Ecology Letters, (2004) 7: 669–675 doi: 10.1111/j.1461-0248.2004.00625.x

REPORT

Minimizing the cost of environmental management decisions by optimizing statistical thresholds

Scott A. Field,[1]* Andrew J. Tyre,[1] Niclas Jonzén,[1] Jonathan R. Rhodes[1,2] and Hugh P. Possingham[1]

Abstract

Environmental management decisions are prone to expensive mistakes if they are triggered by hypothesis tests using the conventional Type I error rate (α) of 0.05. We derive optimal α-levels for decision-making by minimizing a cost function that specifies the overall cost of monitoring and management. When managing an economically valuable koala population, it shows that a decision based on $\alpha = 0.05$ carries an expected cost over \$5 million greater than the optimal decision. For a species of such value, there is never any benefit in guarding against the spurious detection of declines and therefore management should proceed directly to recovery action. This result holds in most circumstances where the species' value substantially exceeds its recovery costs. For species of lower economic value, we show that the conventional α-level of 0.05 rarely approximates the optimal decision-making threshold. This analysis supports calls for reversing the statistical 'burden of proof' in environmental decision-making when the cost of Type II errors is relatively high.

Keywords

Koala, management, optimal monitoring, statistical power, statistical significance, Type I error, Type II error.

Ecology Letters (2004) 7: 669–675

INTRODUCTION

Monitoring data play a central role in informing the decisions of those charged with managing the environment. Active adaptive management (Shea *et al.* 2002) presupposes a monitoring regime that is capable of detecting environmental trends in spite of the levels of process and observation uncertainty inherent in the system. In other words, it implies that analysis of monitoring data will achieve sufficient statistical power to enable informed decision-making. But what exactly is 'sufficient'? In the frequentist statistical paradigm, the traditional answer has been found by adherence to the 'five-eighty convention' (Di Stefano 2003), in which the significance level (α, the chance of making a false positive, or 'Type I' error) is fixed at 0.05 and a non-significant result regarded as definitive if the resultant statistical power ($1 - \beta$, the chance of avoiding a false negative, or 'Type II' error) is 0.8 or higher, at a specified effect size. In practice, however, this power target is very rarely achieved (Sedlmeier & Gigerenzer 1989; Anderson *et al.* 2000; Jennions & Moller 2003).

Even if this convention were strictly followed, numerous authors have noted that it is likely to have disastrous consequences for the environment (Gray 1990; Peterman 1990a,b; Peterman & M'Gonigle 1992; Taylor & Gerrodette 1993; Mapstone 1995; Keough & Mapstone 1997; Dayton

[1]*The Ecology Centre, University of Queensland, St Lucia 4072, Queensland, Australia*
[2]*School of Geography, University of Queensland, Planning and Architecture, St Lucia 4072, Queensland, Australia*
**Correspondence and Present address: Dr Scott A. Field, Department of Environmental Biology, Benham Building, University of Adelaide, North Terrace 5005, SA, Australia*
E-mail: scott.field@adelaide.edu.au
Present address: Andrew J. Tyre, School of Natural Resource Sciences, University of Nebraska-Lincoln, Lincoln, NE 68583-0819, USA
Present address: Niclas Jonzén, Department of Theoretical Ecology, Ecology Building, Lund University, SE-223 62 Lund, Sweden

2001; Di Stefano 2001). The reason is that failing to detect an environmental effect (a Type II error) may result in serious damage to the environment that is long-term and/or irreversible, such as the collapse of fish stocks (Peterman 1990b; Dayton 2001), the extinction of threatened species (Taylor & Gerrodette 1993) or the pollution of water supplies (Mapstone 1995; Keough & Mapstone 1997; Di Stefano 2003). On the other hand, mistakenly concluding there is an effect (a Type I error) will usually cause relatively minor short-term economic impacts (Dayton 2001). Therefore, it has been argued that the statistical 'burden of proof' (Gray 1990; Dayton 2001), traditionally weighted in favour of avoiding spurious effects, should be adjusted by environmental managers to ensure that real instances of environmental damage are not overlooked.

Given this imperative, the literature on how to do so is surprisingly sparse and its application virtually non-existent. Although several authors have proposed methods for balancing the burden of proof (Nagel & Neef 1977; Cascio & Zedeck 1983; Mapstone 1995; Power *et al.* 1995; Fox 2001; Murphy & Myors 2004), the problem has yet to be formulated fully and correctly, and no such method has been widely implemented in published studies (Appendix A). There remains a clear need to formulate the problem in a decision-theoretic framework that sets an appropriate objective (minimizing the overall cost of the management decision taken), and uses the relative costs of Type I and Type II errors to find statistical thresholds that meet this objective.

Here we achieve this by specifying and minimizing an expression for the expected overall cost of a management decision triggered by a hypothesis test. We apply it to a case study of managing an economically valuable threatened species, the koala (*Phascolarctos cinereus*) in eastern Australia, and show that applying conventional statistical practice would result in an expected economic loss in excess of \$5 million. We also examine the general implications for less economically valuable species and identify a range of error cost ratios for which optimal α-values exist. Our results have implications for managers from a broad range of applied scientific disciplines, including fisheries, conservation biology, forestry and environmental impact assessment.

MATERIAL AND METHODS

An expression for the cost of a management decision requires four main components: (1) an estimate of the probability that a deleterious change (in our case study, a regional koala population decline) has occurred; (2) the probability that analysis of monitoring data will correctly diagnose whether that change has occurred; (3) the monetary costs of actions triggered by the conclusions of the analysis (sometimes referred to as 'utility' or 'loss'); and (4) specification of a relationship between α and β. By constructing a decision tree that incorporates these components (Fig. 1), multiplying down the branches and summing across their termini, we can derive an expression for the overall expected cost of monitoring and management:

$$E[C] = (1-\delta)\alpha R + \delta[(1-\beta)R + \beta V] + M \quad (1)$$

where $E[C]$ is the expected total cost of monitoring and management, δ represents prior expectation of a deleterious change occurring, α is the Type I error rate (probability of falsely detecting a change), β is the Type II error rate (probability of missing a real change), R is the cost of recovery action, V is economic loss associated with a deleterious change, and M is the economic cost of carrying out monitoring.

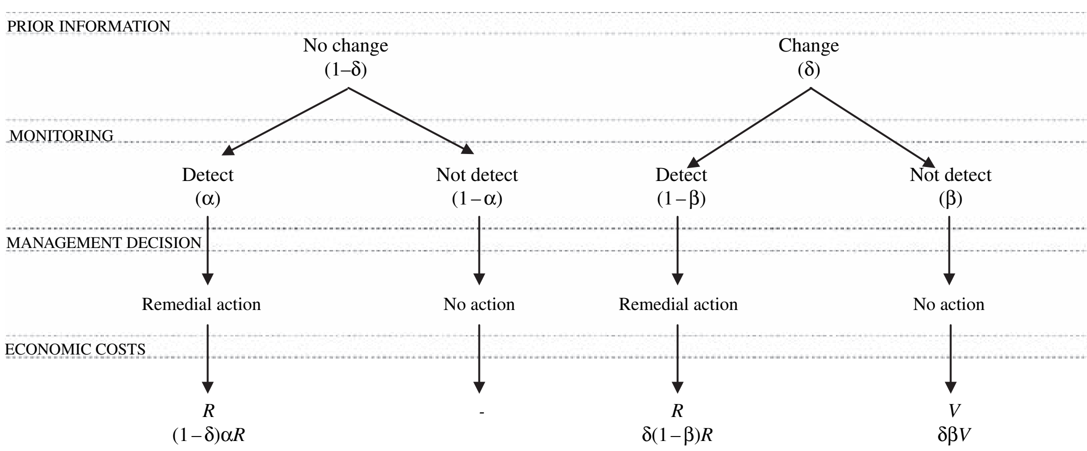

Figure 1 Decision tree for calculating the total expected cost of an environmental management decision based on monitoring data. δ, probability that a deleterious change has occurred; α, Type I error rate; β, Type II error rate; R, cost of remedial action; V, economic value of environmental resource lost or damaged as a result of an undetected change.

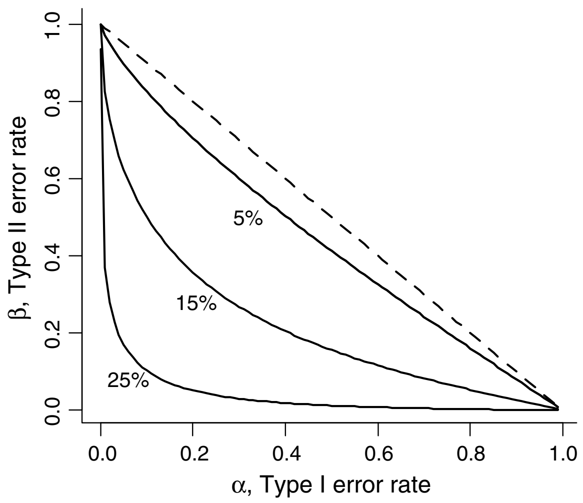

Figure 2 Relationship between β and α generated by fitting a zero-inflated binomial model to simulated presence–absence monitoring data. Parameter settings were occupancy (p) = 0.5, detectability (q) = 0.5, number of sites sampled (n) = 100, and number of repeat visits to sites (m) = 3. The curves represent three levels of decline (d): 5, 15 and 25%. The dashed line indicates the line $\alpha = 1 - \beta$, the limiting state for the relationship as effect size approaches zero.

In order to evaluate this cost function, a relationship between α and β must be specified, which in turn requires that an *a priori* power analysis be performed. This relationship may vary with different statistical models, target effect sizes and sample sizes, but retains the same general form, with a negative, decreasing slope and $\alpha < 1 - \beta$ (Fig. 2). We used a relationship derived from fitting a zero-inflated binomial (ZIB) model (Hall 2000; MacKenzie *et al.* 2002; Tyre *et al.* 2003) to simulated presence–absence-monitoring data for a declining population (Appendix B).

We now illustrate the application of the cost function to management of the koala, an iconic marsupial that is declining in parts of eastern Australia (Lunney *et al.* 2000). A major remaining koala population is on the northern coast of New South Wales at Coffs Harbour, an area that in 1995–1996 attracted $196 million in direct expenditure from nature-based tourism (Hamilton *et al.* 2000). It is also an area of rapid urban development, which poses an immediate threat to koala population viability (Lunney *et al.* 2000). Enabling the coexistence of koalas with an expanding human population has therefore become a major issue in local conservation planning (Hamilton *et al.* 2000; Lunney *et al.* 2000). With this in mind, we applied our cost function to the task of calculating an optimal α for monitoring to trigger recovery action.

To parameterize the cost function, we used an economic evaluation of koala conservation planning in Coffs Harbour City Council (Hamilton *et al.* 2000). In the most conservative scenario, the minimum estimated loss in tourist revenue (expressed as net present value, NPV) associated with a decline in the local koala population was V = $21 million. The authors did not specify the size of the decline, so we used the moderate effect size of 15% from Fig. 2. The cost of implementing a koala management plan that would prevent these economic losses was estimated to be R = $833,000 (Hamilton *et al.* 2000). Based on prior experience, we estimated the NPV cost of surveying to be M = $56,000. We initially set the probability of a decline occurring (δ) at an intermediate level of 50%.

RESULTS

The resulting cost function shows that the overall cost continually decreases as the probability of spuriously detecting a decline, α, is relaxed towards one (Fig. 3). The curve is particularly steep when α is low, such that even a modest increase in α from 0.05 to 0.20 produces an expected saving in excess of $2 million. However, the key result from this figure is that there is no economic benefit in trying to detect Type I errors and thus determine whether a suspected decline is real or spurious. The economically rational (though not necessarily ecologically sensible) conclusion is that monitoring to trigger recovery action in this situation is superfluous; the manager should assume that a decline is occurring and intervene directly.

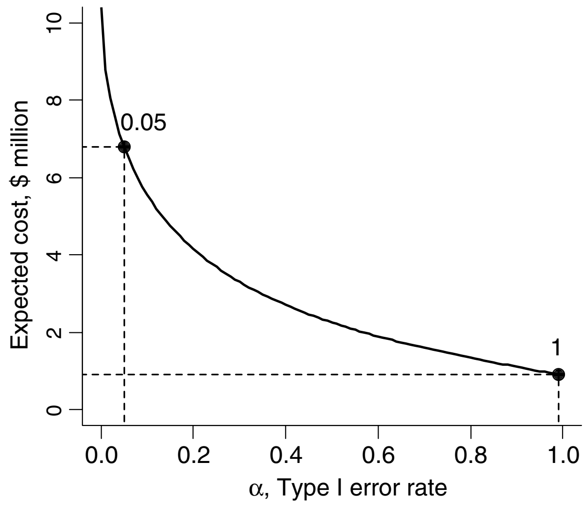

Figure 3 Expected cost ($ million) of monitoring and management of Coffs Harbour koalas as a function of Type I error rate, α. Intersection of dashed horizontal and vertical lines with curves indicate the expected costs at $\alpha = 0.05$ and at the cost function minimum of $\alpha = 1$.

Sensitivity analysis for the koala example showed that this result is very robust. Varying the cost of monitoring simply shifted the curve slightly up or down vertically and had no qualitative impact. Only when the probability of decline, δ, was very small (<12%), or when the size of the decline was very large (>23%), did an optimum α-value less than one appear. In other words, if one was interested only in protecting apparently secure populations and/or detecting catastrophically large declines, employing conventional hypothesis testing to trigger recovery action could be sensible. However, in both these cases, the profile of the cost function was still very flat in the vicinity of the optimum, such that the economic benefits from monitoring over proceeding to immediate recovery action are practically negligible. The simple message is that when the economic costs of Type II errors are as overwhelmingly high as they are for koalas, monitoring using a frequentist hypothesis test is not a cost-efficient way of deciding whether management intervention is needed.

Recognising that the koala represents an extreme case in terms of economic value, we also examined model outcomes for much lower species' values, which are likely to be closer to reality for the majority of threatened species. With economic values (V) slightly above recovery costs (R) and other parameters as for the koala example, a range of optimal α-values were found (Fig. 4). As the ratio $V : R$ was increased from one to five, the optimal α level rose from zero to one, and the total

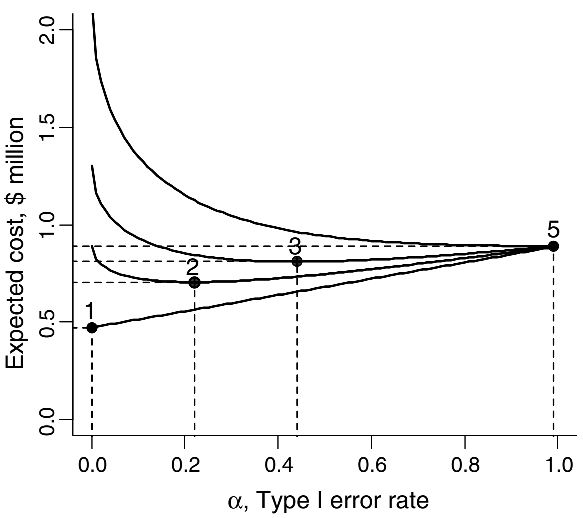

Figure 4 Expected cost ($ million) of monitoring and management as a function of α when Type II errors are between one and five times the cost of Type I errors. Intersection of dashed horizontal and vertical lines with curves indicate the optimal α-value and associated expected costs for that ratio (indicated by numbers 1, 2, 3 and 5) of Type II to Type I error costs.

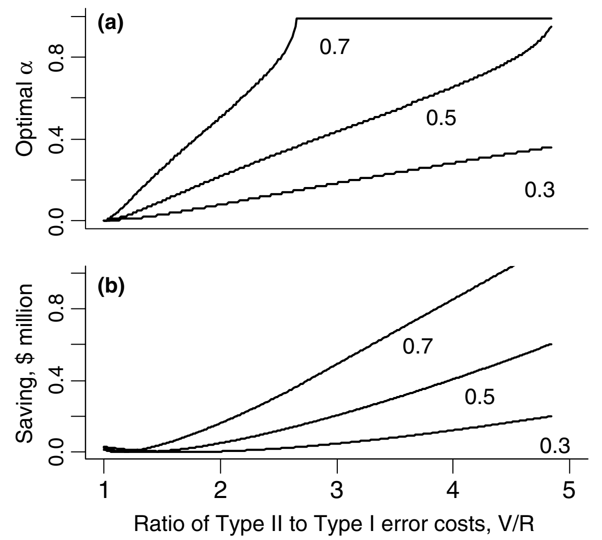

Figure 5 (a) Optimal α-value and (b) expected cost saving ($ million) obtained by using the optimal α rather than $\alpha = 0.05$, as a function of the ratio of Type II to Type I error costs, for three different probabilities of a decline occurring (δ): 0.3, 0.5 and 0.7.

expected cost at the optimum rose from \$0.47 to \$0.89 million (Fig. 4).

Thus there exists a narrow window of $V : R$ ratios – between 1 and 4.85 for the baseline parameter settings – for which an optimal α-value exists. Within this range, the optimal α increases approximately linearly (Fig. 5a) and the cost saving by using the optimal α (rather than $\alpha = 0.05$) increases either side of $V : R = 1.07$ (Fig. 5b). Above $V : R = 4.85$, the cost of Type II errors (failing to detect a decline) is so high that they should never be tolerated (optimal $\alpha = 1$) and management intervention would be prudent regardless of what population trend the monitoring data indicated. When the probability of decline, δ, is increased to 0.7, this point is reached more quickly ($V : R = 2.60$) and the cost saving is greater, whereas when it is decreased to 0.3, the optimal α always remains below 0.4 and the cost saving is relatively small (Fig. 5a,b).

DISCUSSION

These results make an emphatic point about the flawed logic of using low, fixed significance levels in monitoring for environmental management. In the koala example, adhering to the $\alpha = 0.05$ convention rather than acting without monitoring would come at an expected cost in excess of \$5 million. We should make it clear we are not advocating the position that monitoring in such situations is worthless; there are a host of reasons why monitoring data can be useful aside from simply determining whether a binary change in conservation status has occurred. However, to the extent that this objective is driving data collection, and a frequentist hypothesis test is to be the arbiter for management decisions, resources could be better spent elsewhere.

It is also worth noting that in cases where an optimum was present, the cost function profile was quite flat for high α, but rose steeply for low α. This means that there is always a greater penalty for choosing an α that is too low, as opposed to one that is too high. This supports the suggestion by some authors (Gray 1990; Dayton 2001) that a more advisable rule of thumb for hypothesis testing might be to reverse the conventional practice and set a low, fixed target value for β, rather than α. We share Di Stefano's (2003) disdain for arbitrary thresholds of any kind, but in environmental management scenarios for which relative costs of Type I and Type II errors are not known, our analysis still suggests that favouring a low β over a low α would be prudent.

We emphasise the point that the expected cost of management decisions cannot be computed without specifying some prior expectation of how likely it is that the change of interest (in our case a koala population decline) will occur. This practice is unfamiliar in frequentist statistics, but is a core component of Bayesian analysis, in which a prior expectation is combined with the present data to form a 'posterior' distribution for the parameter under scrutiny (Dorazio & Johnson 2003). This point has been recognised by some authors who have dealt with balancing significance and power in frequentist analysis (Cascio & Zedeck 1983; Fox 2001) and adds weight to the argument that Bayesian statistics are a more natural choice for problems in applied management and ecology (Ellison 1996; Wade 2000; Dorazio & Johnson 2003; Wintle *et al.* 2003; Ellison 2004). In our opinion, a Bayesian approach, in which the estimated magnitude of the effect in question can be iteratively updated as more data become available, is a more natural option for a manager who must make a sequence of decisions about how to apportion resources between monitoring and management as the state of the system, and the quality of information about it, changes through time. Bayesian methods are becoming increasingly accessible to ecologists and we actively encourage managers to investigate their application. We will address the task of formulating this problem within a Bayesian framework in future work.

Even so, and despite continuing criticism of frequentist hypothesis testing (Yoccoz 1991; Anderson *et al.* 2000; Eberhardt 2003), it seems likely that many managers tasked with the problem we have addressed here will continue to use it for the foreseeable future. This being the case, we reiterate the need for managers to embrace a decision–theory approach to the analysis of monitoring data, which includes an *a priori* power analysis and takes into account the economic costs associated with inference errors about the environmental trend of interest. Estimates of such costs, or at least their relative magnitude, should be readily obtained in most fields of natural resource management. Setting decision thresholds in this way is not a new concept, but is one that has been almost completely neglected in practice. This omission is not for lack of discussion about the problem in the literature, nor exhortation to remedy it (Bernstein & Zalinski 1983; Millard 1987; Gray 1990; Peterman 1990a,b; Peterman & M'Gonigle 1992; Taylor & Gerrodette 1993; Fairweather 1994; Keough & Mapstone 1997; Dayton 2001; Di Stefano 2001; Fox 2001; Di Stefano 2003). We acknowledge that the process of arriving at thresholds acceptable to all stakeholders may often be difficult, and will require much closer cooperation between professional ecologists and managers than is perhaps typical. However, given the magnitude of what is at stake financially and environmentally, we encourage those in a position to influence environmental decision-making to seriously address the problem.

ACKNOWLEDGEMENTS

We thank Sama Low Choy, Fiona Fidler, Petra Kuhnert, Mick McCarthy and Brendan Wintle for comments on the manuscript. This work was supported by funding from the Australian Research Council, the South Australian Department of Environment and Heritage and the Australian Koala Foundation.

REFERENCES

Anderson, D.R., Burnham, K.P. & Thompson, W.L. (2000). Null hypothesis testing: problems, prevalence, and an alternative. *J. Wildl. Manag.*, 64, 912–923.

Bernstein, B.B. & Zalinski, J. (1983). An optimum sampling design and power tests for environmental biologists. *J. Environ. Manag.*, 16, 35–43.

Cascio, W.F. & Zedeck, S. (1983). Open a new window in rational research planning: adjust alpha to maximize statistical power. *Personnel Psychol.*, 36, 517–526.

Dayton, P.K. (2001). Reversal of the burden of proof in fisheries management. *Science*, 279, 821–822.

Di Stefano, J. (2001). Power analysis and sustainable forest management. *Forest Ecol. Manag.*, 154, 141–153.

Di Stefano, J. (2003). How much power is enough? Against the development of an arbitrary convention for statistical power calculations. *Funct. Ecol.*, 17, 707–709.

Dorazio, R.M. & Johnson, F.A. (2003). Bayesian inference and decision theory – a framework for decision making in natural resource management. *Ecol. Appl.*, 13, 556–563.

Eberhardt, L.L. (2003). What should we do about hypothesis testing? *J. Wildl. Manag.*, 67, 241–247.

Ellison, A.M. (1996). An introduction to Bayesian inference for ecological research and environmental decision-making. *Ecol. Appl.*, 6, 1036–1046.
Ellison, A.M. (2004). Bayesian inference in ecology. *Ecol. Lett.*, 7, 509–520.
Fairweather, P.G. (1994). Improving the use of science in environmental assessment. *Aust. Zool.*, 29, 217–223.
Field, S.A., Tyre, A., Ball, S.J. & Possingham, H.P. (2001). Observer error and statistical power: evaluating survey reliability for conservation management. In: *Modsim 2001: International Congress of Modelling and Simulation* (ed. Littleboy M.). MSSANZ, Canberra, pp. 831–836.
Field, S.A., Tyre, A. & Possingham, H.P. (2004). Optimizing allocation of monitoring effort under economic and observational constraints. *J. Wildl. Manag.*, in press.
Fox, D.R. (2001). Environmental power analysis – a new perspective. *Environmetrics*, 12, 437–449.
Gray, J. (1990). Statistics and the precautionary principle. *Mar. Pollut. Bull.*, 21, 174–176.
Hall, D. (2000). Zero-inflated Poisson and binomial regression with random effects: a case study. *Biometrics*, 56, 1030–1039.
Hamilton, C., Lunney, D. & Matthews, A. (2000). An economic evaluation of local government approaches to koala conservation. *Aust. J. Environ. Manag.*, 7, 158–169.
Jennions, M.D. & Moller, A.P. (2003). A survey of the statistical power of research in behavioral ecology and animal behavior. *Behav. Ecol.*, 14, 438–445.
Keough, M. & Mapstone, B. (1997). Designing environmental monitoring for pulp mills in Australia. *Water Sci. Technol.*, 35, 397–404.
Lunney, D., Matthews, A., Moon, C. & Ferrier, S. (2000). Incorporating habitat mapping into practical koala conservation on private lands. *Conservation Biol.*, 14, 669–680.
MacKenzie, D.I., Nichols, J.D., Lachman, G.B., Droege, S., Royle, J.A. & Langtimm, C.A. (2002). Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, 83, 2248–2255.
Mapstone, B. (1995). Scalable decision rules for environmental impact studies: effect size, type 1, and type 2 errors. *Ecol. Appl.*, 5, 401–410.
Millard, S.P. (1987). Environmental monitoring, statistics, and the law – room for improvement. *Am. Statistician*, 41, 249–253.
Murphy, K.R. & Myors, B. (2004) *Statistical Power Analysis. A Simple and General Model for Traditional and Modern Hypothesis Tests*. 2nd edn. Lawrence Erlbaum Associates, Mahwah, New Jersey.
Nagel, S.S. & Neef, M. (1977). Determining an optimum level of statistical significance. *Eval. Stud. Rev. Annu.*, 2, 146–158.
Neyman, J. & Pearson, E. (1933). The testing of statistical hypotheses in relation to probabilities *a priori*. *Proc. Cambridge Philos. Soc.*, 29, 492–510.
Peterman, R. (1990a). The importance of reporting statistical power: the forest decline and acidic deposition example. *Ecology*, 71, 2024–2027.
Peterman, R. (1990b). Statistical power analysis can improve fisheries research and management. *Can. J. Fish. Aqua. Sci.*, 47, 2–15.
Peterman, R. & M'Gonigle, M. (1992). Statistical power analysis and the precautionary principle. *Mar. Pollut. Bull.*, 24, 231–234.
Power, M., Power, G. & Dixon, D. (1995). Detection and decision-making in environmental effects monitoring. *Environ. Manag.*, 19, 629–639.
Sedlmeier, P. & Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies. *Psychol. Bull.*, 105, 309–316.
Shea, K., Possingham, H.P., Murdoch, W.W. & Roush, R. (2002). Active adaptive management in insect pest and weed control: Intervention with a plan for learning. *Ecol. Appl.*, 12, 927–936.
Taylor, B. & Gerrodette, T. (1993). The uses of statistical power in conservation biology: the vaquita and northern spotted owl. *Conservation Biol.*, 7, 489–500.
Tyre, A.J., Possingham, H.P. & Lindenmayer, D.B. (2001). Inferring process from pattern: can territory occupancy provide information about life history parameters? *Ecol. Appl.*, 11, 1722–1737.
Tyre, A.J., Tenhumberg, B., Field, S.A., Niejalke, D., Parris, K. & Possingham, H.P. (2003). Improving precision and reducing bias in biological surveys: Estimating false-negative error rates. *Ecol. Appl.*, 13, 1790–1801.
Wade, P.R. (2000). Bayesian methods in conservation biology. *Conservation Biol.*, 14, 1308–1316.
Wald, A. (1950) *Statistical Decision Functions*. Chelsea, New York.
Wintle, B.A., McCarthy, M.A., Volinsky, C.T. & Kavanagh, R.P. (2003). The use of Bayesian model averaging to better represent uncertainty in ecological models. *Conservation Biol.*, 17, 1579–1590.
Yoccoz, N. (1991). Use, overuse and misuse of significance tests in evolutionary biology and ecology. *Bull. Ecol. Soc. Am.*, 72, 106–111.

Editor, Fangliang He
Manuscript received 24 March 2004
First decision made 3 May 2004
Manuscript accepted 18 May 2004

APPENDIX A

Lack of a decision-theoretic approach to setting α and β

Consideration of how to set optimal signficance levels can be traced back to Egon Pearson and Jerzy Neyman (Neyman & Pearson 1933) and Abraham Wald (Wald 1950). Numerous authors have since proposed methods for doing so in various decision-making contexts (Nagel & Neef 1977; Cascio & Zedeck 1983; Mapstone 1995; Power *et al.* 1995; Fox 2001; Murphy & Myors 2004). However, each of these lacks one or more of the components of our formulation. For example, the proposal whose context was most similar to our decision problem (Mapstone 1995) considered how to set α and β when using an indicator species to detect environmental impacts from effluent discharges. The author suggested setting α and β such that their ratio equalled the ratio of Type I to Type II error costs. Although a vast improvement over using the arbitrary rule of $\alpha = 0.05$, we would argue that this approach is deficient in two respects: (1) the objective should be not to equalize the costs of the two kinds of errors, but to minimize their sum; and (2) that some prior expectation of the probability of the impact of interest occurring must be taken into account. Furthermore, a literature survey indicated that none of the existing proposals for setting optimal significance levels have been widely embraced by ecologists in practice. We surveyed the last 5 years (1998–2003) of monitoring literature published in the journals Austral Ecology, Conservation Biology, Ecological Applications and Journal of Applied Ecology. These journals were chosen to cover a wide geographical range of monitoring studies, with a focus on applied population management. We searched the journal abstracts for the word 'trend' and at least one of the following words: 'monitor', 'estimate', 'decline', 'decrease' and 'increase' (including all grammatical forms).

The search gave a total of 147 hits and after excluding papers not testing for temporal trends or being out of scope for other reasons (e.g. not dealing with temporal patterns at all, summary and review papers, etc.), we found 46 relevant papers. Of these, only eight estimated statistical power and not a single paper used decision-theory to set the α and β levels.

APPENDIX B

Relationship between α and β

To estimate β for a given α, we simulated a 'virtual ecologist' (VE) (Tyre *et al.* 2001) conducting presence–absence sampling on a declining population and performing a hypothesis test on the resulting data (Field *et al.* 2001, 2004). The population began with an initial occupancy p and then declined linearly by a total amount d over the study period. The VE made m repeat visits at each of three time periods (beginning, middle and end of the decline) to n sites and had a probability q (detectability) of sighting the species on each visit, given that it was present. The VE then arrived at a conclusion about whether a decline had occurred by fitting a zero-inflated binomial (ZIB) model (Hall 2000; MacKenzie *et al.* 2002; Tyre *et al.* 2003) to the data and testing for significance at a given α. Statistical power ($1-\beta$) could then be found by repeating this process a large number of times at each α and calculating the proportion of times that a decline was detected.

To avoid time-consuming simulations, we approximated the power of the ZIB model by fitting a generalized linear model (GLM) to the likelihood ratio between the decline model and the null model. The likelihood ratios follow a non-central chi-square distribution with one degree of freedom where the non-centrality parameter depends on p, q, d, n and m. The mean of a non-central chi-square with 1 degree of freedom is $1 + \lambda$, and the variance is $2 + 4\lambda$ where λ is the non-centrality parameter. Therefore, we fit a quasi-likelihood model with a log link and variance function proportional to the mean. After some trial and error we used the following model structure

$$\begin{aligned} f = {} & b_0 + b_1 n + b_2 d + b_3 m + b_4 q + b_5 p_0^2 \\ & + b_6 p_0^2 + b_7 nd + b_8 mq \end{aligned} \tag{A1}$$

where f is the linear predictor in the GLM. The quadratic term for p_0 was included after preliminary fits indicated that power was highest at $p_0 \approx 0.5$ and declined for both higher and lower values. We used the statistical software *R* version 1.6.2 to fit the model.

The data to fit the model were created using a latin hypercube with 1100 uniformly distributed values between the indicated bounds: $n(10, 300)$, m $(1.4, 8)$, $p_0(0.2, 0.95)$, $q(0.2, 0.95)$, and $d(-0.5, -0.1)$. We discarded runs where numerical error resulted in negative likelihood ratios (all with absolute values <0.001) or where the minimiser failed to converge (usually when simulated sample sizes were small). This resulted in a total of 1040 points to fit the approximating model.