

How experimental biology and ecology can support evidence-based decision-making in conservation: avoiding pitfalls and enabling application

Steven J. Cooke^{1,2,*}, Kim Birnie-Gauvin^{1,2}, Robert J. Lennox², Jessica J. Taylor^{1,2}, Trina Rytwinski^{1,2}, Jodie L. Rummer³, Craig E. Franklin⁴, Joseph R. Bennett¹ and Neal R. Haddaway⁵

¹Canadian Centre for Evidence-Based Conservation and Environmental Management, Department of Biology and Institute of Environmental Science, Carleton University, 1125 Colonel By Drive, Ottawa, Ontario, Canada K1S 5B6

²Fish Ecology and Conservation Physiology Laboratory, Department of Biology, Carleton University, 1125 Colonel By Drive, Ottawa, Ontario, Canada K1S 5B6

³Australian Research Council (ARC) Centre of Excellence for Coral Reef Studies, James Cook University, Townsville, QLD 4811, Australia

⁴School of Biological Sciences, The University of Queensland, Brisbane, QLD 4072, Australia

⁵EviEM, Stockholm Environment Institute, Box 24218, 10451 Stockholm, Sweden

*Corresponding author: Email: Steven_Cooke@carleton.ca

Policy development and management decisions should be based upon the best available evidence. In recent years, approaches to evidence synthesis, originating in the medical realm (such as systematic reviews), have been applied to conservation to promote evidence-based conservation and environmental management. Systematic reviews involve a critical appraisal of evidence, but studies that lack the necessary rigour (e.g. experimental, technical and analytical aspects) to justify their conclusions are typically excluded from systematic reviews or down-weighted in terms of their influence. One of the strengths of conservation physiology is the reliance on experimental approaches that help to more clearly establish cause-and-effect relationships. Indeed, experimental biology and ecology have much to offer in terms of building the evidence base that is needed to inform policy and management options related to pressing issues such as enacting endangered species recovery plans or evaluating the effectiveness of conservation interventions. Here, we identify a number of pitfalls that can prevent experimental findings from being relevant to conservation or would lead to their exclusion or down-weighting during critical appraisal in a systematic review. We conclude that conservation physiology is well positioned to support evidence-based conservation, provided that experimental designs are robust and that conservation physiologists understand the nuances associated with informing decision-making processes so that they can be more relevant.

Key words: Evidence, experimental biology and ecology, relevance, reliability

Editor: Gary Burness

Received 14 December 2016; Revised 21 June 2017; Editorial Decision 23 June 2017; accepted 15 July 2017

Cite as: Cooke SJ, Birnie-Gauvin K, Lennox RJ, Taylor JJ, Rytwinski T, Rummer JL, Franklin CE, Bennett JR, Haddaway NR (2017) How experimental biology and ecology can support evidence-based decision-making in conservation: avoiding pitfalls and enabling application. *Conserv Physiol* 5(1): cox043; doi:10.1093/conphys/cox043.

Introduction

Humans have dominated the planet for much of the last century (Vitousek *et al.*, 1997), leading to massive environmental change, unprecedented population declines that have triggered the need to designate many organisms as Threatened or Endangered (Mace *et al.*, 2008), and loss of biodiversity (Cardinale *et al.*, 2012; Pimm *et al.*, 2014) to the point now that the current Epoch is referred to as ‘the Anthropocene’ (Crutzen, 2006). The financial resources and government capacity required to make this a ‘good Anthropocene’ (Dalby, 2016) are, however, scarce. We depend on decision-makers to ensure their actions are informed by scientific evidence; yet, it is apparent that environmental practitioners’ decisions are most heavily influenced by past experience, tradition or input from co-workers (Pullin *et al.*, 2004). Practitioners are simply failing to make use of the full body of scientific knowledge, and this can lead to management decisions that fail to achieve the desired outcome, waste precious resources, and/or time (e.g. imperiled populations; Lindenmayer *et al.*, 2013). This series of circumstances has resulted in vast repositories of information being neglected and/or overlooked entirely by decision-makers and practitioners (Cook *et al.*, 2013). The reasons behind this so-called ‘knowledge-action’ gap are numerous and complex (Cook *et al.*, 2013). Literature on the sociology of knowledge reveals that the ways in which practitioners and decision makers consume knowledge and assess knowledge claims is very different than those that typically generate new knowledge (i.e. scientists). Accordingly, scientists and other claimants need to make conscious decisions about whose expectations they hope to meet in their scientific communications and engagement activities and plan their scientific activities accordingly if their aim is to ‘be relevant’ (Young *et al.*, 2016).

Taking the lead from the evidence-based approach that has revolutionized the health sciences (i.e. Cochrane Reviews; see <http://www.cochranelibrary.com/cochrane-database-of-systematic-reviews/>) and led to standardized practices in the evaluation of evidence, the field of conservation and environmental management is undergoing a paradigm shift toward procedures that provide the critical syntheses needed to properly inform policy and management (Sutherland *et al.*, 2004). One of the largest barriers to evidence-based environmental management is the disconnect between science and policy (Pullin and Knight, 2012). Policy-makers desire a high degree of certainty, which is highly unlikely in science, especially if one views a single empirical study rather than integrating all available evidence. Although traditional approaches to reviewing existing evidence may be familiar to many decision-makers and researchers, such methods are highly susceptible to a suite of biases, including selection bias and publication bias (Sutton *et al.* 2000). Systematic review methods (Pullin and Stewart, 2006; The Cochrane Collaboration, 2011; Collaboration for Environmental Evidence, 2013; Campbell Collaboration, 2016) incorporate steps to reduce the risk of error and bias and have been specifically developed to address the importance of context and the risks of bias in less reliable review methods.

Systematic review methods thus aim to ensure access to the best available evidence, yielding more-efficient and less-biased platforms for decision-making than informal reviews (Pullin and Knight, 2009a, b).

Systematic reviews depend on the availability of empirical studies to draw upon as part of the evidence synthesis process. Although scientific studies in the realm of biology and ecology can take many forms, those based on rigorous experimentation are considered to be particularly valuable. In the purest sense, an experiment is a procedure used to verify, refute or validate a hypothesis. What is particularly powerful about rigorous and well-designed experiments is their ability to identify cause-and-effect relationships and response thresholds through careful systematic manipulation of various factors of interest, while controlling against the intrusion of other factors that might otherwise confound the results and interpretation. Experiments can occur in the laboratory or field environment. Other tenets of experimental biology include transparent and repeatable methods, consideration of bias and appropriate use of statistical analyses in a hypothesis-testing framework. Relative to other types of studies (e.g. observational and correlative), experiments provide more certainty with respect to effects. In the context of systematic reviews, a well-designed and executed study is an essential component of evidence synthesis because of its ability to control and describe confounders and sources of heterogeneity. That is, ideally one would assemble a series of independent studies and, often using meta-analytical approaches, aggregate studies to seek more general trends across the broader literature base. Of course, not all experiments are of the highest validity (for a variety of reasons). Therefore, another important part of systematic review is a detailed, critical appraisal of the internal validity (quality) and external validity (generalizability) of each study to determine its utility in the evidence synthesis process.

As systematic reviews become more commonplace in conservation science and environmental management, it is worthwhile to reflect on the characteristics of study design that would be considered indicative of quality. Substantial effort has now been expended on a suite of systematic reviews relating to various conservation and environmental topics. Many of these topics involve studies with classical experimental designs, e.g. biomani-pulation for the treatment of eutrophication (Bernes *et al.* 2015). Across these reviews, thousands of studies and experimental designs have been appraised in an attempt to integrate reliable evidence into syntheses to aid decision-making. Our authorship team including those with experience as systematic reviewers and those that identify as experimental biologists, have observed a variety of study designs in experimental biology that can, through differential execution, facilitate or limit the utility of individual studies in syntheses. Here, we summarize some of the most common issues with experimental design that limit relevance to evidence synthesis and management decisions, and propose constructive solutions to improve the utility of experimental biology and ecology research in the increasingly popular field of evidence synthesis.

Increasingly, those working in the realms of biology (e.g. molecular biologists, physiologists and ethologists) and ecology (e.g. community, population and ecosystem ecologists) are conducting research related to understanding our natural world in an applied context. Such research is leading to the development of a number of nascent disciplines (e.g. conservation physiology; [Cooke *et al.*, 2013a](#), the intersection of conservation behaviour and physiology; [Cooke *et al.*, 2014a](#)). Even the most fundamental work can be used to support conservation and management activities, directly or indirectly; however, there may be a long lag (years or decades) between knowledge generation and its application or synthesis into synoptic analyses. To assist experimental biologists and ecologists with being ‘relevant’ to conservation and management practitioners and in supporting evidence-based conservation, we provide a perspective on how experimental biology and ecology underpins and promotes evidence-based decision-making in conservation. To do so, we first reflect on the process of evidence synthesis and critical appraisal of study design in systematic reviews. We then present and synthesize key aspects of study quality, referencing ways in which experimental biologists and ecologists can refine their experimental designs and methods to be more likely to be included in systematic reviews. We conclude with a forward-looking perspective on the potential for experimental biology and ecology to revolutionize conservation and management by creating the evidence base that is needed to make informed decisions. We want to be clear that we are not implying that the experimental biology community is doing ‘poor’ science. Rather, the needs and expectations of practitioners can be markedly different than what the ‘typical’ experimental biologist does. Therefore, to ‘be relevant’, it is necessary to consider science through the lens of the potential end users (e.g. stakeholders, managers and policy makers), which will extend the science beyond the scientific community.

Characteristics of experimental biology and ecology that facilitate evidence-based decisions in conservation

Cause and effect

Experimental biology and ecology are particularly relevant in the context of evidence-based research, given their focus on identifying causal relationships. The field often integrates experimental, descriptive and theoretical methods to answer study questions, but experimentation is the most powerful tool for determining causality ([Werner, 1998](#)). Experimental biology frequently yields quantitative data, which describe mechanisms and processes that can be understood as ‘causally interacting entities’ ([Craver, 2007](#); [Weber, 2012](#)). Researchers use model organisms or experimental systems ([Weber, 2012](#)) to determine causality by identifying appropriate predictor

variables to be manipulated, establishing proper controls, and neutralizing any confounding factors.

A significant benefit to the use of experimental biology in conservation and environmental management is its ability to isolate the effects of stressors (or ‘causes’) by performing experiments under controlled conditions, which allows scientists to separate the effects of a particular variable from other possible effects ([Seebacher and Franklin, 2012](#)). This requires a study with two situations: one in which the phenomenon under investigation occurs, and one in which it does not ([Weber, 2012](#)). Typically, only a single aspect of the study is altered to establish whether the variation in that aspect is attributed to changes in the response variable. This approach takes advantage of the regularity exhibited by biological mechanisms and assumes that these mechanisms will go from a start point to an end point whenever specific conditions permit ([Weber, 2004](#)). Studies of cause and effect are not without their challenges; [Nichols *et al.* \(2017\)](#) suggested that natural variability, multiple stressors, the difficulty of performing rigorous experiments, and the time and money required to undertake such studies collectively make cause and effect studies in field settings somewhat uncommon.

The benefits of increasing power to detect causality in experimental research are especially relevant to the critical appraisal process of systematic review. During critical appraisal, the analyst determines the level of confidence that can be placed on a dataset before incorporating it. In the medical field, a hierarchical approach was proposed to classify the value of data for inclusion in a systematic review ([Stevens and Milne, 1997](#)). Similarly, in conservation, the highest level of evidence comes from ‘at least one properly designed, randomized controlled trial of appropriate size’ ([Pullin and Knight, 2003](#)). Descriptive studies (e.g. observational and expert opinion) are assigned the lowest ranking in this hierarchy ([Concato, 2004](#)) and are likely to be excluded if they do not meet the quality standards set out in the systematic review protocol. Well-designed experiments are needed to generate the evidence necessary to support conservation.

Ability to explore complex interactions

Biological systems interact with one another in ways that are often difficult to predict ([Green and Sadedin, 2005](#)). Experimental biology is relevant to evidence-based science in part due to its ability to explore these complex interactions. The majority of experimental studies isolate and focus on the impacts of a single variable ([Blaustein and Kiesecker, 2002](#)). Results of these studies provide an essential foundation to understanding natural systems; however, they are often not representative of natural settings ([Christensen *et al.*, 2006](#)). For example, stressors do not often occur in isolation and can interact in complex and unexpected ways ([Christensen *et al.*, 2006](#); [Altshuler *et al.*, 2011](#)). Multiple stressors can be additive, synergistic or antagonistic ([Folt *et al.*, 1999](#)), making it difficult to formulate adequate unbiased hypotheses and design appropriate studies. Although it may not always

be possible to manipulate or isolate variables of interest, experimental biology allows the testing of the combined effects of such stressors, helping to illuminate the nature of these complex interactions (Di Santo, 2015). Although often-times challenging to conduct in field settings, experiments that reveal ecological interactions are especially relevant for conservation practitioners (Darling and Côté, 2008; Crain *et al.*, 2009).

It has been argued that the best way to approach the study of complex interactions is through the use of a multi-disciplinary approach (Altshuler *et al.*, 2011) as well as the combination of both holistic and reductionist approaches (Lidicker, 1988). By doing so, researchers integrate genetic, molecular, physiological, ecological and evolutionary approaches for a more complete perspective of the functioning of biological systems. Such approaches have been used to study amphibian population declines (Blaustein and Kiesecker, 2002; van Uitregt *et al.*, 2007), freshwater ecosystem threats (Altshuler *et al.*, 2011), human-induced stressors in marine systems (Crain *et al.*, 2009), and climatic effects on macroalgal recruitment (Lotze and Worm, 2002), among others. These studies have frequently demonstrated complex additive, antagonistic and synergistic effects that will have considerable impacts on the conservation of global ecosystems (Coors and De Meester, 2008). To add to this complexity, ecological carryover effects are also prescribed as crucial considerations in conservation biology. Carryover effects occur in any situation in which an individual's previous history and experience explains their current performance in a given situation (O'Connor *et al.*, 2014). Failure to assess carryover effects can lead to misguided conclusions and cause further damage to imperiled populations (O'Connor and Cooke, 2015; Ceccato *et al.*, 2016). Recognizing the potential for carryover effects and incorporating that concept into experimental design (e.g. monitoring over longer periods or different life-stages) is important to enhance relevance, particularly in our multiple-stressor world.

Evidence synthesis in systematic reviews

One of the core elements of a systematic review that differs from other forms of evidence synthesis is the emphasis on critical appraisal of study quality. Although one might assume that papers obtained from the primary peer-reviewed literature are of adequate quality, the reality is that contemporary peer review is not perfect (Smith, 2006). Indeed, if the focus is on where a paper ends up being published rather than if it gets published at all, the process can become distracted (Peres-Neto, 2016). This is particularly problematic, given the rise of predatory journals with little to no peer review (Bartholomew, 2014). In addition, it is generally agreed upon that peer-review itself is a game of chance. So, if a 'poor quality' article is submitted enough times, just by chance alone, it is likely to get published (Neff and Olden, 2006). For these reasons, there is dire need for critical appraisal of study internal validity and

exclusion of low quality results. Not only does a critique of study quality determine whether a paper may be included or excluded in synthesis, it also can form the basis for differential weighting of studies during meta-analysis (e.g. Detsky *et al.*, 1992). Quoting from the Collaboration for Environmental Evidence (CEE) guidelines 'study quality assessment requires a number of decisions about the absolute and relative importance of different sources of bias and data quality elements common to environmental data, particularly the appropriateness of temporal and spatial scales. It is therefore vital that the assessment process be standardized and as transparent and repeatable as possible. Quality is a relative term and its measurement and scale are very dependent on the question being addressed. It may be helpful to breakdown the concept of quality into two separate units; study reliability and study relevance' (Collaboration for Environmental Evidence, 2013).

There are no specific standards for quality assessment when it comes to systematic reviews in the realm of conservation and environmental management. This may not be surprising given that this is also the case for health and medicine, where the use of systematic reviews has a long history (Pullin and Knight, 2001). Bilotta *et al.* (2014) suggested that appraisals of quality must have construct validity, provide consistent results among different reviewers, be broadly applicable, and be easy to implement. However, Katrak *et al.* (2004) concluded that there was no 'gold standard' critical appraisal tool for any study design (e.g. experimental, observational, diagnostic and qualitative), nor is there any widely accepted generic tool that can be applied equally well across an array of study types. Of particular interest was, as the authors noted, the fact that there were more critical appraisal methods for experimental than observational study designs (e.g. randomized clinical trials; see Crombie, 1996, e.g. of one critical appraisal approach). The most important critical appraisal elements identified for use in health science by Katrak *et al.* (2004) included random allocation of treatments, appropriateness of outcome measures used, sample size justification/power calculations, study design (whether it was reported) and assessor blinding. Although not all of these are entirely relevant to conservation and management-oriented studies, the concepts are consistent with the aims.

Pitfalls in experimental approaches when aims are to contribute to evidence-based conservation

Study reliability

Reliability is the extent to which the design of a given study minimizes susceptibility to bias. There are four primary considerations.

Selection bias

Selection bias stems from the manner in which treatment groups (including control groups) are amassed (Kunz and

Oxman, 1998). A randomized distribution of experimental units (sites or subjects) to treatments is important to avoid selection bias, though it may not always be feasible. This problem commonly occurs in management situations where control sites are not analogous to treatment sites (i.e. controls are often considered protected areas; [Collaboration for Environmental Evidence, 2013](#)). Under these circumstances, temporal comparisons of samples obtained before and after the establishment of protected areas (also known as ‘before-after control-impact’, BACI) is considered the most powerful experimental design (based on considerations of relative inferential strength), and has been successful, e.g. in evaluating the effectiveness of protected areas ([Osenberg and Schmitt, 1996](#); [Francini-Filho and Moura, 2008](#)) and road mitigation measures ([Rytwinski et al., 2016](#)), and investigating flow regulation services of wetlands ([Kadykalo and Findlay, 2016](#)). Concerns of selection bias have been documented for decades ([Blackwell and Hodges, 1957](#)), but continue to occur. This form of bias constitutes a major problem for experiments that occur sequentially, because such experimental design does not permit for randomization ([Wei, 1978](#)). In some cases, selection bias can be overcome by using a person to select the subjects who is not involved in the experiment itself ([Blackwell and Hodges, 1957](#)). However, the most important source of selection bias originates from baseline difference among treatment groups—groups that differ initially cannot reveal treatment differences. This is also known as non-causal association, and can pose a threat to the validity of a study ([Lipsitch et al., 2010](#)). Selection bias is not limited to the selection of subjects during an experiment. It can also result from the selection of regions or areas of study, often biased toward easily accessible locations ([Phillips et al., 2009](#)). Increased awareness of selection biases and their effects on study reliability can substantially improve experimental reliability. Assigning treatment groups with randomization, stratification or pairing can reduce selection bias, and considering the most relevant approach to a given study can increase the quality of an experiment.

Performance bias

Performance bias affects conservation research when treatment groups receive different standards of care, which can affect the outcome of the treatment itself ([Collaboration for Environmental Evidence, 2013](#)). In medicine, performance bias arises when subjects known to be in a treatment group are provided a different level of attention or care during follow up, asked more or less detailed questions about symptoms, etc. Any systematic difference that alters the balance of the experiment, the validity of the various treatment groups, and the provision of treatment is performance bias. Performance bias can similarly manifest in ecological studies, e.g. if ambient temperatures are monitored more closely in treatment tanks than in control tanks or if the standard of handling care is different for control group subjects than for those undergoing treatment. Both scenarios could alter the effects of the treatment on the treatment group. When the

person collecting the data does so blindly ([Kardish et al., 2015](#)), such that they are incapable of differentiating between control and treatment group individuals, performance bias can be mitigated, and the standard of care is more likely to be equivalent. However, blinding is often impractical in ecological experiments, and is therefore, making performance bias difficult to exclude from many studies.

Measurement bias

Measurement or detection bias occurs when knowledge of the intervention alters adequate evaluation of the results ([Collaboration for Environmental Evidence, 2013](#)). This form of bias is most often addressed through blinding ([Schultz et al., 1995](#)), though, again, blinding is often not possible in the context of biology and ecology. It is often impossible to blind researchers to the systems being studied by the very nature of experimental biology often being field-based. However, whenever blinding is possible, it should become standard practice ([Philipson and DeSimone, 1997](#)). For example, where manipulative experimental studies alter environmental conditions for an organism ([Munday et al., 2016](#) for good example) in a controlled (e.g. laboratory) or semi-controlled (e.g. mesocosm) environment, measurement bias could be mitigated for by ensuring the individuals taking measurements of organisms are unaware of the treatment applied (e.g. water chemistry differences). Alternatively, where differences between treatment groups are obvious (e.g. water turbidity), where possible sampling and recording can be separated, such that measurement bias can be mitigated during measurement of the target outcome, whilst sampling (e.g. extraction of blood samples, video and photograph recordings) need not be blinded ([Hess et al., 2015](#)). Such activities may be done without substantial additional resource requirements.

The use of appropriate technology can also minimize measurement bias by ensuring that the techniques are chosen to best fit subjects and/or species. Some tools used to measure physiological parameters are highly sensitive and should be assessed to ensure they measure the relevant metrics to answer the question at hand. For example, point-of-care devices enable researchers to assess various blood-based physiological parameters of wild animals in remote locations only with careful validation of device performance ([Stoot et al., 2014](#)). To answer questions related to ecological parameters, evaluating population growth and other demographics are fundamentally necessary ([Freckleton et al., 2006](#)). However, the reliability of population estimates is variable and often depends on the observer (detection bias; [Collaboration for Environmental Evidence, 2013](#)) or the population model used, which often has a detection limit that cannot achieve those required to implement changes in many systems ([Hovestadt and Nowicki, 2008](#)).

Attrition bias

Attrition bias refers to differences in withdrawal rates among groups of a study, leading to incomplete outcome data ([Jüni and Egger, 2005](#)). This may result when subjects do not

survive and may change the characteristics of the experimental groups (e.g. as has been observed in rodent biomedical research; [Holman *et al.*, 2016](#)). Analysts that encounter attrition must consider the reasons for attrition and whether the data are likely to be missing at random or missing due to a reason related to fundamental differences between treatment groups ([Harrell, 2015](#)). For example, the researcher may choose to exclude anyone who drops out of an experiment or that disappears from a study area, excluding interesting data that might be related to the treatment (i.e. informative missing). Attrition can therefore compromise randomization ([Leon *et al.*, 2006](#)), and may lead to other forms of bias. Attrition is sometimes unavoidable, and remains outside the realm of what researchers can control. Accurate reporting of attrition and justification for the approach used to account for the issue can guide decisions made by systematic reviewers. Nonetheless, the comparison of unequal samples may interfere with the validity of studies ([Leon *et al.*, 2006](#)).

Study relevance

Relevance is often considered in terms of the external validity of the study ([The Cochrane Collaboration, 2011](#)); how transferable is it to the context of the question being systematically reviewed? However, appraising study relevance can be more subjective than appraising study quality. Here we discuss four primary considerations.

Scale-mismatch

Ecological research is often conducted at the organismal or even down to the sub-organismal scale, but managers typically focus on populations. Exceptions, however, would be for endangered species, for which every individual matters. Therefore, scientists must often extrapolate results to scales that are relevant to management ([Cooke *et al.*, 2014b](#)), even if they are not prepared or trained to do so. This has traditionally been a problem when using experimental evidence in conservation ([Carpenter, 1996](#); [Fausch *et al.*, 2002](#)), despite the fact that scale-dependency of ecological processes is well known ([Levin, 1992](#); [Shea and Chesson, 2002](#); [Holland *et al.*, 2004](#)). However, more experiments are being done at larger temporal and spatial scales ([Hautier *et al.*, 2014, 2015](#)). Experimental manipulation is key to these lessons, as is coordination of researchers across experimental sites ([Stokstad, 2011](#)).

Experimental levels lack relevance

Biologists and ecologists must examine the structure and dynamics of biology from the molecular level, all the way through to populations. Innovative physiological tools available for field studies mean that research is not restricted to the laboratory ([Costa and Sinervo, 2004](#); [Stoot *et al.*, 2014](#)). Yet, logistical and time constraints can still limit researchers in their approach to study a particular system or aspect of a system. For example, angling event simulations ([Donaldson *et al.*, 2011](#)) or the use of supra-physiological hormone injections ([Sopinka *et al.*, 2015](#)) have been widely used in recent years to

mimic the natural responses of animals to challenges that they encounter in the wild. These approaches are intended to control treatment variability, but may also lead to a mismatch between study results and true physiological responses ([Cooke *et al.*, 2013b](#)). For example, cortisol/corticosterone implants are intended to mimic 'chronic' stress in the wild, yet it is unclear the extent to which chronic stress occurs in the wild and what it even looks like in terms of responses ([Boonstra, 2013](#)). Indeed, experimental approaches (usually dose-dependent) may be extreme and unrealistic by design, which could make them irrelevant to the natural system. Another common example is laboratory research related to climate change where several static temperatures are selected (e.g. low, medium and high) and organisms are monitored at those temperatures. In most field environments, there is significant thermal heterogeneity in space and time (e.g. diel variation, seasonal fluctuation and thermal refuge); these conditions are difficult to replicate in the laboratory but can be incorporated in experiments ([Nay *et al.*, 2015](#); [Habary, 2016](#)). Moreover, if conducting research that is intended to be relevant to thermal aspects of climate change, treatments (i.e. exposure) need to be sufficiently long enough for species to acclimate through phenotypic plasticity and potentially across generations such that epigenetic mechanisms and adaptive responses can be identified ([Donelson and Munday, 2015](#); [Donelson *et al.*, 2016](#); [Munday *et al.*, 2016](#)). Consideration of how experimental conditions reflect the experience of wild animals is therefore key to the relevance of studies to systematic review and also to management.

Experimental interventions are impractical

Experimental ecology and conservation science can strive to develop solutions that are relevant to the problem and realistic to the end users. Research funding must be directed toward practical solutions that can be acceptable to stakeholders. Co-development of the research agenda among stakeholders (e.g. landowners, public interest groups and recreational/commercial users) is essential to ensuring that solutions considered in experiments could, in addition to being effective, be scaled (i.e. be mass-produced and/or quickly and easily implemented) and well-received. To do so, experiments must consider possible economic ramifications of experimental interventions. For example, [Cairns *et al.* \(2013\)](#) studied freshwater turtle bycatch exclusion devices and compared the number of turtles successfully excluded as well as differences in retention of target fish species in nets set by commercial fishers, recognizing that the exclusion devices would only be adopted if they did not interfere with normal operation of the nets or reduce target yields. Similar considerations must be made when assessing the effectiveness of conservation interventions so that resources are allocated to creating actionable interventions that are acceptable to all stakeholders.

Use of surrogates

Much conservation and environmental management experimental research aims to understand how environmental alterations will affect species that are already at risk or those

living in extreme stenothermal conditions (e.g. Antarctic fishes). These animals are typically very difficult to obtain due to protective laws or expensive and time-consuming retrieval methods, necessitating the use of surrogates (reviewed in Caro and O'Doherty, 1999). Surrogates offer an avenue for experimentation to reduce the cost of some problems (e.g. animal models in medical science; Shanks *et al.*, 2009). In ecology, surrogate species offer an easy alternative (i.e. they are often cultured or raised in captivity) to wild animals; however, surrogates remain different from the species of interest, which can be an important pitfall to experimental science. Studies that fail to consider sex, physiological differences and population variation will provide irrelevant information to management of wild populations. In many scenarios, a single surrogate species is unlikely to satisfy all the criteria for the aim of a study (Lambeck, 1997). In many cases, surrogates are chosen on the basis of historical precedent, charisma, or ease of management (Mealy and Horn, 1981; Sidle and Suring, 1986), and not necessarily for their match to a key species of interest.

Although there are limitations to using surrogate species or populations to develop evidence, they are necessary and useful in many scenarios where it is impossible to attain the necessary sample size for experimentation using the target species/population (Raby *et al.*, 2015). Replication of studies using every possible population of every species is logistically impossible and unnecessary when surrogates can provide relevant information. Understanding when and in what context surrogates are relevant is crucial to conducting high quality research for the many species at risk of extinction. Appropriate consideration of the relationship between the target and surrogate can improve the relevance of such approaches (Wenger, 2008). Differences between target and surrogate species/populations may exist at both genetic and molecular levels as well as at whole animal scales. For example, Ebner *et al.* (2009) developed tag attachment methods on the surrogate *Macquaria ambigua* for deployment in the closely related but endangered *Macquaria australasica* and found that the surrogate had not provided relevant information. Certainly, conservation triage is useful for identifying the important candidate species and populations for funding and research (Bottrill *et al.*, 2009).

Repeatability, accuracy and statistical power

Insufficient details for repeatability and lack of transparency

Despite the recognized importance of providing sufficient detail for repeatability (Nakagawa and Schielzeth, 2010), lack of such detail is one of the most common pitfalls in experimental biology and ecology. Repeatability is defined as the proportion of the total variance that can be reproduced through repeated measures of the same subjects (Lessells and Boag, 1987), and is often used to evaluate the accuracy,

quality and transparency of data, which is of primary concern to researchers (Garamszegi *et al.*, 2009). Unfortunately, the literature in experimental science often uses diffuse and technical calculations to quantify repeatability (Nakagawa and Schielzeth, 2010), which can discourage others to repeat their experiment. A similar pitfall in experimental biology and ecology is the lack of transparency and replicability, though opportunities to maximize transparency have grown considerably with the advance of online publishing (Wicherts *et al.*, 2012) and repositories (e.g. Dryad and FigShare), which facilitates inclusion of raw data, code scripts and other extensive details that support the findings and improves replicability (Roche *et al.*, 2015). It has even been argued that the openness of the scientific realm is what makes it so successful, suggesting that if researchers were more transparent in their way of practicing and publishing science, the fields of biology and ecology would hugely benefit through a reduction of low quality research and gross errors (Wicherts *et al.*, 2012).

An emerging challenge to repeatable science is in the methodology that is essential to reproducing the results of a study. Methodological details are used by systematic reviewers for critical appraisal and quantitative details (e.g. means, variability, sample size and direction of the trend) and are needed for meta-analyses (Haddaway and Verhoeven, 2015). Optimally, field biologists would design studies alongside statisticians with an analytical tool pre-specified to test the hypothesis. Otherwise, bias can be introduced by serial implementation of possible statistical methods until one is identified that yields desired results (p-hacking; Simonsohn *et al.*, 2014; Head *et al.*, 2015). This is not necessarily a devious practice, but one that can arise when the underlying mathematics of statistical models are poorly understood and a belief that models exhibiting the desired results are more likely to be accurate. Furthermore, this can lead to selective reporting of results within a study when researchers fail to report that a statistical test was conducted when the outcome is weak, negative or absent, hindering interpretation of study findings and weakening meta-analytical syntheses (Jennions *et al.*, 2013; Parker *et al.* 2016). Communicating the modelling approach selected by the analysts can support the reliability of the findings. Accurate reporting of exact *P*-values (rather than $P > 0.05$), presenting results for all tested predictor and response variables, and all subgroups of data, listing assumptions tested to validate the statistical method implemented, and citing statistical software and, in some cases, the specific functions or packages used to produce results improves the understanding of the interpretation and allows systematic reviewers to verify the relevance of results when appraising studies for systematic review.

Statistical power

A related issue is statistical power (Peterman, 1990; Anderson *et al.*, 2001). In many cases, experimental manipulations are undertaken with a few individuals from a single species to inform management decisions for populations or

even multiple species. Particularly with larger animals or species threatened with extinction, population sizes and logistics often do not allow sufficient independent samples to yield robust statistical tests. One potential solution to this issue is relaxing the traditional (and relatively arbitrary) alpha value of 0.05 (Barber and Ogle, 2014; Spanos, 2014). This makes particular sense for issues where the precautionary principle should apply. For example, it may be too costly to sample sufficient numbers of a threatened species to show that an experimental environmental manipulation leads to a ‘significant’ decline using $P < 0.05$. Other researchers have argued for a more ‘information theoretic approach’, which differs from traditional frequentist probability statistics (Burnham and Anderson, 2014). Another solution is the use of larger collaborative, distributed experiments (see ‘Scaling’ above). For example, the approach of coordinated distributed experiments can significantly increase the quality of study designs and boost sample sizes (Fraser *et al.*, 2013) because researchers can more easily pool money for research across a number of projects and plan more comprehensive programmes that achieve superior outcomes.

Pseudoreplication

Pseudoreplication occurs when experimental replicates are not statistically independent (Hurlbert, 1984). Pseudoreplication is a controversial issue in ecology, and the subject of considerable debate (Oksanen, 2001; Quinn and Keough, 2002). An unambiguous example is repeat sampling of an individual. For example, in a study investigating how fish respond to varying levels of oxidative stress, repeat samples of the same individual at different stress levels are not independent experimental replicates. A more ambiguous example is an experiment is divided into two tanks, a treatment and control, where 100 fish are measured from each tank. Although the tanks may be the same design, using the same water source and same fish population, there is always a possibility of unmeasured differences that could be responsible for systematic differences between tanks (e.g. proximity to the laboratory door). On the other hand, it is likely impossible to have 100 tanks for each treatment and sample one fish from each tank. Thus, there is often a sacrifice in replication to account for various aspects of experimental design (Oksanen, 2001).

In larger-scale experiments, spatial pseudoreplication can be an additional problem (Hurlbert, 1984). Almost all environmental drivers of biological phenomena are spatially structured. An experimental design that does not account for this spatial structure (e.g. one where many replicates are concentrated in a small portion of the study area, and few in others) can lead to biased effect sizes and statistical significances (Fortin and Dale, 2005; Beale *et al.*, 2010). Likewise, environmental drivers are almost always temporally auto-correlated. Not accounting for this can result in hidden effects that bias results. It must, however, be accepted by experimenters, reviewers and systematic reviewers that all experiments are pseudoreplicated to some extent and that all

data rely on some level of interpretation of causality (Oksanen, 2001). Although not all instances of pseudoreplication can be accounted for, there are statistical techniques available to account for lack of independence in some study designs that result in pseudoreplication. Mixed effect models incorporate experimental units in which repeated replicates (e.g. repeated measurements on an individual and multiple samples from a tank) are considered as random effects, each with their own expected error distribution, which can be analyzed alongside the fixed effects of the experimental treatments (Bolker *et al.*, 2007; Zuur *et al.*, 2010). Models incorporating auto-correlation of response variables in space can also be used to alleviate spatial pseudoreplication (Beale *et al.*, 2010). However, there is no substitute for foresight in experimental design to provide the most independent test of experimental treatments as possible (Quinn and Keough, 2002).

Experimental design and controls

Lack of appropriate experimental controls

Control is one of the most basic scientific concepts and is one of the great strengths of experimental approaches to conservation. Incorporating a proper control can provide the power necessary to separate correlation from causation in many studies. Control groups for experimentation may be generated using spatially distinct units, e.g. by comparing sites inside and outside protected areas (Pullin *et al.*, 2013; Twardek *et al.*, 2017, in press) or with temporally distinct units or baselines, such as in BACI designs (Underwood, 1992). Control groups should ideally originate from exactly the same population as the treatment group in a randomized way, but perfect controls may be impossible, particularly in field experiments (Smokorowski and Randall, 2017). Thinking carefully about what appropriate controls are and what factors of the experiment require isolation is necessary to achieving robust results from experimentation. Obtaining baseline physiological values is extremely difficult and potentially impossible for wild animals, particularly where the intervention or impact is already in place when experiments begin. However, there are a variety of techniques recommended for physiological control that involve rapid capture and sampling, e.g. prior to the manifestation of some primary and secondary stress hormones (Pankhurst, 2011). Another example where proper control is difficult is in animal tagging (i.e. biotelemetry; Hussey *et al.*, 2015; Kays *et al.*, 2015). Information about animal behaviour, physiology and ecology in the wild can be used for conservation to inform restoration (Lapointe *et al.*, 2013), combat invasive species (Lennox *et al.*, 2016), or study effects of human impacts on wildlife (Donaldson *et al.*, 2008). All animals must be captured, handled (in some cases anaesthetized), and physically attached to or surgically implanted with a tag to do so. These handling/tagging effects can bias findings by affecting the natural behaviour of animals (Wilson and McMahon, 2006; Jepsen *et al.*, 2015), and control groups

are logistically difficult or impossible to obtain in some experimental scenarios (Cooke *et al.* 2013b). Laboratory studies prior to tagging can be used to compare survival and activity of tagged and untagged animals in captivity to determine whether the tagged animals will be representative of the population.

Too much control

Control groups for experimentation must be developed with the broader question in mind so that the control and treatment groups in an experiment are established to provide relevant results. Over-control and isolation of too few potential predictor variables limits discovery of the interesting interactions that occur in nature. Systems-based thinking is necessary to establish experiments that can produce relevant, scalable, transferrable results to real world problems in ecology and conservation.

Conclusion

As a crisis discipline, conservation science strives to develop evidence that is timely, reliable and relevant to management (Soulé, 1985). With technological advancements in remote monitoring of behaviour and physiology, the gap between experimental and field approaches to conservation is narrowing. Novel experimental approaches now possible at larger, more ecologically relevant scales have more power to detect causality in the field and separate complex interactions that exist in nature. Although an expanding human footprint and rapidly changing environment have led to increasingly complex and difficult conservation questions, experimental biology is capable of producing the evidence needed to address many of the key questions so that solutions can be found across a range of relevant scales (Cooke *et al.*, 2014b).

Developing research projects that generate reliable evidence is key but must rely on a solid partnership between scientists, managers and stakeholders. Not only must managers be able to communicate the pressing questions to scientists, they must be able to acknowledge how evidence at various scales can inform policy at the present scale. However, it is generally the responsibility of scientists to consider what scale to work on in a given situation with the available tools. Although the use of experimental biology has permitted great progress in understanding how biological systems interact with one another, we are still far from completely understanding how rapid environmental changes will interact with ecosystem function, population dynamics and individual physiologies (Altshuler *et al.*, 2011). The need to better understand the interactive impacts of multiple stressors is still considered one of the most pressing questions in ecology and conservation (Sala *et al.*, 2000; Crain *et al.*, 2009).

Evidence is increasingly important in managing all aspects of governance, and conservation is becoming one of the leading disciplines in developing protocols and standards for

evidence (Dicks *et al.*, 2014). Scientists working in the realm of conservation physiology can have great impacts on conservation (Madliger *et al.*, 2016), but must be aware of the pitfalls outlined above. Educating scientists as to what constitutes the good experimental evidence that is necessary to inform management and policy will be key. In addition, proper reporting protocols for communicating evidence in the literature will be positive for management requiring support for their decision making, scientists striving to produce relevant research, and for the natural systems that all parties must be working collaboratively to manage and conserve.

Acknowledgements

We thank the Society for Experimental Biology for supporting travel for this article to be presented at their annual meeting in Brighton.

Funding

S.J.C., R.J.L., K.B.G., J.J.T. and T.R. are supported by the Canada Research Chairs Program and the Natural Sciences and Engineering Research Council of Canada (NSERC). Bennett is supported by an NSERC Discovery Grant. J.L.R. is supported by an Australian Research Council (ARC) Discovery Award as well as the ARC Centre of Excellence for Coral Reef Studies. N.R.H. is supported by Mistra EviEM, part of the Stockholm Environment Institute.

References

- Altshuler I, Demiri B, Xu S, Constantin A, Yan ND, Cristescu ME (2011) An integrated multi-disciplinary approach for studying multiple stressors in freshwater ecosystems: Daphnia as a model organism. *Integr Comp Biol* 51:623–633.
- Anderson DR, Burnham KP, Gould WR, Cherry S (2001) Concerns about finding effects that are actually spurious. *Wildlife Soc Bull* 29: 311–316.
- Barber JJ, Ogle K (2014) To P or not to P? *Ecology* 95:621–626.
- Bartholomew RE (2014) Science for sale: the rise of predatory journals. *J R Soc Med* 107:384–385.
- Beale CM, Lennon JJ, Yearsley JM, Brewer MJ, Elston DA (2010) Regression analysis of spatial data. *Ecol Lett* 13:246–64.
- Bernes C, Carpenter SR, Gårdmark A, Larsson P, Persson L, Skov C, Speed JD, Van Donk E (2015) What is the influence of a reduction of planktivorous and benthivorous fish on water quality in temperate eutrophic lakes? A systematic review. *Environ Evidence* 4:7.
- Bilotta GS, Milner AM, Boyd I (2014) On the use of systematic reviews to inform environmental policies. *Environ Sci Policy* 42:67–77.
- Blackwell D, Hodges JL Jr (1957) Design for the control of selection bias. *Ann Math Stat* 28:449–460.

- Blaustein AR, Kiesecker JM (2002) Complexity in conservation: lessons from the global decline of amphibian populations. *Ecol Lett* 5: 597–608.
- Bolker BM, Okuyama T, Bjørndal KA, Bolten AB (2007) Incorporating multiple mixed stocks in mixed stock analysis: ‘many-to-many’ analyses. *Mol Ecol* 16:685–95.
- Boonstra R (2013) Reality as the leading cause of stress: rethinking the impact of chronic stress in nature. *Funct Ecol* 27:11–23.
- Bottrill MC, Joseph LN, Carwardine J, Bode M, Cook C, Game ET, Grantham H, Kark S, Linke S, McDonald-Madden E, et al. (2009) Finite conservation funds mean triage is unavoidable. *Trends Ecol Evol* 24:183–184.
- Burnham KP, Anderson DR (2014) P values are only an index to evidence: 20th- vs. 21st-century statistical science. *Ecology* 95:627–630.
- Cairns NA, Stoot LJ, Blouin-Demers G, Cooke SJ (2013) Refinement of bycatch reduction devices to exclude freshwater turtles from commercial fishing nets. *Endanger Species Res* 22:251–261.
- Campbell Collaboration (2016) Campbell Systematic Reviews: Policies and Guidelines. Version 1.2 [updated November 2016]. Noonan E, Goudie S, eds. Available from www.campbellcollaboration.org (accessed April 2017).
- Cardinale BJ, Duffy JE, Gonzalez A, Hooper DU, Perrings C, Venail P, Narwani A, Mace GM, Tilman D, Wardle DA, et al. (2012) Biodiversity loss and its impact on humanity. *Nature* 486:59–67.
- Caro TM, O’Doherty G (1999) On the use of surrogate species in conservation biology. *Conserv Biol* 13:805–814.
- Carpenter SR (1996) Microcosm experiments have limited relevance for community and ecosystem ecology. *Ecology* 77:677–80.
- Ceccato E, Cramp RL, Seebacher F, Franklin CE (2016) Early exposure to UV-B radiation decreases immune function later in life. *Conserv Physiol* 4:cow037.
- Christensen MR, Graham MD, Vinebrooke RD, Findlay DL, Paterson MJ, Turner MA (2006) Multiple anthropogenic stressors cause ecological surprises in boreal lakes. *Global Change Biol* 12:2316–22.
- Collaboration for Environmental Evidence (2013) Guidelines for systematic reviews in environmental management. Version 4.2. Environmental Evidence.
- Concato J (2004) Observational versus experimental studies: what’s the evidence for a hierarchy? *NeuroRx* 1:341–347.
- Cook CN, Mascia MB, Schwartz MW, Possingham HP, Fuller RA (2013) Achieving conservation science that bridges the knowledge–action boundary. *Conserv Biol* 27:669–678.
- Cooke SJ, Donaldson MR, O’Connor CM, Raby GD, Arlinghaus R, Danylchuk AJ, Hanson KC, Hinch SG, Clark TD, Patterson DA, et al. (2013b) The physiological consequences of catch-and-release angling: perspectives on experimental design, interpretation, extrapolation, and relevance to stakeholders. *Fish Manag Ecol* 20: 268–287.
- Cooke SJ, Sack L, Franklin CA, Farrell AP, Beardall J, Wikelski M, Chown SL (2013a) What is conservation physiology? Perspectives on an increasingly integrated and essential science. *Conserv Physiol*. doi:10.1093/conphys/cot001.
- Cooke SJ, Blumstein DT, Buchholz R, Caro T, Fernández-Juricic E, Franklin CE, Metcalfe J, O’Connor CM, St. Clair C, Sutherland WJ, et al. (2014a) Physiology, behaviour, and conservation. *Physiol Biochem Zool* 87:1–14.
- Cooke SJ, Killen SS, Metcalfe JD, McKenzie DJ, Mouillot D, Jørgensen C, Peck MA (2014b) Conservation physiology across scales: insights from the marine realm. *Conserv Physiol*. doi:10.1093/conphys/cou024.
- Coors A, De Meester L (2008) Synergistic, antagonistic and additive effects of multiple stressors: predation threat, parasitism and pesticide exposure in *Daphnia magna*. *J Appl Ecol* 45:1820–8.
- Costa DP, Sinervo B (2004) Field physiology: physiological insights from animals in nature. *Ann Rev Physiol* 66:209–238.
- Crain CM, Halpern BS, Beck MW, Kappel CV (2009) Understanding and managing human threats to the coastal marine environment. *Ann NY Acad Sci* 1162:39–62.
- Craver C (2007) *Explaining the Brain: Mechanisms and Mosaic Unity of Neuroscience*. Oxford University Press, Oxford.
- Crombie IK (1996) *The Pocket Guide to Critical Appraisal: A Handbook for Health Care Professionals*. BMJ Publishing Group, London.
- Crutzen PJ (2006) The “anthropocene”. *Earth System Science in the Anthropocene*. Springer, Berlin Heidelberg. 13–18.
- Dalby S (2016) Framing the anthropocene: the good, the bad and the ugly. *Anthropocene Rev* 3:33–51.
- Darling ES, Côté IM (2008) Quantifying the evidence for ecological synergies. *Ecol Lett* 11:1278–1286.
- Detsky AS, Naylor CD, O’Rourke K, McGeer AJ, L’Abbé KA (1992) Incorporating variations in the quality of individual randomized trials into meta-analysis. *J Clin Epidemiol* 45:255–265.
- Di Santo V (2015) Ocean acidification exacerbates the impacts of global warming on embryonic little skate, *Leucoraja erinacea* (Mitchill). *J Exp Mar Biol Ecol* 463:72–78.
- Dicks LV, Walsh JC, Sutherland WJ (2014) Organising evidence for environmental management decisions: a ‘4S’ hierarchy. *Trends Ecol Evol* 29:607–613.
- Donaldson MR, Arlinghaus R, Hanson KC, Cooke SJ (2008) Enhancing catch-and-release science with biotelemetry. *Fish Fish* 9:79–105.
- Donaldson MR, Hinch SG, Patterson DA, Hills J, Thomas JO, Cooke SJ, Raby GD, Donaldson LA, Robichaud D, Farrell AP, et al. (2011) The consequences of angling and beach seine capture on the physiology, post-release behaviour and survival of adult sockeye salmon during upriver migration. *Fish Res* 108:133–141.
- Donelson JM, Munday PL (2015) Transgenerational plasticity mitigates the impact of global warming to offspring sex ratios. *Glob Change Biol* 21:2954–62.

- Donelson JM, Wong M, Booth DJ, Munday PL (2016) Transgenerational plasticity of reproduction depends on rate of warming across generations. *Evol Appl* 9:1072–81.
- Ebner BC, Lintermans M, Jekabsons M, Dunford M, Andrews W (2009) A cautionary tale: surrogates for radio-tagging practice do not always simulate the responses of closely related species. *Mar Freshwater Res* 60:371–378.
- Fausch KD, Torgersen CE, Baxter CV, Li HW (2002) Landscapes to riverscapes: bridging the gap between research and conservation of stream fishes. *BioScience* 52:483–98.
- Folt CL, Chen CY, Moore MV, Burnaford J (1999) Synergism and antagonism among multiple stressors. *Limnol Oceanogr* 44:864–77.
- Fortin MJ, Dale MR (2005) *Spatial Analysis: A Guide for Ecologists*. Cambridge University Press, Cambridge, UK.
- Francini-Filho RB, Moura RL (2008) Evidence for spillover of reef fishes from a no-take marine reserve: an evaluation using the before-after control-impact (BACI) approach. *Fish Res* 93:346–356.
- Fraser LH, Henry HA, Carlyle CN, White SR, Beierkuhnlein C, Cahill JH Jr, Casper BB, Cleland E, Collins SL, Dukes JS, et al. (2013) Coordinated distributed experiments: an emerging tool for testing global hypotheses in ecology and environmental science. *Front Ecol Environ* 11:147–155.
- Freckleton RP, Watkinson AR, Green RE, Sutherland WJ (2006) Census error and the detection of density dependence. *J Anim Ecol* 75:837–851.
- Garamszegi LZ, Calhim S, Dochtermann N, Hegyi G, Hurd PL, Jørgensen C, Kutsukake N, Lajeunesse MC, Pollard KA, Schielzeth H, et al. (2009) Changing philosophies and tools for statistical inferences in behavioral ecology. *Behav Ecol* 20:1363–1375.
- Green DG, Sadedin S (2005) Interactions matter—complexity in landscapes and ecosystems. *Ecol Complexity* 2:117–30.
- Habary A (2016) Adapt, move or die—how will tropical coral reef fishes cope with ocean warming? *Global Change Biol* 23:566–577.
- Harrell F. (2015) *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*. Springer-Verlag, New York.
- Hautier Y, Seabloom EW, Borer ET, Adler PB, Harpole WS, Hillebrand H, Lind EM, MacDougall AS, Stevens CJ, Bakker JD, et al. (2014) Eutrophication weakens stabilizing effects of diversity in natural grasslands. *Nature* 508:521–5.
- Hautier Y, Tilman D, Isbell F, Seabloom EW, Borer ET, Reich PB (2015) Anthropogenic environmental changes affect ecosystem stability via biodiversity. *Science* 348:336–40.
- Haddaway NR, Verhoeven JTA (2015) Poor methodological detail precludes experimental repeatability and hampers synthesis in ecology. *Ecol Evol* 5:4451–4454.
- Head ML, Homan L, Lanfear R, Kahn AT, Jennions MD (2015) The extend of consequences of p-hacking in science. *PLoS Biol* 12: e1002106.
- Hess S, Wenger AS, Ainsworth TD, Rummer JL (2015) Exposure of clownfish larvae to suspended sediment levels found on the Great Barrier Reef: impacts on gill structure and microbiome. *Sci Rep* 5: 10561.
- Holland JD, Bert DG, Fahrig L (2004) Determining the spatial scale of species' response to habitat. *BioScience* 54:227–233.
- Holman C, Piper SK, Grittner U, Diamantaras AA, Kimmelman J, Siegerink B, Dirnagl U (2016) Where have all the rodents gone? The effects of attrition in experimental research on cancer and stroke. *PLoS Biol* 14:e1002331.
- Hovestadt T, Nowicki P (2008) Process and measurement errors of population size: their mutual effects on precision and bias of estimates for demographic parameters. *Biodivers Conserv* 17: 3417–3429.
- Hurlbert SH (1984) Pseudoreplication and the design of ecological field experiments. *Ecol Monograph* 54:187–211.
- Hussey NE, Kessel ST, Aarestrup K, Cooke SJ, Cowley PD, Fisk AT, Harcourt RG, Holland KN, Iverson SJ, Kocik JF, et al. (2015) Aquatic animal telemetry: a panoramic window into the underwater world. *Science*. doi:10.1126/science.1255642.
- Jennions MD, Lortie CJ, Rosenberg MS, Rothstein HR (2013) Publication and related biases. In Koricheva J, Gurevitch J, Mengersen K (eds). *Handbook of Meta-analysis in Ecology and Evolution*. Princeton University Press, USA. 207–236.
- Jepsen N, Thorstad EB, Havn T, Lucas MC (2015) The use of external electronic tags on fish: an evaluation of tag retention and tagging effects. *Anim Biotelem* 3:1.
- Jüni P, Egger M (2005) Empirical evidence of attrition bias in clinical trials. *Int J Epidemiol* 34:87–88.
- Kadykalo AN, Findlay CS (2016) The flow regulation services of wetlands. *Ecosystem Services* 20:91–103.
- Kardish MR, Mueller UG, Amador-Vargas S, Dietrich EI, Ma R, Barrett B, Fang CC (2015) Blind trust in unblinded observation in ecology, evolution, and behavior. *Front Ecol Evol* 3:51.
- Katrak P, Bialocerkowski AE, Massy-Westropp N, Kumar VS, Grimmer KA (2004) A systematic review of the content of critical appraisal tools. *BMC Med Res Methodol* 4:22.
- Kays R, Crofoot MC, Jetz W, Wikelski M (2015) Terrestrial animal tracking as an eye on life and planet. *Science* 348(6240): aaa2478.
- Kunz R, Oxman AD (1998) The unpredictability paradox: review of empirical comparisons of randomised and non-randomised trials. *Br Med J* 317:1185–1190.
- Lambeck RJ (1997) Focal species: a multi-species umbrella for nature conservation. *Conserv Biol* 11:849–857.
- Lapointe NW, Thiem JD, Doka SE, Cooke SJ (2013) Opportunities for improving aquatic restoration science and monitoring through the use of animal electronic-tagging technology. *BioScience* 63: 390–396.

- Lennox RJ, Blouin-Demers G, Rous AM, Cooke SJ (2016) Tracking invasive animals with electronic tags to assess risks and develop management strategies. *Biol Invasions* 18:1219–1233.
- Leon AC, Mallinckrodt CH, Chuang-Stein C, Archibald DG, Archer GE, Chartier K (2006) Attrition in randomized controlled clinical trials: methodological issues in psychopharmacology. *Biol Psychiatry* 59: 1001–1004.
- Lessells CM, Boag PT (1987) Unrepeatable repeatabilities: a common mistake. *Auk* 104:116–121.
- Levin SA (1992) The problem of pattern and scale in ecology: the Robert H. MacArthur award lecture. *Ecology* 73:1943–67.
- Lidicker WZ Jr (1988) The synergistic effects of reductionist and holistic approaches in animal ecology. *Oikos* 1:278–81.
- Lindenmayer DB, Piggott MP, Wintle BA (2013) Counting the books while the library burns: why conservation monitoring programs need a plan for action. *Front Ecol Environ* 11:549–555.
- Lipsitch M, Tchetgen ET, Cohen T (2010) Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology* 21:383–388.
- Lotze HK, Worm B (2002) Complex interactions of climatic and ecological controls on macroalgal recruitment. *Limnol Oceanogr* 47: 1734–41.
- Mace GM, Collar NJ, Gaston KJ, Hilton-Taylor CR, Akçakaya HR, Leader-Williams NI, Milner-Gulland EJ, Stuart S (2008) Quantification of extinction risk: IUCN's system for classifying threatened species. *Conserv Biol* 22:1424–42.
- Madliger CL, Cooke SJ, Crespi EJ, Funk JL, Hultine KR, Hunt KE, Rohr JR, Sinclair BJ, Suski CD, Willis CK, et al. (2016) Success stories and emerging themes in conservation physiology. *Conserv Physiol* 4: cov057.
- Mealy SP, Horn JR (1981) Integrating wildlife habitat objectives into the forest plan. *Trans N Am Wildl Nat Resour Conf* 46:488–500.
- Munday PL, Welch MJ, Allan BJ, Watson SA, McMahon SJ, McCormick MI. (2016) Effects of elevated CO₂ on predator avoidance behaviour by reef fishes is not altered by experimental test water. *PeerJ* 4:e2501.
- Nakagawa S, Schielzeth H (2010) Repeatability for Gaussian and non-Gaussian data: a practical guide for biologists. *Biol Rev* 85:935–956.
- Nay TJ, Johansen JL, Habary A, Steffensen JF, Rummer JL (2015) Behavioural thermoregulation in a temperature-sensitive coral reef fish, the five-lined cardinalfish (*Cheilodipterus quinquelineatus*). *Coral Reefs* 34:1261–5.
- Neff BD, Olden JD (2006) Is peer review a game of chance? *BioScience* 56:333–340.
- Nichols SJ, Peat M, Webb JA (2017) Challenges for evidence-based environmental management: what is acceptable and sufficient evidence of causation? *Freshwater Sci* 36:240–249.
- O'Connor CM, Norris DR, Crossin GT, Cooke SJ (2014) Biological carry-over effects: linking common concepts and mechanisms in ecology and evolution. *Ecosphere* 5(3): article 28. doi:10.1890/ES13-00388.1.
- O'Connor CM, Cooke SJ (2015) Ecological carryover effects complicate conservation. *Ambio* 44:582–591.
- Oksanen L (2001) Logic experiments in ecology: is pseudoreplication a pseudoissue? *Oikos* 94:27–38.
- Osenberg CW, Schmitt RJ (1996) Detecting ecological impacts caused by human activities. In Schmitt RJ, Osenberg CW (eds). *Detecting Ecological Impacts*. Academic Press, San Diego. 3–16.
- Pankhurst NW (2011) The endocrinology of stress in fish: an environmental perspective. *Gen Comp Endocrinol* 170:265–275.
- Parker TH, Forstmeier W, Koricheva J, Fidler F, Hadfield JD, En Chee Y, Kelly CD, Gurevitch J, Nakagawa S (2016) Transparency in ecology and evolution: real problems, real solutions. *Trends Ecol Evol* 31: 711–719.
- Peres-Neto PR (2016) Will technology trample peer review in ecology? Ongoing issues and potential solutions. *Oikos* 125:3–9.
- Peterman RM (1990) Statistical power analysis can improve fisheries research and management. *Can J Fish Aquat Sci* 47:2–15.
- Phillips SJ, Dudík M, Elith J, Graham CH, Lehmann A, Leathwick J, Ferrier S (2009) Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecol Appl* 19:181–197.
- Phillipson T, DeSimone J (1997) Experiments and subject sampling. *Biometrika* 84:619–630.
- Pimm SL, Jenkins CN, Abell R, Brooks TM, Gittleman JL, Joppa LN, Raven PH, Roberts CM, Sexton JO (2014) The biodiversity of species and their rates of extinction, distribution, and protection. *Science* 344:1246752.
- Pullin AS, Knight TM (2001) Effectiveness in conservation practice: pointers from medicine and public health. *Conserv Biol* 15:50–54.
- Pullin AS, Knight TM (2003) Support for decision making in conservation practice: an evidence-based approach. *J Nat Conserv* 11: 83–90.
- Pullin AS, Knight TM (2009a) Data credibility: a perspective from systematic reviews in environmental management. *New Directions Eval* 122:65–74.
- Pullin AS, Knight TM (2009b) Doing more good than harm—building an evidence-base for conservation and environmental management. *Biol Conserv* 142:931–934.
- Pullin AS, Knight TM (2012) Science informing policy—a health warning for the environment. *Environ Evidence* 1:15.
- Pullin AS, Stewart GB (2006) Guidelines for systematic review in conservation and environmental management. *Conserv Biol* 20: 1647–1656.

- Pullin AS, Knight TM, Stone DA, Charman K (2004) Do conservation managers use scientific evidence to support their decision-making? *Biol Conserv* 119:245–252.
- Pullin AS, Bangpan M, Dalrymple S, Dickson K, Haddaway NR, Healey JR, Hauari H, Hockley N, Jones PG, Knight T, *et al.* (2013) Human well-being impacts of terrestrial protected areas. *Environ Evidence* 2:19.
- Quinn GP, Keough MJ (2002) *Experimental Design and Data Analysis for Biologists*. Cambridge University Press, Cambridge, UK.
- Raby GD, Clark TD, Farrell AP, Patterson DA, Bett NN, Wilson SM, Willmore WG, Suski CD, Hinch SG, Cooke SJ, *et al.* (2015) Facing the river gauntlet: understanding the effects of fisheries capture and water temperature on the physiology of coho salmon. *PLoS One* 10:e0124023.
- Roche DG, Kruuk LEB, Lanfear R, Binning SA (2015) Public data archiving in ecology and evolution: how well are we doing? *PLoS Biol* 13: e1002295.
- Rytwinski T, Soanes K, Jaeger JAG, Fahrig L, Findlay CS, Houlahan J, van der Ree R, van der Grift EA (2016) How effective is road mitigation at reducing road-kill? A meta-analysis. *PLoS One* 11: e0166941.
- Sala OE, Chapin FS, Armesto JJ, Berlow E, Bloomfield J, Dirzo R, Huber-Sanwald E, Huenneke LF, Jackson RB, Kinzig A, *et al.* (2000) Global biodiversity scenarios for the year 2100. *Science* 287:1770–4.
- Schultz KF, Chalmers I, Hayes RJ, Altman DG (1995) Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *J Am Med Assoc* 273: 408–412.
- Seebacher F, Franklin CE (2012) Determining environmental causes of biological effects: the need for a mechanistic physiological dimension in conservation biology Introduction. *Phil Trans R Soc B* 367: 1607–1614.
- Shea K, Chesson P (2002) Community ecology theory as a framework for biological invasions. *Trends Ecol Evol* 17:170–6.
- Shanks N, Greek R, Greek J (2009) Are animal models predictive for humans? *Philos Ethics Humanit Med* 4:2.
- Sidle WB, Suring LH (1986) *Management Indicator Species for the National Forest Lands in Alaska*. Technical publication R10-TP-2. U.S. Forest Service, Juneau, Alaska.
- Simonsohn U, Nelson LD, Simmons JP (2014) P-curve: a key to the file-drawer. *J Exp Psychol* 143:534–547.
- Smith R (2006) Peer review: a flawed process at the heart of science and journals. *J R Soc Med* 99:178–182.
- Smokorowski KE, Randall RG (2017) Cautions on using the Before-After-Control-Impact design in environmental effects monitoring programs. *FACETS* 2:212–232. doi:10.1139/facets-2016-0058.
- Sopinka NM, Patterson LD, Redfern JC, Pleizier N, Belanger C, Midwood JD, Crossin GT, Cooke SJ (2015) Manipulating glucocorticoids in wild animals: basic and applied perspectives. *Conserv Physiol* 3:cov031.
- Soulé ME (1985) What is conservation biology? *BioScience* 35:727–734.
- Spanos A (2014) Recurring controversies about *P* values and confidence intervals revisited. *Ecology* 95:645–651.
- Stevens A, Milne R (1997) The effectiveness revolution and public health. In Scally G (ed). *Progress in Public Health*. Royal Society of Medicine Press, London. 197–225.
- Stokstad E (2011) Open-source ecology takes root across the world. *Science* 334:308–9.
- Stoot LJ, Cairns NA, Cull F, Taylor JJ, Jeffrey JD, Morin F, Mandelman JW, Clark TD, Cooke SJ (2014) Use of portable blood physiology point-of-care devices for basic and applied research on vertebrates: a review. *Conserv Physiol* 2:cou011.
- Sutherland WJ, Pullin AS, Dolman PM, Knight TM (2004) The need for evidence-based conservation. *Trends Ecol Evol* 19:305–8.
- Sutton AJ, Duval SJ, Tweedie RL, Abrams KR, Jones DR (2000) Empirical assessment of effect of publication bias on meta-analyses. *Br Med J* 320:1574–7.
- The Cochrane Collaboration (2011) *Cochrane Handbook for Systematic Reviews of Interventions*. Version 5.1.0 [updated March 2011]. Higgins J, Green S, eds. www.cochrane-handbook.org (accessed April 2017).
- Twardek WM, Elvidge CK, Wilson AD, Algera DA, Zolderdo AJ, Loughheed SC, Cooke SJ (2017) Do protected areas mitigate the effects of fisheries-induced evolution on parental care behaviour of a teleost fish? *Aquat Conserv: Mar Freshw Ecosyst*. doi: 10.1002/aqc.2718.
- Underwood AJ (1992) Beyond BACI: the detection of environmental impacts on populations in the real, but variable, world. *J Exp Mar Biol Ecol* 161:145–178.
- van Uitregt VO, Wilson RS, Franklin CE (2007) Cooler temperatures increase sensitivity to ultraviolet B radiation in embryos and larvae of the frog *Limnodynastes peronii*. *Global Change Biol* 13:1114–1121.
- Vitousek PM, Mooney HA, Lubchenco J, Melillo JM (1997) Human domination of Earth's ecosystems. *Science* 277:494–9.
- Weber M (2012) Experiment in biology. Stanford Encyclopedia of Philosophy.
- Weber M (2004) *Philosophy of Experimental Biology*. Cambridge University Press, Cambridge, UK.
- Wei LJ (1978) On the random allocation design for the control of selection bias in sequential experiments. *Biometrika* 65:79–84.
- Wenger SJ (2008) Use of surrogates to predict the stressor response of imperiled species. *Conserv Biol* 22:1564–1571.
- Werner EE (1998) Ecological experiments and a research program in community ecology. In Reserits WJ Jr, Bernardo J (eds). *Experimental Ecology: Issues and Perspectives*. Oxford University Press, Oxford. 3–26.

- Wicherts JM, Kievit RA, Bakker M, Borsboom D (2012) Letting the daylight in: reviewing the reviewers and ways to maximize transparency in science. *Front Comput Neurosci* 6:1–9.
- Wilson RP, McMahon CR (2006) Measuring devices on wild animals: what constitutes acceptable practice? *Front Ecol Environ* 4: 147–154.
- Young N, Corriveau M, Nguyen VM, Cooke SJ, Hinch SG (2016) How do potential knowledge users evaluate new claims about a contested resource? Problems of power and politics in knowledge exchange and mobilization. *J Environ Manage* 184:380–388.
- Zuur AF, Ieno EN, Elphick CS (2010) A protocol for data exploration to avoid common statistical problems. *Methods Ecol Evol* 1:3–14.