# Evaluation Options for Wildlife Management and Strengthening of Causal Inference

JIM HONE, V. ALISTAIR DRAKE, AND CHARLES J. KREBS

*Wildlife management aims to halt and then reverse the decline of threatened species, to sustainably harvest populations, and to control undesirable impacts of some species. We describe a unifying framework of three feasible options for evaluation of wildlife management, including conservation, and discuss their relative strengths of statistical and causal inference. The first option is trends in abundance, which can provide strong evidence a change has occurred (statistical inference) but does not identify the causes. The second option assesses population outcomes relative to management efforts, which provides strong evidence of cause and effect (causal inference) but not the trend. The third option combines the first and second options and therefore provides both statistical and causal inferences in an adaptive framework. We propose that wildlife management needs to explicitly use causal criteria and inference to complement adaptive management. We recommend incorporating these options into management plans.*

*Keywords: adaptive management, biodiversity conservation, causality, population trends, strength of inference*

**Many wildlife species are managed to conserve** them, to sustainably harvest them, or to control them, such as reducing their undesirable impacts (Caughley 1980). Ideally, management should be both effective in achieving its aims and efficient in its use of scarce resources and should also be responsive to evidence of its impacts through monitoring. Therefore, management can be a learning process termed *adaptive management* (Walters and Holling 1990, Westgate et al. 2013).

In conservation, the aim is to halt and then reverse declines in the abundance and distribution of threatened species (Caughley 1994, Green 2002, Akcakaya et al. 2018). Conservation programs need a clear, quantifiable, aim (Tear et al. 2005) so their effectiveness can be assessed. Vague aims such as "to improve the trajectories of priority threatened species by 2031" (Australian government 2021) or "the risk of extinction is reduced for all priority species" (Australian government 2022) could be interpreted as implying an increase in abundance or a slower rate of decline of abundance and therefore provide only weak guidance for managers trying to achieve positive conservation outcomes. Assessment usually requires monitoring (Lindenmayer and Gibbons 2012, Legge et al. 2018) to provide information on the status of the population being managed and whether the management aim is achieved. The costs of management options should also be monitored and evaluated (Reddiex and Forsyth 2006) and should be used to estimate effort–outcome relationships (Hone et al. 2017, 2018) and the return on investment (Murdoch et al. 2007).

Monitoring a species may be sufficient to determine its status according to the conventional categories of the International Union for the Conservation of Nature (IUCN 2012)—such as whether it is Threatened, Endangered, or Critically Endangered. Monitoring is needed to assess trends (figure 1) which should be inferred from formal analyses. For example, woodland birds in Australia have often been reported to be declining in abundance, but a detailed review of studies showed only 14 out of 44 of these studies incorporated formal statistical analysis of trends (Rayner et al. 2014). Similarly, there have been limited evaluations of the effects of vertebrate pest control on the pests or their impacts (Reddiex and Forsyth 2006, Reddiex et al. 2006) and of many conservation efforts (Williams et al. 2020). On a positive note, there are many reported examples of positive trends following conservation efforts (Garnett et al. 2018), and these are being recognized in the Green Status of Species (Akcakaya et al. 2018, Grace et al. 2021). Trends use statistical inference to provide evidence of change in abundance (figure 1) but do not identify the causes (time is not a cause; Harre 1972).

We recognize two broad classes of graphical representation of management outcomes, conceptual graphs and

*BioScience* 73: 48–58. 
https://doi.org/10.1093/biosci/biac105

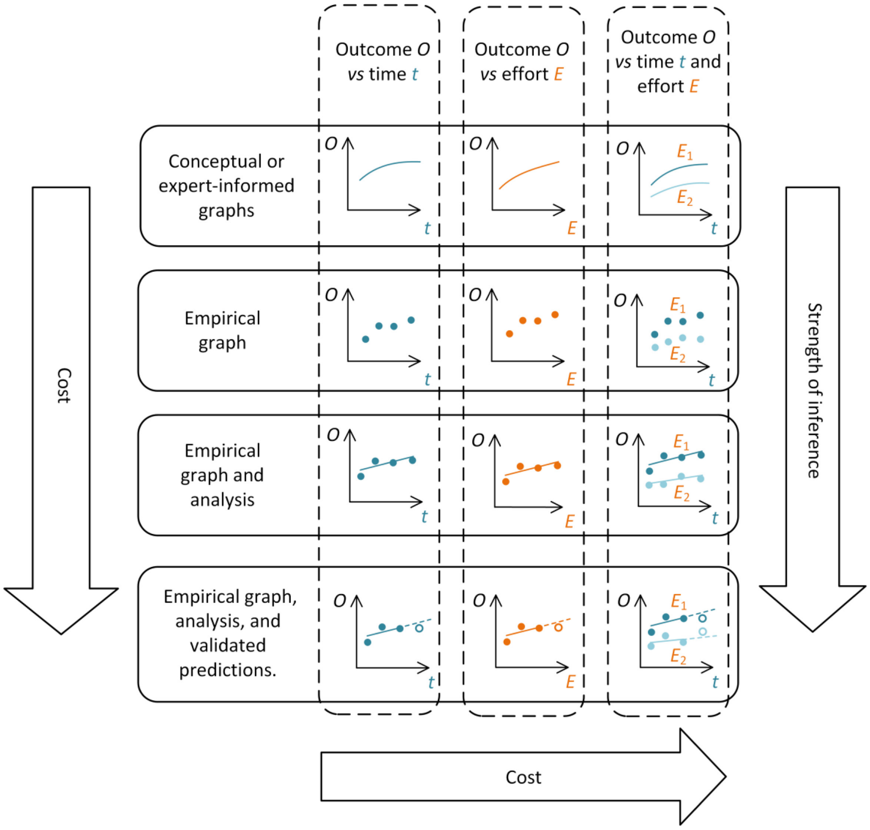


***Figure 1. A framework of schematic graphs showing three options, as columns, for evaluating wildlife management. From the top down within the column of outcomes (O) versus time (t), there is increasingly stronger statistical inference of a trend. From the top down within the middle column of outcomes versus efforts (E), there is increasingly stronger statistical inference of parameter values and stronger causal inference that management efforts caused the outcomes. From the top down within the right column of outcomes versus both time and efforts, there is increasingly stronger statistical inference of trends and parameter values, and stronger causal inference that management efforts caused the trends. In the bottom row, open circles are predictions that are validated independently (dashed lines). Examples of options in each column are illustrated in figures 2, 3, and 4 and in tables 1, 2, and 3 respectively. Costs of options increase down each column and across each row. For simplicity, aims and measures of precision are not explicit but should be recognized.***

empirical graphs (figure 1). Conceptual graphs have no data and detail hypothetical trends or relationships of what has happened, what is happening now, and what is expected to happen in the future (figure 1). Conceptual graphs have an obvious role at the planning stage of management and can also be thought of as models of alternative hypotheses to be evaluated as the program proceeds. They may reflect intentions, established knowledge, or expert opinion. Expert opinions or judgements can be useful (Joseph et al. 2009), especially if procedures to reduce bias are incorporated (Speirs-Bridge et al. 2010). In contrast, empirical graphs are based on data obtained through monitoring (figure 1). In all these graphs, the outcome measure—for example, abundance or density—is presented on the vertical axis (as a dependent variable), and time or the level of management effort is on the horizontal axis (independent variable). Causal theory is clear that time is not a cause (Harre 1972), and, therefore, graphs of outcomes as a function of time show trends and not causes. Empirical graphs with analyzed data and graphs incorporating validated predictions provide progressively stronger statistical and causal inference that the outcomes did occur and why; in figure 1, this progression increases from a minimum at top left to a maximum at bottom right.

The aim of this article is to describe a unifying framework that incorporates three options—the option of evaluating wildlife management on the basis of assessment of trends but not causes, an alternative option on the basis of evidence that management causes observed outcomes but without assessing trends, and a third option that combines evidence of both trends and causes. We distinguish between statistical inference, with its focus on parameter estimation and inferences from a sample, and causal inference, with its use of logic and evidence. For example, statistical inference using linear regression focuses on estimates of the slope and intercept and the precision of each. In contrast, causal inference focuses on the extent and direction of change in an outcome when management effort changes (increases or decreases) and therefore links cause and effect. We show how strength of inference—that is, the confidence we have in both parameter estimates and causality—varies within each management option, and how scientists and managers can benefit from explicit consideration of cause and effect.

## Demonstrating that an outcome is caused by management—causal inference

The problem of demonstrating that a wildlife management outcome, whether beneficial or detrimental, is a consequence of management efforts, also called *actions* or *investments*, is an instance of the general scientific problem of demonstrating causality or cause and effect. Causality is based on logic and its principles of reasoning (Harre 1972, Williams 1997, Williams et al. 2002) and is a fundamental issue in science and other research (Pearl 2009, Imbens 2020). Many studies have explicitly described features that are required or desirable to demonstrate cause and effect. Those criteria include temporality (a cause occurring before its effect), experiment, a plausible mechanism, consistency (repeatability), coherence, a dose–response relationship, specificity, strength, and analogy (Hill 1965). Other evidence includes an increase in prediction bias when a possible cause is removed (Granger 1969) and convergence to a higher correlation between observed and predicted parameter values with an increased data set (Sugihara et al. 2012). Repeatability, consistency, and coherence have been examined using triangulation (Munafo and Smith 2018) to encourage greater evidence of causality from multiple sources. More basic evidence is that both a stressor (a management effort in our context) and a response (an outcome) occurred (Nichols et al. 2017).

Management that demonstrates more of these criteria can make stronger causal inferences (Cox and Wermuth 2004) about the link between efforts and outcomes; that is, managers have more confidence of a causal relationship and therefore have more reliable knowledge (Romesburg 1981). Levels of causal inference were described starting at level 0 when they were based only on association (e.g., correlation between observed quantities), level 1 when they were based on experimental results, and level 2 when they were based on a unique mechanism (Cox and Wermuth 2004). We propose a third level when inferences are also supported by validated quantitative predictions. We interpret the levels as cumulative; level 2 incorporates level 1 and so on. We recognize that wildlife managers and scientists cannot always do randomized, replicated, manipulative experiments, especially with threatened species. However, validating predictions has been encouraged in wildlife research as part of obtaining reliable knowledge (Sells et al. 2018). In other scientific disciplines—notably, astronomy and meteorology—randomized manipulative experiments are typically not possible. However, predictions can be made from hypotheses and then validated in separate observational studies (Lipton 2005, Sells et al. 2018), and these disciplines evidently produce reliable scientific knowledge. A famous example in science is the validated prediction in Einstein's theory of general relativity of light bending, by observations of stars close to the sun during a solar eclipse (Dyson et al. 1920, Nature Physics 2019). Levels 0 to 3 match particular rows in the righthand column in figure 1; row 2 is level 0, showing association in empirical graphs, row 3 combines the first and second levels, showing empirical graphs with analysis and incorporating temporality, experiment, and a mechanism; and the bottom row is our level 3, showing an empirical graph, analysis, and validated predictions. An analogous hierarchy of experimental designs used in medicine ranked them on their ability to lead from weaker to stronger causal inferences (Harris et al. 2006) and was similar to that described earlier by Manly (1992) in a general scientific context.

Adaptive management is used in wildlife management to integrate monitoring and learning into a process of improving the effectiveness and efficiency of on-ground efforts and actions (Walters and Holling 1990, Williams et al. 1996, 2002). Criteria for demonstrating adaptive management have been listed as identification of management goals and options, rigorous statistics such as including experiments, monitoring, and adjusting management in response to results (Westgate et al. 2013). These components and criteria are similar to but not identical to the causal criteria identified earlier; adaptive management is more explicit about goals and options and changes to management as information accumulates. Wildlife managers need to have confidence in the reliability of results they use to adjust management. The framework of figure 1 can help managers assess their level of confidence for decision-making and how it can be increased by changing management toward having features matching those in the figure's bottom righthand corner. In the context of wildlife management, a focus on inferring cause and effect from manipulative experimentation seems appropriate, because management implies the use of interventions rather than making passive observations alone. However, there will be situations in which observations are appropriate, most obviously prior to more active management. We now briefly describe the options.

## Option 1: Evaluation by using trends

Species and community management can assess temporal trends in abundance or related parameters by using statistical inference—namely, parameter estimation and inferences from a sample to a wider population (figure 1's left column). Trends, such as marked declines in abundance, can be used to decide whether management actions or efforts should be initiated or, if they are already underway, modified. Trends are used to assess conservation status categories, such as Endangered (IUCN 2012).

**A conceptual graph of outcomes.** A conceptual graph can describe expected trends over time (figure 2a, table 1). An interesting example is used in the Saving Our Species program of the New South Wales government (Brazill-Boast 2018, Office of Environment and Heritage 2018). The outcome metrics are described as *on track* if they are between the two curved lines in the figure, an approach borrowed from quality control charts in industry (Burgman et al. 2012). The target outcome, also called the aim, objective, or goal, is explicit. The figure is not explicit about how to interpret outcomes above the on-track lines, although presumably managers should be pleased outcomes are ahead of the time schedule. The strength is that the concept is simple and being explicit is available for all to see and use. Such a graph is called a *response to management curve* (Office of Environment and Heritage 2018, Mayfield et al. 2020); to show the response explicitly, an annotation is needed to show the point at which management commenced and preferably separate curves showing the expected outcomes with and without management should be included from that point.

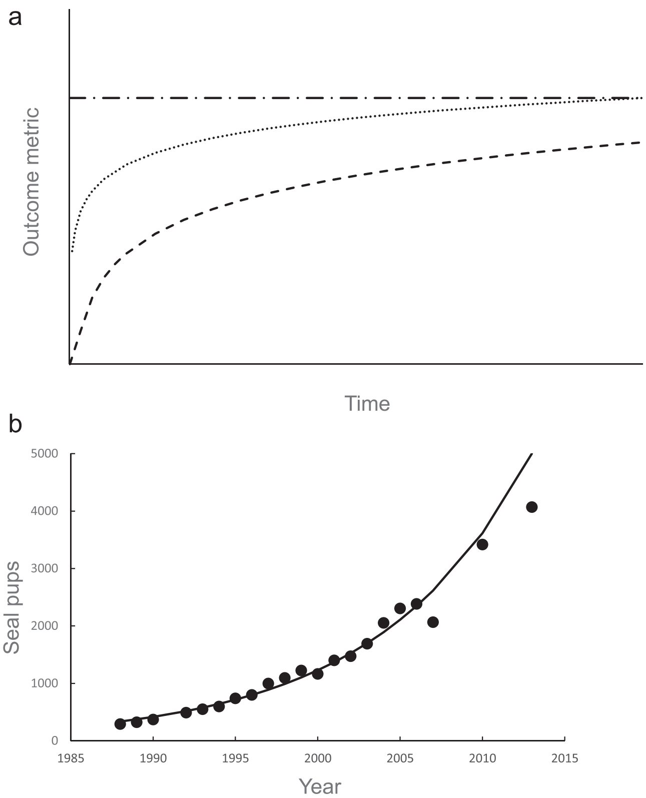


***Figure 2. Examples of graphs illustrating evaluation of trends in wildlife species. (a). Conceptual graph of outcomes over time. Outcomes are on track between the dashed line and dotted line, and the long-term target is shown by the dashed and dotted line. After Office of Environment and Heritage (2018) and Brazill-Boast (2018). (b). Annual estimates of abundance of pups of long-nosed fur seals (solid circles) in part of southern Australia and the fitted exponential population growth curve (solid line). Measures of precision are not shown in the figure but were reported in the original paper. After Shaughnessy and Goldsworthy (2015).***

**Outcomes over time with no analysis.** This is a simple graph or table of outcomes over time, such as of abundance or distribution, obtained from monitoring. It is empirical, although with no measures of precision or formal estimation of trend (figure 1); it is found in the literature with surprising frequency (Reddiex et al. 2006, Rayner et al. 2014). The latter study used a scoring system to reflect the inferential status of trends in woodland birds from casual (no analysis) to rigorous (formal analysis) and included assessing whether formal data analysis had occurred (Rayner et al. 2014). This management option highlights a gap in the threatened species conservation categories of the IUCN. The data deficient category (IUCN 2012) is appropriate when there is insufficient data to assess trend. When data are available but are not analyzed and any trends are ambiguous in the absence of analysis, then the category would be better labelled as *data or analysis deficient*. A strength is that the graph or table is simple and is explicit about the outcome measure, which can be related to a management aim or target if one has been set.

**Outcomes over time analyzed.** Some studies report trends in abundance with analysis (figure 1) and measures of precision. A single species example demonstrating a high level of precision is the positive annual instantaneous population growth ($r$ = 0.11, 95% confidence interval = 0.10–0.12) of pups (figure 2b) of long-nosed fur seals (*Arctocephalus forsteri*), a previously threatened population, at Cape du Couedic in South Australia (Shaughnessy and Goldsworthy 2015). Additional examples are shown in table 1.

The observed annual rate of population growth, $r$, and its precision can be estimated to determine whether the instantaneous rate, $r$, is negative, suggesting population decline, or less than or equal to the maximum growth rate ($r_m$); $0 < r \leq r_m$ (table 1). The latter approach is analogous to quality control charts defining boundaries of change (Manly 2001, Burgman et al. 2012). The parameter, $r_m$, can be obtained from the literature, or estimated using published methods (Duncan et al. 2007, Hone et al. 2010). The strength of the approach suggests it should be incorporated into the standard procedures and Recovery Plans for managing threatened species. Introducing an arbitrary cap on $r_m$, is undesirable, such as in the Threatened Species Index, which caps annual finite population growth ($\lambda$) at 10 (Bayraktarov et al. 2020 and their supplemental material, p. 34) independent of a species' maximum rate of growth. Threatened species can have a maximum annual finite growth rate greater than 10 and, therefore, an $r_m$ greater than 2.30 (as $\lambda = e^r$, that is $10 = e^{2.30}$; Hone et al. 2010).

**Predicted trends validated.** Quantitative predictions can be made using data collected previously, and those predictions compared with independent data (Lipton 2005, Possingham et al. 2012), including data to be obtained in the future (figure 1). Ideally, the predictions are validated—that is, shown to be unbiased (no significant difference between observed and predicted estimates)—and precise (low variability). Examples from wildlife management are shown in table 1. A strength of the option is that it provides evidence of reliable information on which to base decisions. Limitations are that the option requires high quality (unbiased and precise) monitoring data and analysis, which requires more funds, and an extended duration.

The conceptual and empirical graphs of trends in the left column of figure 1 provide no evidence that trends are caused by management efforts. The next two options generate causal inference about the effects of management.

## Option 2: Evaluation of outcomes and management efforts

A second set of options (figure 1, middle column) focuses on relationships between management outcomes and

*Table 1. Examples of wildlife studies using statistical inference and reporting outcomes as a function of time (trends) as is shown in figure 1's left column.*

| Type of data or information | Species | Sources |
|---|---|---|
| Conceptual graph | Generic not species specific | Brazill-Boast (2018), Office of Environment and Heritage (2018), Mayfield and colleagues (2020) |
| Data with analysis, single species | Helmeted honeyeater (*Lichenostomus melanops cassidix*) | Smales and colleagues (2009), Hone (2014) |
| | Long-nosed fur seal (*Arctocephalus forsteri*) | Shaughnessy and Goldsmith (2015) |
| | Imperial cormorant (*Leucocarbo atriceps*) | Yorio and colleagues (2020) |
| | Northern spotted owl (*Strix occidentalis caurina*) | Anthony and colleagues (2006), Forsman and colleagues (2011), Dugger and colleagues (2016) |
| Data with analysis, multiple species | Woodland birds | Rayner and colleagues (2014) |
| | Amphibians, reptiles, birds, mammals | Leung and colleagues (2017) |
| | Birds | Department of Environment, Food and Rural Affairs (2018), |
| | Birds | Rosenberg and colleagues (2019) |
| Data with analysis, including checking whether $r \leq r_m$ | African elephant (*Loxodonta africana*) | Caughley (1974), Foley and Faust (2010), Morrison and colleagues (2018), Louw and colleagues (2021) |
| | Yellow-footed rock wallaby (*Petrogale xanthopus*) | Sharp and colleagues (2014) |
| | Humpback whale (*Megaptera novaeangliae*) | Harrison and Woinarski (2018) |
| | Boodie (*Bettongia lesueur*) | Treloar and colleagues (2021) |
| Validated predictions | Reindeer (*Rangier tarandus*) | McCallum (2000) |
| | House mouse (*Mus domesticus*) | Krebs and colleagues (2004) |
| | Lynx (*Lynx canadensis*) | Hone and colleagues (2007) |
| | Soay sheep (*Ovis aries*) | Coulson and colleagues (2008) |

efforts and therefore involves both statistical and causal inference. The core issue in the present article was highlighted in a question asked of wildlife conservation managers by the then New Zealand prime minister in 2019 (Kiri Reihana, University of Waikato, Hamilton, New Zealand, personal communication, 3 December 2019): "How much would it cost to get a 2% increase in kiwi populations?" As for trends, the progression from conceptual graphs to data to analyzed data and to validated predictions corresponds to increasing strength of inference that the observed outcomes occurred (statistical inference) and were caused by management efforts (causal inference). The temporality causal criterion requires that management efforts must precede outcomes.

**A conceptual graph of outcomes versus management efforts.** A conceptual graph (figure 3a) relates management outcomes ($y$-axis, dependent variable) to efforts ($x$-axis, independent variable; Hone et al. 2017). Alternative management aims should also be shown. Examples are listed in table 2. A strength of the conceptual graph is its use as a planning tool, which can be revised as a program proceeds. It shows management levels including no management explicitly (figure 3a). Time is not shown explicitly, although this can be indicated by labelling.

**Outcomes and management effort reported with no analysis.** Measurements of outcomes at different levels of effort are graphed (figure 1). A simplified version includes a comparison of efforts and no efforts. A strength is that the graph has empirical data over a range of differing management efforts, which includes no management (an experimental control). A limitation is the lack of statistical analysis.

**Outcomes and management efforts reported and analyzed.** Desired outcomes, such as abundance, or its change ($r$) over a defined period, are related to management efforts through both graphing and analysis (figure 1). Examples are listed in table 2, demonstrating such an approach is feasible. Data to estimate an effort–outcome relationship can be observational, using natural experiments, such as in the rhinoceros (figure 3b) and elephant examples (Leader-Williams and Albon 1988). However, to provide stronger causal inference manipulative experiments can be used, such as in the five field experiments of wildlife management reported by Hone and colleagues (2017, their table 2). A strength of the option is that the extent

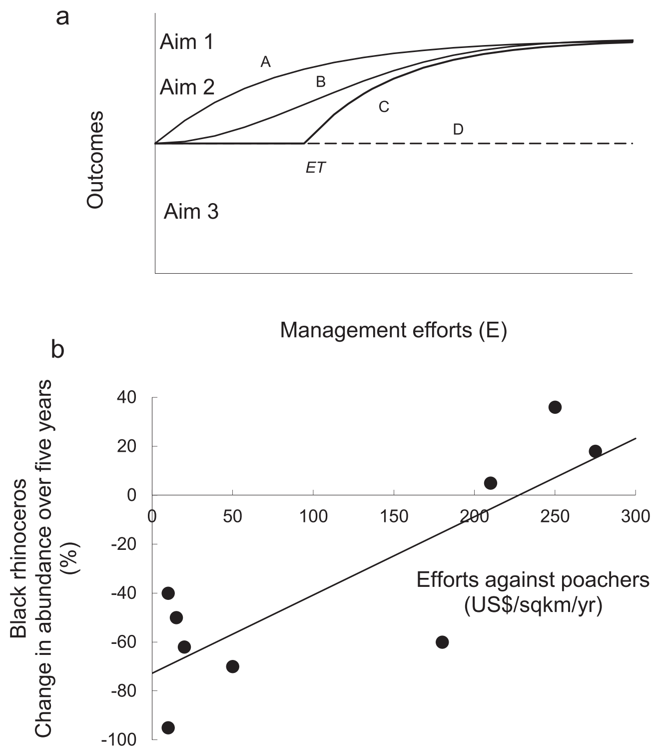


***Figure 3. Examples of graphs illustrating evaluation of effort–outcome relationships in wildlife management. (a) Conceptual graph of possible cause and effect relationships (A, B, and C represent alternative hypotheses) between management outcomes and efforts. Alternative C has a threshold level of effort (ET) before outcomes change. The intercept on the y axis corresponds to the outcome when there is no management, and no response to management efforts (the null hypothesis) is shown as the horizontal dashed line (D). Three alternative management aims are shown. Modified from Hone and colleagues (2017). (b). The percentage change in abundance of black rhinoceros over 5 years in nine African countries with different mean annual levels of effort against poachers ($R^2$ = 0.68). Source: Adapted from Leader-Williams and Albon (1988).***

to which desired management outcomes or targets are related to management efforts is explicit. Once an empirical relationship is estimated, the threshold level of effort required for population growth can be estimated from the intercept on the *x*-axis corresponding to no change in abundance.

**Predicted outcomes are validated.** Observed outcomes, such as abundance, are graphed and analyzed to predict effects of future management and these quantitative predictions are compared with independent data (figure 1). A strength is that the validation of predictions gives strong evidence of cause and effect (Cox and Donnelly 2011) and increases the confidence of managers in their choice of a level of management to achieve their aims.

## Option 3: Evaluation of outcomes related to both trends and management efforts

A third set of options for evaluating species management measures trends in outcomes relative to differing management efforts (figure 1 right column). Like option 2, it incorporates both statistical and causal inference. What makes option 3 different from option 2 is that it incorporates variation in both the time and the effort dimensions.

**Conceptual graph of outcomes over time with differing management.** A simple component of an adaptive management program, with management efforts at one site and no management efforts at another site, after which efforts are switched, is illustrated in figure 4a. Later, the more successful management approach is adopted at both sites, an example of learning in active adaptive management (figure 4a). The alternative lines in figure 4a are models of different hypotheses and are therefore testable predictions of the effects of different management efforts or actions. If only one population can be managed then different levels of effort can be applied sequentially. Over any time period, the rate of population growth can be compared to the maximum possible growth rate ($r_m$; figure 4a, dotted line). The management aim is explicit (figure 4a, the horizontal dashed and double-dotted line) and can be shown in the conceptual graph as a line, implying a fixed-point aim such as a population size of 3500, or as an interval, such as between 3000 and 4000, as was described in general terms by Mayfield and colleagues (2020). Examples are listed in table 3.

**Outcomes over time for sites with differing management and no analysis.** Empirical data of trends in outcomes with differing efforts are graphed (figure 1) or tabulated. A strength is that the observed outcomes can be compared visually to the desired outcome or target. A limitation is that there is no recognition of precision and no statistical analysis to estimate the trends and significance of association. This option corresponds to level 0 of causality of Cox and Wermuth (2004) and the casual inference of Rayner and colleagues (2014).

**Outcomes over time for sites with differing management and analyzed.** This can range from a simple experimental design (figure 1), such as before–after–control–impact (BACI) through to a randomized, replicated, and controlled experiment, depending on the resources available. Examples are listed in table 3, including the responses of yellow-footed rock wallabies (*Petrogale xanthopus*) to control of red fox (*Vulpes vulpes*; figure 4b; Sharp et al. 2014). If only one population occurs the sequential management outcomes could be evaluated by a regression discontinuity design (Manly 1992, Butsic et al. 2017). If the costs of achieving the increases are recorded, then the return on investment (Murdoch et al. 2007) and diminishing returns on investment (Grantham et al. 2008, Hone 2013) could be estimated. This option corresponds to levels 1 and 2 of causality of Cox and Wermuth (2004).

*Table 2. Examples of wildlife studies using statistical and causal inference and reporting outcomes as a function of management efforts as is shown in figure 1's middle column.*

| Type of data or information | Species and topics | Sources |
|---|---|---|
| Conceptual graph | Species-investment curve | Wilson and colleagues (2007) |
| | Benefits and costs of duck hunting and geographic scale | Johnson and colleagues (2015) |
| | Effort–outcomes relationship | Hone and colleagues (2017) |
| Data with analysis, single species | Black rhinoceros (*Diceros bicornis*) and African elephant (*Loxodonta africana*) and antipoaching efforts | Leader-Williams and Albon (1988) |
| | Yellow-eyed penguin (*Megadyptes antipodes*) and management efforts | Busch and Cullen (2009) |
| Data with analysis, multiple species | Bird species richness and feral pig (*Sus scrofa*) control | Hone (2012) |
| Validated predictions | Additive and compensatory hunting mortality models for mallard (*Anas platyrhynchos*) | Nichols (1991) |

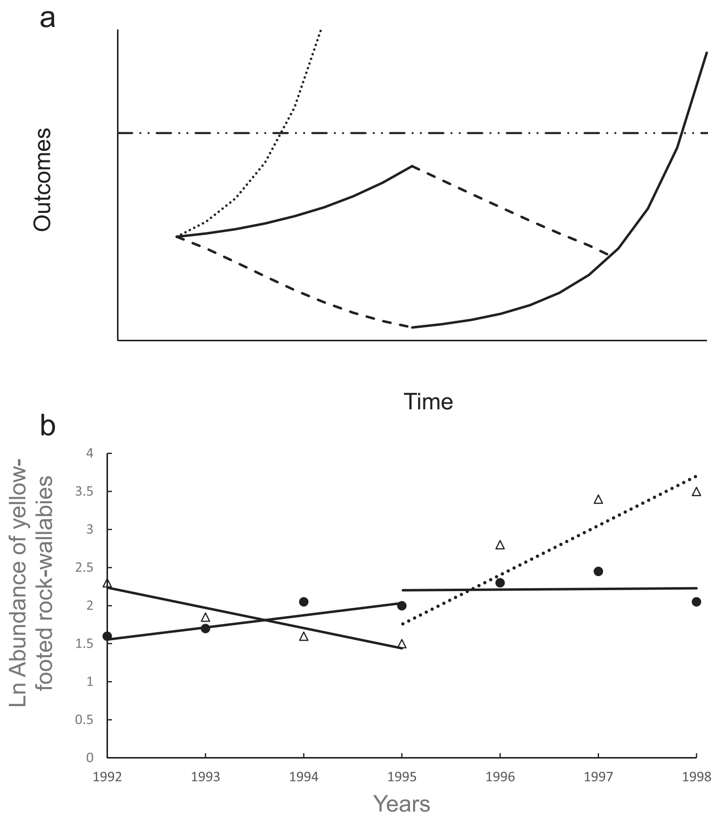


***Figure 4. Examples of graphs illustrating evaluation of trends in effort–outcome relationships in wildlife management. (a). A conceptual graph of adaptive management outcomes over time, showing a scenario in which initially management efforts or actions are applied (solid lines) or not applied (dashed lines) simultaneously, then management is crossed over, and then later both sites are managed. Management can be applied sequentially, as in mallard management in North America, and that is shown by following one line from left to right. For illustrative purposes, exponential population growth and decline, and no diminishing returns are assumed. Maximum population growth ($r_m$) is shown as the dotted line, and the management aim as the horizontal dashed and double dotted line. (b) Trends in abundance indices (on a ln = $\log_e$ scale) of yellow-footed rock wallabies at two sites (Gap Range, the solid circles, and Coturaundee Range, the open triangles) before fox control (1992 to 1995). These trends were not significantly different. Trends were significantly different during 1995 to 1998 at the one site without fox control (Gap Range, the solid circles) compared with the one site with fox control (Coturaundee Range, the open triangles). The fitted regression lines show trends without (the solid lines) and with (the dotted line) fox control. Source: Adapted from Sharp and colleagues (2014).***

*Table 3. Examples of wildlife studies using statistical and causal inference and reporting management outcomes as a function of both time and management efforts as is shown in figure 1's right column.*

| Type of data or information | Species and topics | Source |
|---|---|---|
| Conceptual graph | Wildlife with/without harvest | Walters and Hilborn (1978) |
| | Duck dynamics and harvest versus no harvest | Frith (1979 |
| | Conservation of forest birds | Green (2002) |
| | Conservation legacy, gain, and recovery potential | Akcakaya and colleagues (2018) |
| Data with analysis, single species | Black-flanked rock wallaby (*Petrogale lateralis*) and red fox (*Vulpes vulpes*) control | Hone (1994),<br>Kinnear and colleagues (1998) |
| | Kokako (*Callaeas cinerea wilsoni*) and pest control | Innes and colleagues (1999) |
| | Bovine TB in cattle and badger (*Meles meles*) control | Donnelly and colleagues (2006),<br>Jenkins and colleagues (2010) |
| | Grand skink (*Oligosoma grande*) and vertebrate pest control | Norbury and colleagues (2014) |
| | Northern spotted owl and barred owl (*Strix varia*) removal | Diller and colleagues (2016),<br>Dugger and colleagues (2016) |
| | Hihi (*Notiomystis cincta*) and parasite control | Mather and colleagues (2021) |
| | African elephant during and after culling | Louw and colleagues (2021) |
| Data with analysis, multiple species | Trends in native forest birds and vertebrate pest control | Innes and colleagues (2010) |
| Validated predictions | Mallard and harvesting | Nichols and colleagues (2015, 2019)<br>Zhao and colleagues (2016) |
| | Rhinoceros (*Ceratotherium simum simum*) and poaching | Haas and Ferreira (2016) |

**Predicted outcomes in trends of differing management are validated.** The trends of outcomes with differing management are compared with the management aim, and quantitative predictions made from the analysis are validated (figure 1). Repeated demonstrations that predictions are unbiased strengthens causal inferences. An example is of North American midcontinental mallards (*Anas platyrhynchos*), whose abundance varies over years as inferred from annual population estimates and their 95% confidence intervals (table 3; Nichols et al. 2015, 2019). Predictions were generated by alternative hypotheses including of the effects of hunting on survival rates in an adaptive management framework with hunting varying sequentially across years. Therefore, the mallard management is similar to the sequential variation of management illustrated in figure 4a, but without the simultaneous management. The sequential implementation of alternative management actions, predictions, evaluations and therefore learning, is a Bayesian approach that could be applied to management of other species. An extra evaluation of mallard management could estimate observed annual growth rates, such as $\lambda$, because these can be used as metrics of population health and to assess the effects of management actions (Nichols and Hines 2002). A limitation is that this approach may be too demanding of resources for small or inadequately funded programs.

## Discussion

We have described how wildlife management, including conservation, can be evaluated using three options—namely, outcomes over time (figure 1, lefthand column), outcomes relative to effort (figure 1, middle column), and a combination of these (figure 1, righthand column). Across and within each option, we recognize different strengths of statistical and causal inference and these lead to identification of the favored approach described in the bottom righthand corner of figure 1—data analyzed to estimate trends and experimental estimation of effects of management efforts with validation of predictions. We encourage adaptive wildlife management; however, criteria for demonstrating such management are not identical to those proposed for demonstrating strong causal inference, so we recommend that wildlife managers link their efforts and outcomes to explicit causal criteria.

Our classification of evaluation options according to time and effort and the consideration of both statistical and causal inference help to reveal the strengths and limitations of each. Like Lipton (2005) and Possingham and colleagues (2012), we recommend greater generation and validation of predictions as occurs in other scientific disciplines, such as astronomy and meteorology, where scope for adaptive manipulative experiments is limited. Validating predictions can occur across times at one site, across sites at one time, or a combination. Implementation of options incorporating time and effort dimensions (bottom righthand corner of figure 1) in future wildlife management will increase the reliability of knowledge (Romesburg 1981, Sells et al. 2018) and support a transition from “no clear evidence” to “clear evidence” (Garnett et al. 2019) of management effectiveness. Implementing the stronger options also supports the learning needed for more effective adaptive management, and can require greater cooperation between scientists and managers, which we encourage. The empirical examples presented of mammals, birds, reptiles and amphibians in a wide range of locations around the world show the options are feasible and have broad applicability. We suggest the framework has application to other related disciplines, such as fisheries management.

Our classification of evaluation options can be drawn on by wildlife managers and scientists to support their decision-making in several ways. They can identify which approach they have been using and choose which option to use in future for stronger statistical and causal inference. They can determine whether full use has been made of the monitoring data collected over time, in particular through the estimation of confidence levels and intervals that form the basis of sound decision-making. They can assess whether the program has a clear, quantifiable aim (Tear et al. 2005) and evaluate whether it is achieved. They can predict the outcomes of management and use a different data set to assess predictive bias. They can assess the feasibility of the options described in the present article and, if needed, seek additional resources, in order to strengthen inferences about trends and management effectiveness. They can identify the consequences of not adopting an option in terms of reduced strength of inference and less well-informed adaptive management. In conclusion, we recommend incorporating the framework of figure 1 and its formal process of choosing an evaluation option into the procedures of wildlife management including conservation.

## Acknowledgments

We thank the Institute for Applied Ecology, the University of Canberra, the Australasian Wildlife Management Society and the Southern African Wildlife Management Association for support. Sally Box, Peter Bridgewater, Nick Dexter, Sue Nichols, Astrida Upitis, Rudi van Aarde, and anonymous referees are thanked for useful discussions, comments and suggestions. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. JH received travel and conference funds from the Australasian Wildlife Management Society and the Southern African Wildlife Management Association, which are gratefully acknowledged.

## Disclosure statement

There are no competing interests to declare.

## References cited

Akcakaya HR et al. 2018. Quantifying species recovery and conservation success to develop an IUCN green list of species. Conservation Biology 32: 1128–1138.

Anthony RG et al. 2006. Status and trends in demography of northern spotted owls, 1985–2003. Wildlife Monographs 163: 1–48.

Australian government. 2021. Threatened Species Strategy 2021–2031. Commonwealth of Australia.

Australian government. 2022. Threatened Species Action Plan towards Zero Extinctions 2022–2032. Commonwealth of Australia.

Bayraktarov E et al. 2020. A threatened species index for Australian birds. Conservation Science and Practice 3: e322.

Brazill-Boast J. 2018. Saving our species: A cost-effective, large-scale monitoring and evaluation program for threatened species. Pages 225–238 in Legge S, Lindenmayer DB, Robinson NM, Scheele BC, Southwell DM, Wintle BA. eds. Monitoring Threatened Species and Ecological Communities. CSIRO.

Burgman M et al. 2012. An endpoint hierarchy and process control charts for ecological monitoring. Pages 71–78 in Lindenmayer D, Gibbons P. eds. Biodiversity Monitoring in Australia. CSIRO.

Busch J, Cullen R. 2009. Effectiveness and cost-effectiveness of yellow-eyed penguin recovery. Ecological Economics 68: 762–776.

Butsic V, Lewis DJ, Radeloff VC, Baumann M, Kuemmerle T. 2017. Quasi-experimental methods enable stronger inferences from observational data in ecology. Basic and Applied Ecology 19: 1–10.

Caughley G. 1974. Bias in aerial survey. Journal of Wildlife Management 38: 921–933.

Caughley G. 1980. Analysis of Vertebrate Populations, reprinted with corrections. Wiley.

Caughley G. 1994. Directions in conservation biology. Journal of Animal Ecology 63: 215–244.

Coulson T et al. 2008. Estimating the functional form for the density dependence from life history data. Ecology 89: 1661–1674.

Cox DR, Donnelly CA. 2011. Principles of Applied Statistics. Cambridge University Press.

Cox DR, Wermuth N. 2004. Causality: A statistical view. International Statistical Reviews 72: 285–305.

Department of Environment, Food and Rural Affairs. 2018. Wild Bird Populations in the UK, 1970 to 2017. UK Department of Environment, Food and Rural Affairs.

Diller LV et al. 2016. Demographic response of northern spotted owls to barred owl removal. Journal of Wildlife Management 80: 691–707.

Donnelly CA et al. 2006. Positive and negative effects of widespread badger culling on tuberculosis in cattle. Nature 439: 843–846.

Dugger KM et al. 2016. The effects of habitat, climate, and barred owls on long-term demography of northern spotted owls. Condor 118: 57–116.

Duncan RP, Forsyth DM, Hone J. 2007. Testing the metabolic theory of ecology: Allometric scaling exponents in mammals. Ecology 88: 324–333.

Dyson FW, Eddington AS, Davidson C. 1920. A determination of the deflection of light by the sun's gravitational field, from observations made at the total eclipse of May 29, 1919. Philosophical Transactions of the Royal Society A 220: 291–333.

Foley CAH, Faust LJ. 2010. Rapid population recovery in an elephant population *Loxodonta africana* recovering from poaching in Tarangire National Park Tanzania. Oryx 44: 205–212.

Forsman ED et al. 2011. Population Demography of Northern Spotted Owls. Studies in Avian Biology no. 40. University of California Press.

Frith HJ. 1979. Wildlife Conservation, revised ed. Angus and Robertson.

Garnett S, Latch P, Lindenmayer D, Woinarski J. 2018. Recovering Australian Threatened Species. A Book of Hope. CSIRO.

Garnett ST et al. 2019. Metrics of progress in the understanding and management of threats to Australian birds. Conservation Biology 33: 456–468.

Grace MK et al. 2021. IUCN launches Green Status of Species: A new standard for species recovery. Oryx 55: 651–652.

Granger CWJ. 1969. Investigating causal relations by using econometric models and cross-spectral methods. Econometrica 37: 424–438.

Grantham HS, et al. 2008. Diminishing returns on investment for biodiversity data in conservation planning. Conservation Letters 1: 190–198.

Green R. 2002. Diagnosing causes of population declines and selecting remedial actions. Pages 139–156 in Norris K, Pain DJ, eds. Conserving Bird Diversity: General Principles and Their Application. Cambridge University Press.

Haas TC, Ferreira SM. 2016. Combating rhino horn trafficking: The need to disrupt criminal networks. PLOS ONE 11: e0167040.

Harre R. 1972. The Philosophies of Science: An Introductory Survey. Oxford University Press.

Harris AD et al. 2006. The use and interpretation of quasi-experimental studies in medical informatics. Journal of the American Informatics Association 13: 16–23.

Harrison PL, Woinarski JCZ. 2018. Recovery of Australian subpopulations of humpback whale. Pages 5–12 in Garnett S, Latch P, Lindenmayer D, Woinarski J, eds. Recovering Australian Threatened Species: A Book of Hope. CSIRO.

Hill AB. 1965. The environment and disease: Association or causation? Proceedings of the Royal Society of Medicine 58: 295–300.

Hone J. 1994. Analysis of Vertebrate Pest Control. Cambridge University Press.

Hone J. 2012. Applied Population and Community Ecology: The Case of Feral Pigs in Australia. Wiley Blackwell.

Hone J. 2013. Diminishing returns in bovine tuberculosis control. Epidemiology and Infection 141: 1382–1389.

Hone J. 2014. Estimating wildlife population trends: The case of the helmeted honeyeater. Emu 114: 191–196.

Hone J, Krebs C, O'Donoghue M, Boutin S. 2007. Evaluation of predator numerical responses. Wildlife Research 34: 335–341.

Hone J, Duncan RP, Forsyth DM. 2010. Estimates of maximum annual population growth rates ($r_m$) and their application in wildlife management. Journal of Applied Ecology 47: 507–514.

Hone J, Drake VA, Krebs CJ. 2017. The effort-outcomes relationship in applied ecology: Evaluation and implications. BioScience 67: 845–852.

Hone J, Drake VA, Krebs CJ. 2018. Evaluating wildlife management by using principles of applied ecology: Case studies and implications. Wildlife Research 45: 436–445.

Imbens GW. 2020. Potential outcome and directed acyclic graph approaches to causality: Relevance for empirical practice in economics. Journal of Economic Literature 58: 1129–1179.

Innes J et al. 1999. Successful recovery of North Island kokako *Callaeas cinerea wilsoni* populations, by adaptive management. Biological Conservation 87: 201–214.

Innes J, Kelly D, Overton JMc, Gillies C. 2010. Predation and other factors currently limiting New Zealand forest birds. New Zealand Journal of Ecology 34: 86–114.

[IUCN] International Union for Conservation of Nature. 2012. Red List Categories and Criteria, ver. 3.1, 2nd ed. IUCN Species Survival Commission.

Jenkins HE, Woodroffe R, Donnelly CA. 2010. The duration of the effects of repeated widespread badger culling on cattle tuberculosis following the cessation of culling. PLOS ONE 5: e9092.

Johnson FA et al. 2015. Multilevel learning in the adaptive management of waterfowl harvests. 20 years and counting. Wildlife Society Bulletin 39: 9–19.

Joseph LN, Maloney RF, Possingham HP. 2009. Optimal allocation of resources among threatened species: A project prioritization protocol. Conservation Biology 23: 328–338.

Kinnear JE, Onus ML, Sumner NR. 1998. Fox control and rock-wallaby population dynamics. II. An update. Wildlife Research 25: 81–95.

Krebs CJ et al. 2004. Can outbreaks of house mice in south-eastern Australia be predicted by weather models? Wildlife Research 31: 465–474.

Leader-Williams N, Albon S. 1988. Allocation of resources for conservation. Nature 336: 533–535.

Legge S et al. 2018. Monitoring Threatened Species and Ecological Communities. CSIRO.

Leung B, Greenberg DA, Green DM. 2017. Trends in mean growth and stability in temperate vertebrate populations. Diversity and Distributions 23: 1372–1380.

Lindenmayer D, Gibbons P. 2012. Biodiversity Monitoring in Australia. CSIRO.

Lipton P. 2005. Testing hypotheses: Predictions and prejudice. Science 307: 219–221.

Louw AS et al. 2021. Elephant population responses to increased density in Kruger National Park. Koedoe 63: a1660.

Manly BFJ. 1992. The Design and Analysis of Research Studies. Cambridge University Press.

Manly BFJ. 2001. Statistics for Environmental Science and Management. Chapman and Hall.

Mather E et al. 2021. Testing management alternatives for controlling nest parasites in an endangered bird. Animal Conservation 24: 580–588.

Mayfield HJ, et al. 2020. Estimating species response to management using an integrated process: A case study from New South Wales, Australia. Conservation Science and Practice 2: e269.

McCallum H. 2000. Population Parameters: Estimation for Ecological Models. Blackwell.

Morrison TA, et al. 2018. Informing aerial total counts with demographic models: Population growth of Serengeti elephants not explained purely by demography. Conservation Letters 11: 1–8.

Munafo MR, Smith GD. 2018. Repeating experiments is not enough. Nature 553: 399–401.

Murdoch W et al. 2007. Maximising return on investment in conservation. Biological Conservation 139: 357–388.

Nature Physics. 2019. A century of correct predictions. Nature Physics 15: 415.

Nichols JD. 1991. Responses of North American duck populations to exploitation. Pages 498–525 in Perrins CM, Lebreton JD, Hirons GJM, eds. Bird Population Studies: Relevance to Conservation and Management. Oxford University Press.

Nichols JD, Hines JE. 2002. Approaches for the direct estimation of $\lambda$, and demographic contributions to $\lambda$, using capture-recapture data. Journal of Applied Statistics 29: 539–568.

Nichols JD, Johnson FA, Williams BK, Boomer GS. 2015. On formally integrating science and policy: Walking the walk. Journal of Applied Ecology 52: 539–543.

Nichols JD, Kendall WL, Boomer GS. 2019. Accumulating evidence in ecology: Once is not enough. Ecology and Evolution 9: 13991–14004.

Nichols SJ, Peat M, Webb JA. 2017. Challenges for evidence-based environmental management: What is acceptable and sufficient evidence of causation? Freshwater Science 36: 240–249.

Norbury G, Hutcheon A, Reardon J, Daigneault A. 2014. Pest fencing or pest trapping: A bio-economic analysis of cost-effectiveness. Austral Ecology 39: 795–807.

Office of Environment and Heritage. 2018. Saving our Species. Saving our Species Monitoring, Evaluation and Reporting. Guidelines for Conservation Projects. New South Wales Office of Environment and Heritage.

Pearl J. 2009. Causal inference in statistics: An overview. Statistical Surveys 3: 96–146.

Possingham HP, Wintle BA, Fuller RA, Joseph LN. 2012. The conservation return on investment from ecological monitoring. Pages 49–61 in Lindenmayer D, Gibbons P, eds. Biodiversity Monitoring in Australia. CSIRO.

Rayner L, Lindenmayer DB, Gibbons P, Manning A. 2014. Evaluating empirical evidence for decline in temperate woodland birds: A nationally threatened assemblage of species. Biological Conservation 171: 145–155.

Reddiex B, Forsyth DM. 2006. Control of pest mammals for biodiversity protection in Australia: II. Reliability of knowledge. Wildlife Research 33: 711–717.

Reddiex B et al. 2006. Control of pest mammals for biodiversity protection in Australia: I. Patterns of control and monitoring. Wildlife Research 33: 691–709.

Romesburg HC. 1981. Wildlife science: Gaining reliable knowledge. Journal of Wildlife Management 45: 293–313.

Rosenberg KV, et al. 2019. Decline of the North American avifauna. Science 366: 120–124.

Sells SN et al. 2018. Increased scientific rigor will improve reliability of research and effectiveness of management. Journal of Wildlife Management 82: 485–494.

Sharp A et al. 2014. Population recovery of the yellow-footed rock-wallaby following fox control in New South Wales and South Australia. Wildlife Research 41: 560–570.

Shaughnessy PD, Goldsworthy SD. 2015. Increasing abundance of pups of long-nosed fur seals (*Arctocephalus forsteri*) on Kangaroo Island, South Australia, over 26 breeding seasons to 2013–14. Wildlife Research 42: 619–632.

Smales IJ, Quin B, Menkhorst PW, Franklin DC. 2009. Demography of the helmeted honeyeater (*Lichenostomus melanops cassidix*). Emu 109: 352–359.

Speirs-Bridge A et al. 2010. Reducing overconfidence in the interval judgements of experts. Risk Analysis 30: 512–523.

Sugihara G et al. 2012. Detecting causality in complex ecosystems. Science 338: 496–500.

Tear TH et al. 2005. How much is enough? The recurrent problem of setting measurable objectives in conservation. BioScience 55: 835–849.

Treloar S, Lohr C, Hopkins AJM, Davis RA. 2021. Rapid population expansion of boodie (burrowing bettong, *Bettongia lesueur*) creates potential for resource competition with mala (rufous hare-wallaby, *Lagorchestes hirsutus*). Ecological Management and Restoration 22: 54–57.

Walters CJ, Hilborn R. 1978. Ecological optimization and adaptive management. Annual Review of Ecology and Systematics 9: 157–188.

Walters CJ, Holling CS. 1990. Large-scale management experiments and learning by doing. Ecology 71: 2060–2068.

Westgate MJ, Likens GE, Lindenmayer DB. 2013. Adaptive management of biological resources: A review. Biological Conservation 158: 128–139.

Williams BK. 1997. Logic and science in wildlife biology. Journal of Wildlife Management 61: 1007–1015.

Williams BK, Johnson FA, Wilkins K. 1996. Uncertainty and the adaptive management of waterfowl harvests. Journal of Wildlife Management 60: 223–232.

Williams BK, Nichols JD, Conroy MJ. 2002. Analysis and Management of Animal Populations. Modeling, Estimation and Decision Making. Academic Press.

Williams DR, Balmford A, Wilcove DS. 2020. The past and future of conservation science in saving biodiversity. Conservation Letters 13: e12720.

Wilson KA et al. 2007. Conserving biodiversity efficiently: What to do, where, and when. PLOS Biology 5: e223.

Yorio P et al. 2020. Population trends of imperial cormorants (*Leucocarbo atriceps*) in northern coastal Argentine Patagonia over 26 years. Emu Austral Ornithology 120: 114–122.

Zhao Q, Silverman E, Fleming K, Boomer GS. 2016. Forecasting waterfowl population dynamics under climate change: Does the spatial variation of density-dependence and environmental effects matter? Biological Conservation 194: 80–88.

---

*Jim Hone (jim.hone@canberra.edu.au), Alistair Drake (alistair.drake@canberra.edu.au), and Charles Krebs (krebs@zoology.ubc.ca) are researchers in the Institute for Applied Ecology at the University of Canberra, in Canberra, Australian Capital Territory, Australia. Drake and Krebs also have appointments at the University of New South Wales, in Canberra, Australian Capital Territory, Australia, and at the University of British Columbia, in Vancouver, British Columbia, Canada, respectively.*