some cases, comparing several different variants of an action can be more useful than a comparison of treatment versus no treatment (Smith et al., 2014). Different experimental units can be subject to small modifications in aspects such as the frequency, timing, intensity or application of the action (see Example 3 in Box 2). For example, if an invasive plant is rapidly spreading across a protected area and evidence suggests that herbicide is likely to be effective against the species, then an experiment could more usefully compare whether spraying early in the season is more effective than spraying late, rather than simply comparing areas with and without herbicide (e.g. Marushia et al., 2010). This could help inform how a given management budget could be most effectively allocated to deliver a desired outcome. Comparing variants can also be useful in circumstances where controls are not possible, and immediate action needs to be taken across all experimental replicates, for example if leaving an untreated control site could allow the spread of the invasive plant.

4.3 | Allocating treatments

Once the units to be used in an experiment have been identified, the next step is to decide how to allocate the treatment(s) across the units. Doing this carefully reduces the chance that the treatment effects are confounded by any background variation that exists (Johnson, 2002).

Conventional scientific thinking is that the best approach is usually to randomly allocate treatment and controls across the experimental units. This avoids problems of selective bias that might occur if, for example, the experimenter assigns the treatment to the plots nearest the road or the first visitors to arrive on the reserve in the morning. Randomization works well when no prior knowledge about patterns of background variability exists and sample sizes are large, as the characteristics of the treatment and control groups are likely to be similar overall. However, randomization of management interventions across sites can be hard for site staff to accept, as they are used to selecting appropriate management approaches for particular circumstances. The key argument is that randomization of treatment and control across a number of sites – any of which might be suitable for the management being investigated – is the most effective way to measure the impact of that management.

Alternatively, if background variation is known to exist, for example if plant species composition changes along an altitudinal gradient or visitors to a nature reserve that have travelled further tend to arrive later, stratification of replicates is often recommended. Stratification attempts to achieve similar representation in both the experimental and control sites. The experimental replicates are split into, for example, areas of similar altitude or groups of visitors that arrive in the same hour-long period, and an equal number of control and treatment units are then allocated randomly within each group. Similarly, paired

studies use pairs of matched sites with similar conditions and randomly allocate a treatment and control site within each pair.

When the number of experimental replicates is small, as is often the case in conservation, randomization can give clustered results, and it may be worth considering using a regular allocation of treatment and control units (i.e. spacing on a grid system or timing with a fixed interval). A regular allocation of treatments can be a good way of reducing the impact of unknown factors, for example when alternating nest box designs (see Example 1 in Box 2) or providing information to visitors and looking at the effect on their behaviour. Another advantage of using regular allocation is that units will be as widely spaced as possible, which is likely to increase their degree of statistical independence.

4.4 | Dealing with multiple or combined actions

Frequently in conservation, a set of actions is carried out together and the overall response or outcome is monitored. For example, a landscape restoration project or a package of agri-environment measures are likely to comprise multiple actions whose combined impact is evaluated (e.g. Perkins et al., 2011; Jiao et al., 2012). This demonstrates the overall outcome of the project but provides little evidence about the efficacy of the individual actions. Therefore, when including an experiment in a project consisting of a package of measures, it is often useful to break it down into smaller components and investigate one of these actions at a time. For example, a small experimental component of a large forest restoration project could test whether adding mulch around saplings after planting increased their survival until the following year. This will improve effectiveness in the future (knowing if the effort of mulching new trees is worthwhile) with minimal costs and risk to the overall project.

Alternatively, if multiple actions are often carried out together in practice or their effects are likely to interact, it might be more informative to carry out an experiment with a factorial design. The most commonly undertaken factorial design is a 2x2 study in which two actions are tested simultaneously to investigate their single and combined effects. For example, does a combination of cutting and spraying an invasive plant have a greater impact than if the actions were carried out independently (e.g. Miller & Lucero, 2014)? A factorial experiment would look at the effects of all four treatment combinations (cutting alone, spraying alone, both cutting and spraying, and neither cutting nor spraying). Similarly, a 2x2 factorial design was used to test the single and combined effects of prescribed burning and deer exclusion on Scots pine seedling establishment in the Highlands of Scotland (Example 2, Box 2).

5 | SHARING RESULTS REGARDLESS OF OUTCOME

We recommend that a plan of how the results will be made available to others, whether via a scientific paper, report, website or data repository, is stated clearly at the beginning of a project (Table 1, Question

10). This ensures that writing up and sharing results are allocated sufficient resources and are not contingent on the results obtained. The selective publication of conservation studies with positive results has led to problematic biases in the evidence base, with 'successful' studies being much more likely to be written up and published in scientific journals (e.g. Parker et al., 2016; Catalano et al., 2019). Such biases are often not intentionally deceitful; a common cause of publication bias is that non-significant results are considered less interesting and are consequently harder to get published, especially in major journals, leading authors to abandon attempting publication (Csada et al., 1996). However, if an experiment has been well designed, it will yield useful results regardless of whether the action did or did not result in the expected outcome.

A publication plan was devised early on during a heathland restoration project being undertaken by Kent Wildlife Trust. An opportunity to include an experimental trial of bracken control methods was identified, as it was recognized that evidence for the effectiveness of bracken control interventions is limited (Martin et al., 2020). Therefore, rather than implement a single intervention based on inadequate evidence, a BACI experimental design is being used to test three bracken control treatments: (1) cutting, (2) bruising and (3) cutting and scarifying. At the outset, the trial design was discussed with the editor of the *Conservation Evidence Journal* to ensure it met appropriate standards for publication. Data collection will be carried out by staff and volunteers, and the results written up by staff.

6 | THE WAY FORWARD

We believe that there are significant opportunities to increase the generation of useful evidence by including experiments in conservation practice. In this paper, we discuss some of the key requirements to achieving this, namely looking for opportunities to include experimental components when planning management actions, clearly identifying the question to be answered, planning the experimental design and committing to publishing the results regardless of outcome (Berend et al., 2019). Of the many actions undertaken during conservation management, a proportion is likely to be amenable to testing with manipulative experiments to yield informative results. Identifying these opportunities could substantially increase the learning generated and, along with improved collaboration between researchers and practitioners, help to negotiate the research-implementation space (Toomey et al., 2017). Many of these trials are unlikely to be a priority for academic researchers, whether because of their perceived lack of novelty or academic impact, or their sheer quantity. However, practitioners are ideally placed to both identify the questions that, if answered, will make the most difference to the effectiveness and efficiency of conservation management, and to quickly implement the findings (as shown in Box 2).

Although there are many occasions when new evidence can be produced by evaluating conservation management interventions, barriers do exist to experimental tests becoming routine. The major challenge is a lack of time and money to design, implement and write up

experiments as part of conservation management, with resources rarely allocated to these activities (Burbidge et al., 2011). However, a recent survey found that conservation funders see value in including experiments in the projects that they fund, and that 1–3% of a project budget was, on average, considered an appropriate amount to invest in experiments (Tinsley-Marshall et al. unpublished). This indicates that conservationists can feel confident that many funders will look favourably on project proposals containing an experimental component.

In addition, experiments offer the potential to effectively identify what does and does not work, which can create huge savings in resources in the future and repay the short-term investment many times over. As well as using results directly in future adaptive management, there are also advantages to communicating findings externally. If a conservation organization tests one action each year, it will enhance the future effectiveness of that action across its sites. However, if 20 organizations each tested an intervention annually and shared their results, then the conservation community as a whole would soon be much better informed and more effective, with each organization being able to draw on the collective experience and results.

Another barrier may exist if organizations frequently find that a lack of expertise is hindering the effective planning, implementation and analysis of potential experiments (Figure 1). In such cases, solutions include investing in employing, training, contracting or collaborating with scientific staff with these skills to increase the capacity of the organization to carry out more experiments, generate knowledge and improve practice.

We have tried to untangle some of the knottier aspects of experimental design and discussed how to balance the theoretical requirements of statistics with the realities of what can be achieved in the field with the available time and resources. One particularly challenging issue is the number of experimental replicates needed, where it makes sense to maximize as far as practical the number of replicates. However, there may still be value in undertaking an experiment using a very small number of replicates, or even only one, if a question is of high enough importance, the effect is likely to be large and the practicalities of repeating the treatment render it unfeasible. As Hurlbert (2004) stated: 'In the last analysis, every proposed experiment must be judged by its own objectives, design, possibilities, and costs. There should be no automatic rejection of experiments where no treatment replication is proposed'. In the wake of such unreplicated but potentially innovative studies, we also would like to stress the importance of 'replication studies', where results that have been reported based on a small number of replicates are tested again, in order to increase the size and breadth of the evidence base (Johnson, 2002; Biology Staff Editors, 2018).

It is not always appropriate to include an experimental component in a project, and it is important to weigh up the costs of carrying out the experiment against the knowledge that will be gained. New knowledge and insight can be generated from approaches other than manipulative experiments, such as case studies or correlational analyses (e.g. Kaul & Wilsey, 2020). Results of these studies can provide the basis for subsequent experimental tests (e.g. by providing the starting hypothesis) and also directly inform practice. Case studies often improve under-

standing of the impact of a set of actions taken together in a particular context. Correlational studies are useful when a number of results describing the same relationship already exist and can be compared or used to estimate an effect size, but may suffer from issues related to other unknown biases. However, we believe that well-designed experiments are generally the easiest way to demonstrate the effectiveness of a specific intervention.

Publishing non-significant results and/or those based on small sample sizes in established peer-reviewed journals can be challenging. We believe that editors and reviewers of conservation journals are becoming more sensitive to the realities of generating new evidence in the field as well as demonstrating a commitment to overcoming publication bias. If experiments are properly designed, negative or non-significant results should not be cause for a study to be rejected and we strongly support the increasing number of practitioner-focused journals (including *Conservation Science and Practice*, *Ecological Solutions and Evidence* and *Conservation Evidence Journal*) that emphasize the value of sharing useful, well-designed conservation experiences and trials regardless of novelty or statistical significance of the results (Cadotte et al., 2020). The Applied Ecology Resources platform hosted by the British Ecological Society (<https://www.britishecologicalsociety.org/applied-ecology-resources>) provides a platform for practitioners to publish their results, in the form of reports and case studies, whether or not a statistically significant effect was found. Although other barriers that reduce the accessibility of publication to non-academics remain, including publication costs for open access journals and the peer-review process, we believe that publishers and practitioners can work together to explore ways to overcome these (e.g. Cadotte et al., 2020).

In this article, we hope to have demonstrated that, with a little forward planning, useful evidence can often be generated when implementing conservation interventions, with experimentation included as an integral part of management. An increased emphasis on the importance of testing the efficacy of actions and publishing the results could lead to a step change in the breadth and depth of evidence available to everyone working in conservation.

ACKNOWLEDGEMENTS

WJS, NO and AT are funded by Arcadia, a charitable fund of Lisbet Rausing and Peter Baldwin, NO via the Endangered Landscapes Programme, who also part-fund MH, WJS by MAVA, and The David and Claudia Harding Foundation. TA is supported by the Australian Research Council Future Fellowship (FT180100354) and the University of Queensland strategic funding. Thanks to Julia Jones and Jeremy Wilson for comments on previous drafts of this manuscript.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHORS' CONTRIBUTIONS

NO and WJS conceived the idea. NO led the writing of the manuscript, and all authors contributed to the draft and gave approval for publication.

DATA AVAILABILITY STATEMENT

This manuscript does not include any data.

PEER REVIEW

The peer review history for this article is available at <https://publons.com/publon/10.1002/2688-8319.12069>.

ORCID

Nancy Ockendon  <https://orcid.org/0000-0001-5055-1079>

Marc Cadotte  <https://orcid.org/0000-0002-5816-7693>

Harriet Downey  <https://orcid.org/0000-0003-1976-6973>

Paul Tinsley-Marshall  <https://orcid.org/0000-0002-6142-4072>

William J. Sutherland  <https://orcid.org/0000-0002-6498-0437>

REFERENCES

- Berend, K., Haynes, K., and MacKenzie, C. M. (2019). Common garden experiments as a dynamic tool for ecological studies of alpine plants and communities in northeastern North America. *Rhodora*, 121F, 174–212.
- Biology Staff Editors, P. (2018). The importance of being second. *PLoS Biology*, 16, e2005203
- Brooke, M. D. L., Bonnaud, E., Dilley, B. J., Flint, E. N., Holmes, N. D., Jones, H. P., Provost, P., Rocamora, G., Ryan, P. G., Surman, C., & Buxton, R. T. (2018). Seabird population changes following mammal eradications on islands. *Animal Conservation*, 21, 3–12.
- Burbidge, A. H., Maron, M., Clarke, M. F., Baker, J., Oliver, D. L., & Ford, G. (2011). Linking science and practice in ecological research and management: How can we do it better? *Ecological Management & Restoration*, 12, 54–60.
- Cadotte, M. W., Barlow, J., Nuñez, M. A., Pettorelli, N., & Stephens, P. A. (2017). Solving environmental problems in the Anthropocene: the need to bring novel theoretical advances into the applied ecology fold. *Journal of Applied Ecology*, 54, 1–6.
- Cadotte, M. W., Jones, H. P., & Newton, E. L. (2020). Making the applied research that practitioners need and want accessible. *Ecological Solutions and Evidence*, 1, e12000.
- Catalano, A. S., Lyons-White, J., Mills, M. M., & Knight, A. T. (2019). Learning from published project failures in conservation. *Biological Conservation*, 238, 108223.
- Christie, A. P., Amano, T., Martin, P. A., Shackelford, G. E., Simmons, B. I. & Sutherland, W. J. (2019). Simple study designs in ecology produce inaccurate estimates of biodiversity responses. *Journal of Applied Ecology*, 56, 2742–2754.
- Christie, A. P., Amano, T., Martin, P. A., Petrovan, S. O., Shackelford, G. E., Simmons, B. I., Smith, R. K., Williams, D. R., Wordley, C. F. & Sutherland, W. J. (2021). The challenge of biased evidence in conservation. *Conservation Biology*, 35, 249–262.
- Coetzee, B. W., Gaston, K. J. & Chown, S. L. (2014) Local scale comparisons of biodiversity as a test for global protected area ecological performance: A meta-analysis. *PLoS One*, 9, e105824.
- Crawley, M. J. (2015) *Statistics: An introduction using R* (2nd ed.). Wiley.
- Csada, R. D., James, P. C. & Espie, R. H. (1996) The “file drawer problem” of non-significant results: does it apply to biological research? *Oikos*, 76, 591–593.
- Davies, G. M. & Gray, A. (2015) Don't let spurious accusations of pseudoreplication limit our ability to learn from natural experiments (and other messy kinds of ecological monitoring). *Ecology and Evolution*, 5, 5295–5304.
- Douglas, D. J. T., Jones, P. S., Crosher, I., Diack, I. & Littlewood, N. (2019) *Monitoring biodiversity responses to peatland restoration*. IUCN UK Peatland Programme Report. <https://www.iucn-uk-peatlandprogramme.org/sites/default/files/2019-11/COIMON~1.PDF>
- Feinsinger, P. (2001) *Designing field studies for biodiversity conservation*. Island Press.
- Hancock, M. H., Summers, R. W., Amphlett, A. & Willi, J. (2009) Testing prescribed fire as a tool to promote Scots pine *Pinus sylvestris* regeneration. *European Journal of Forest Research*, 128, 319.
- Hurlbert, S. H. (1984) Pseudoreplication and the design of ecological field experiments. *Ecological Monographs*, 54, 187–211.
- Hurlbert, S. H. (2004) On misinterpretations of pseudoreplication and related matters: A reply to Oksanen. *Oikos*, 104, 591–597.
- Hutchings, M. J. (2010) The population biology of the early spider orchid *Ophrys sphegodes* Mill. III. Demography over three decades. *Journal of Ecology*, 98, 867–878.
- Jiao, J., Zhang, Z., Bai, W., Jia, Y., & Wang, N. (2012). Assessing the ecological success of restoration by afforestation on the Chinese loess plateau. *Restoration Ecology*, 20, 240–249.
- Johnson, D. (2002) the importance of replication in wildlife research. *The Journal of Wildlife Management*, 66, 919–932.
- Kaul, A. D., & Wilsey, B. J. (2020). Exotic species drive patterns of plant species diversity in 93 restored tallgrass prairies. *Ecological Applications*, 31, e2252.
- Lehr, R. (1992) Sixteen S-squared over D-squared: A relation for crude sample size estimates. *Statistics in Medicine*, 11, 1099–1102.
- Malpas, L. R., Kennerley, R. J., Hiron, G. J. M., Sheldon, R. D., Ausden, M., Gilbert, J. C. & Smart, J. (2013) The use of predator-exclusion fencing as a management tool improves the breeding success of waders on lowland wet grassland. *Journal for Nature Conservation*, 21, 37–47.
- Mao, J. S., Boyce, M. S., Smith, D. W., Singer, F. J., Vales, D. J., Vore, J. M. & Merrill, E. H. (2005) Habitat selection by elk before and after wolf reintroduction in Yellowstone National Park. *The Journal of Wildlife Management*, 69, 1691–1707.
- Martin, P. A., Rocha, R., Smith, R. K. & Sutherland, W. J. (2020) Shrubland and heathland conservation. In W. J. Sutherland, L. V. Dicks, S. O. Petrovan, & R. K. Smith (Eds), *What works in conservation 2020* (pp. 483–525). Open Book.
- Marushia, R. G., Cadotte, M. W. & Holt, J. S. (2010) Phenology as a basis for management of exotic annual plants in desert invasions. *Journal of Applied Ecology*, 47, 1290–1299.
- Miller T. W., & Lucero, C. (2014) Meadow knapweed (*Centaurea debeauxii*) response to herbicides and mechanical control. *Invasive Plant Science and Management*, 7, 503–510.
- Newell, D. (2019) A test of the use of artificial nest forms in common swift *Apus apus* nest boxes in southern England. *Conservation Evidence*, 16, 24–26.
- Parker, T. H., Forstmeier, W., Koricheva, J., Fidler, F., Hadfield, J. D., Chee, Y. E., Kelly, C. D., Gurevitch, J. & Nakagawa, S. (2016) Transparency in ecology and evolution: Real problems, real solutions. *Trends in Ecology & Evolution*, 31, 711–719.
- Perkins, A. J., Maggs, H. E., Watson, A. & Wilson, J. D. (2011) Adaptive management and targeting of agri-environment schemes does benefit biodiversity: a case study of the corn bunting *Emberiza calandra*. *Journal of Applied Ecology*, 48, 514–522.
- Puttock, A., Graham, H. A., Cunliffe, A. M., Elliott, M. & Brazier, R. E. (2017) Eurasian beaver activity increases water storage, attenuates flow and mitigates diffuse pollution from intensively-managed grasslands. *Science of the Total Environment*, 576, 430–443.
- Rakitin A., & Kramer, D. L. (1996) Effect of a marine reserve on the distribution of coral reef fishes in Barbados. *Marine Ecology Progress Series*, 131, 97–113.
- Rey Benayas, J. M., Newton, A. C., Diaz, A. & Bullock, J. M. (2009) Enhancement of biodiversity and ecosystem services by ecological restoration: A meta-analysis. *Science*, 325, 1121–1124.
- Smith, R. K., Dicks, L. V., Mitchell, R. & Sutherland, W. J. (2014) Comparative effectiveness research: The missing link in conservation. *Conservation Evidence*, 11, 2–6.
- Sutherland, W. J., Dicks, L. V., & Petrovan, S. O. (2020). *What works in conservation 2020*. Open Book.

- Dale, V. H., Swanson, F. J., & Crisafulli, C. M. (Eds.). (2005) *Ecological responses to the 1980 eruption of Mount St. Helens*. Springer.
- Tinsley-Marshall P. et al. (In review) Improving conservation effectiveness through the routine testing of management interventions.
- Toomey, A. H., Knight, A. T. & Barlow, J. (2017) Navigating the space between research and implementation in conservation. *Conservation Letters*, 10, 619–625.
- Williams, D. E., Miller, M. W., Bright, A. J. & Cameron, C. M. (2014) Removal of corallivorous snails as a proactive tool for the conservation of acroporid corals. *PeerJ*, 2, <https://doi.org/10.7717/peerj.680>

How to cite this article: Ockendon, N., Amano, T., Cadotte, M., Downey, H., Hancock, M. H., Thornton, A., Tinsley-Marshall, P., & Sutherland, W. J. (2021). Effectively integrating experiments into conservation practice. *Ecol Solut Evidence*, 2, e12069. <https://doi.org/10.1002/2688-8319.12069>