*Research Note*

# Impact of Criticism of Null-Hypothesis Significance Testing on Statistical Reporting Practices in Conservation Biology

FIONA FIDLER,*§ MARK A. BURGMAN,* GEOFF CUMMING,† ROBERT BUTTROSE,* AND NEIL THOMASON‡

*School of Botany, University of Melbourne, Victoria 3010, Australia
†School of Psychological Science, La Trobe University, Victoria 3086, Australia
‡Department of History and Philosophy of Science, University of Melbourne, Victoria 3010, Australia

**Abstract:** *Over the last decade, criticisms of null-hypothesis significance testing have grown dramatically, and several alternative practices, such as confidence intervals, information theoretic, and Bayesian methods, have been advocated. Have these calls for change had an impact on the statistical reporting practices in conservation biology? In 2000 and 2001, 92% of sampled articles in* Conservation Biology *and* Biological Conservation *reported results of null-hypothesis tests. In 2005 this figure dropped to 78%. There were corresponding increases in the use of confidence intervals, information theoretic, and Bayesian techniques. Of those articles reporting null-hypothesis testing—which still easily constitute the majority—very few report statistical power (8%) and many misinterpret statistical nonsignificance as evidence for no effect (63%). Overall, results of our survey show some improvements in statistical practice, but further efforts are clearly required to move the discipline toward improved practices.*

**Keywords:** Bayesian methods, confidence intervals, statistical power, statistical significance testing

Impacto de la Crítica a las Pruebas de Significancia de Hipótesis Nula en las Prácticas de Reporte Estadístico en Biología de la Conservación

**Resumen:** *En la última década, las críticas a la prueba de significancia estadística han crecido dramáticamente, y se han propuesto varias prácticas alternativas, como los intervalos de confianza, teoría de la información y métodos Bayesianos. ¿Han tenido impacto estos llamados al cambio sobre las prácticas de reporte estadístico en biología de la conservación? En 200 y 2001, 92% de los artículos muestreados en* Conservation Biology *y* Biological Conservation *reportaron pruebas de hipótesis nulas. En 2005 esta cifra cayó a 78%. Hubo incrementos correspondientes en el uso de intervalos de confianza, teoría de la información y técnicas Bayesianas. De los artículos que reportan pruebas de hipótesis nulas—que fácilmente aun constituyen la mayoría—muy pocos (8%) reportan el poder estadístico y muchos mal interpretan la no significancia estadística como evidencia de no efecto (63%). En general, los resultados de nuestro muestreo muestran algunas mejorías en la práctica estadística, pero claramente se requieren mayores esfuerzos para que la disciplina tenga mejores prácticas.*

**Palabras Clave:** intervalos de confianza, métodos Bayesianos, poder estadístico, prueba de significancia estadística

## Introduction

Over the last decade there has been a dramatic increase in criticisms of null-hypothesis significance testing (NHST) (or statistical-significance testing) and calls for reform of traditional statistical practice in conservation biology and ecology (e.g., Johnson 1999, 2002; Anderson et al. 2000; Ellison 2004). The same trend has occurred in psychology, economics, and education (Fidler et al. 2004*a*). Despite widespread agreement that current practices are flawed, some disciplines have shown remarkable resistance to change (e.g., psychology and economics; Altman 2004), as measured by reporting practices in journals. After almost half a century of criticisms, and in the case of psychology, serious editorial and institutional intervention from professional bodies such as the American Psychological Association, statistical-reporting practices in journals remain largely unmoved and heavily dependent on NHST-based procedures (Finch et al. 2004; Fidler et al. 2005).

Has the discipline of conservation biology been equally resistant to change? We examined the criticisms of NHST relevant to conservation biology and some of the alternative practices promulgated in the literature and explored changes in the reporting practices of conservation biologists by examining practices documented in two leading journals. Based on the results of our surveys, we outline some strategies that may accelerate the pace of change, thereby improving the merits of inferences and the effectiveness of science-based decision making in conservation biology.

## Criticisms of and Alternatives to NHST

The literature criticizing NHST is extensive and spans at least half a century. Here we cannot provide a catalog of criticisms; several such reviews have already been published (e.g., Nickerson 2000; Kline 2004). Conservation biology has a particular interest in getting things right: "The consequences of accepting a false null hypothesis can be acute in conservation biology because endangered populations leave very little margin for recovery from incorrect management decisions" (Taylor & Gerrodite 1993). For small populations, waiting for a statistically significant decline before instituting strong protection measures is often tantamount to a guarantee of extinction. Incomplete statistical reporting, particularly low and unknown statistical power, can result in direct, unanticipated, unacceptable environmental damage. Specific consequences of misinterpretation of significance testing are not documented in the articles we sampled, but they include failing to act when action was warranted, unnecessary expenditure when action was not warranted, and provision of incorrect advice that affected policy and planning decisions.

The most outspoken critics of NHST have included advocates of Bayesian methods (e.g., Spiegelhalter et al. 1994; Ellison 1996; Wade 2000) and proponents of likelihood and information theoretic methods (e.g., Anderson et al. 2000; Burnham & Anderson 2002). However, as many, if not more, criticisms and calls for change have come from within the mainstream, error probability, frequentist framework as from outside it (e.g., Cohen 1994; McCloskey 1995). Our arguments against over-reliance on NHST and for alternative practices are conservative in the sense that they reflect minimal changes in practice necessary to overcome major flaws.

In ecology it has been routine to report *p* values since the 1960s (Fidler et al. 2004*a*). Anderson et al. (2000) report that the number of *p* values published in *Ecology* and *Journal of Wildlife Management* has exceeded 3000 every year since 1984 and 1994, respectively. Criticisms of NHST from within a frequentist framework are not necessarily limited to those of misuse or misinterpretation of the test (e.g., neglect of statistical power or misinterpretation of a *p* value). They also legitimately address the irrelevance of dichotomous decision making, neglect of estimation, and the limited emphasis on prior information and cumulative knowledge. Many of these concerns are shared by Bayesian adherents, but they are not intrinsically Bayesian. Bayesian and information theoretic methods have a lot to offer to ecology and conservation biology, and we support their use. Nevertheless, there are other, simple steps that can be taken immediately to overcome problems with NHST without abandoning a classical framework.

Increased use of statistical power has been promoted as a solution to current statistical practice in ecology. Unfortunately, a number of power advocates have failed to distinguish between power calculated either a priori or retrospectively based on the expected effect size, and power calculated retrospectively based on the obtained effect size. Some have even recommended the use of post hoc or retrospective power analysis, based on the observed effect size (e.g., Reed & Blaunstein 1995; Thomas & Juanes, 1996). This practice has been severely and justifiably criticized (e.g., Hoenig & Heisey 2001). Null-hypothesis significance testing with statistical power calculations (based on the expected effect size) is preferable to NHST without power calculations. Even appropriate use of power, however, implies a dichotomous decision and diverts attention from estimation, uncertainty, and the accumulation of scientific knowledge. Equivalence testing has also been recommended as means of reversing the traditional burden of proof in hypothesis testing (Hoenig & Heisey 2001).

Others argue for more general use of confidence intervals, particularly in graphical form (Cumming & Finch 2001, 2005; DiStefano 2003). Confidence intervals make uncertainty explicit and may offer cognitive advantages. They assist meta-analytic thinking—that is, thinking

across the results of independent studies—rather than making dichotomous reject or do-not-reject decisions based on the outcome of single experiments. This relates to acknowledging prior information, with an emphasis on effect size. Results presented as merely statistically significant or not can create the illusion of inconsistency in the literature, particularly when review studies simply tally significant results against nonsignificant results (Parris & McCarthy 2001; McCarthy & Parris 2004). Confidence intervals bind precision information with statistical significance information; the width of a confidence interval is a measure of a study's imprecision and whether or not the interval captures the null-hypothesis value is a basis for statistical significance decisions. Furthermore, confidence intervals provide salient information on the estimated size of the effect. Significance testing, in contrast, provides no direct information on effect size, and full understanding relies on the reporting of extra information, which authors often fail to provide.

## Methods

We surveyed the statistics reported in *Conservation Biology* and *Biological Conservation*. We recorded instances of statistical items (Table 1) reported in 50 articles in each

journal published in 2001–2002 and 50 articles in each journal published in 2005. Our sample included only articles with empirical data and did not include meta-analyses or methodological or theoretical articles in our analysis. We calculated the proportion of articles reporting each statistical item and 95% confidence intervals for those proportions with the method recommended by Newcombe and Altman (2000). We independently double-coded 10% of 2001 and 2002 articles, selected to represent the full range of article types. The accuracy of the first author's coding (assessed by the second and third authors) was 92%. Discrepancies were oversights, rather than disagreements over definitions. By way of comparison of current practices, we also coded 50 articles in *Ecology* and 50 articles in *Journal of Ecology* published in 2005.

## Results

In 2001 and 2002 92% of sampled articles in *Conservation Biology* and *Biological Conservation* reported at least one $p$ value (Table 1). In 2005 this figure dropped to 78%. In 2001 and 2002 7 of the 8 articles with no NHST reported only descriptive statistics. In 2005 of the 22 articles that did not use NHST 7 were descriptive, 9 built mathematical models, 4 used Akaike's information criteria

**Table 1. Percentage of articles (and 95% CIs) in *Conservation Biology* and *Biological Conservation* with statistical significance tests, confidence intervals, and figures.**

| | Good practice[a]? | Conservation Biology *and* Biological Conservation *(2001 and 2002)* | | Conservation Biology *and* Biological Conservation *(2005)* | |
|---|---|---|---|---|---|
| | | *articles,% (n)* | *95% CI (%)* | *articles,% (n)* | *95% CI (%)* |
| Any statistical significance test | | 92 (92/100) | 85–96 | 78 (78/100) | 69–85 |
|   nil null hypothesis[b] | × | 79 (73/92) | 70–86 | 97 (76/78) | 91–99 |
|   ambiguous use of *significant*[c] | × | 68 (63/92) | 58–77 | 63 (49/78) | 52–73 |
|   exact $p$ value[d] | √ | 62 (56/92) | 51–70 | 69 (54/78) | 58–78 |
|   $p$ value asterisks (i.e., *, **, ***)[e] | × | 25 (23/92) | 17–35 | 22 (17/78) | 14–32 |
|   nonsignificant result | | 80 (74/92) | 71–87 | 86 (67/78) | 77–92 |
|     statistical power | √√ | 3 (2/74) | 0–9 | 8 (5/67) | 3–16 |
|     indirect reference to power[f] | √ | 30 (22/74) | 21–41 | 30 (20/67) | 20–42 |
|     interpret as "no effect" | × | 47 (35/74) | 35–57 | 63 (42/67) | 51–73 |
| Any confidence interval | √√ | 19 (19/100) | 13–28 | 26 (26/100) | 18–35 |
|   interpreted confidence interval[g] | √√√ | 26 (5/19) | 12–49 | 31 (8/26) | 17–50 |
| Any figure with data | √√ | 77 (77/100) | 68–84 | 69 (69/100) | 59–77 |
|   error bars on figure[h] | √√√ | 40 (31/77) | 30–51 | 51 (35–69) | 39–62 |

*Key: [a]x, not recommended; √, recommended; √√, highly recommended; √√√, very highly recommended.*

*[b]Hypothesis of, for example, no effect, no difference, or zero correlation.*

*[c]Cases in which we could not determine whether the author was speaking substantively or statistically. If the author did not preface* significant *with* statistically, *follow it with a* p *value or test statistic, or otherwise differentiate statistical and substantive interpretations, the practice was recorded as ambiguous.*

*[d]For example,* p = 0.003.

*[e]This practice has been heavily criticized (e.g., Meehl 1978) because it provides even less information than exact* p *values, is usually insufficient for meta-analysis, and has the potential to mislead researchers into thinking that an effect with two stars is more important than an effect with one.*

*[f] For example, noting that the sample size was small.*

*[g]Any mention of CI width, the possible theoretical importance of the upper or lower bound of the interval, overlap between two CIs, or the word* precision *used in relation to a CI.*

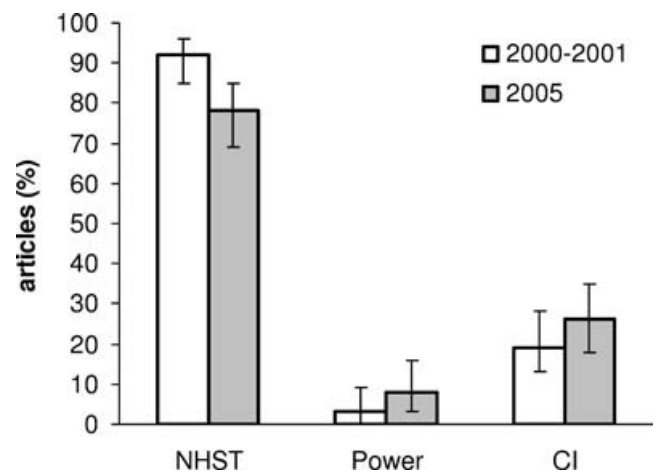*[h]Error bars included standard error and confidence interval bars.*
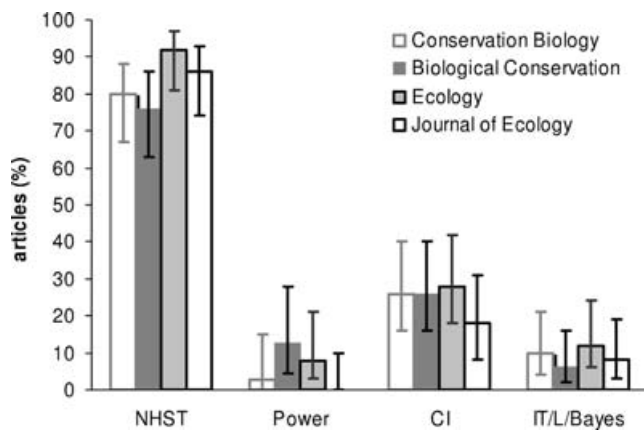
*Figure 1. Percentage of articles reporting null-hypothesis testing (NHST), statistical power (power), confidence intervals (CI), and information theoretic, likelihood, or Bayesian methods (IT/L/Bayes) in 2005 in* Conservation Biology, Biological Conservation, Ecology *and* Journal of Ecology. *Error bars are 95% CIs.*



*Figure 2. Percentage of* Conservation Biology *and* Biological Conservation *in 2000–2001 versus the percentage of articles in 2005 that reported null-hypothesis testing (NHST), statistical power (power), and confidence intervals (CI). Error bars are 95% CIs.*

(AIC) model selection techniques, 1 was Bayesian, and 1 reported confidence intervals as the primary analysis. In addition, of those articles that did report NHST results, 4 supplemented this analysis with AIC, 2 supplemented with maximum likelihood estimates, and 2 supplemented with Bayesian analysis. In *Ecology* and *Journal of Ecology* the rate of NHST in 2005 remained higher than in the conservation biology journals (Fig. 1).

In 2001 and 2002 statistically nonsignificant results were often reported without statistical power: 80% of articles that used NHST reported a nonsignificant result, but only 3% of these reported power. In 2005 there was a slight increase in the reporting of statistical power: 86% of articles that used NHST reported a statistically nonsignificant result, and 8% of those reported statistical power (Fig. 2).

There was an increase in the reporting rate of confidence intervals: up from 19% in 2001–2002 to 26% in 2005. There was also an increase (from 40% to 51%) in the percentage of figures with error bars (primarily SE or CI; Table 1).

In 2001 and 2002 statistically nonsignificant results were interpreted as evidence of "no effect" or "no relationship" in 47% of articles that included a nonsignificant result—despite the rarity of statistical power information. In 2005 this misconception was present in 63% of such articles. This increase cannot be accounted for by the small increase in statistical power reporting. In addition, there remain many cases of *p* values being reported without effect size estimates, measures of variance, or sample size information (Table 2). Similar reporting practices and problems exist in recent articles in *Ecology* and *Journal of Ecology* (Table 3).

## Discussion

Based on our results we conclude that the statistical-reporting practices used by authors publishing in two leading conservation biology journals are changing. However the changes are of a limited extent, and NHST continues to dominate (Fig. 2). Furthermore, when NHST was used, it was rarely supplemented with statistical power and often important additional information about effect size and variation was missing. Clearly, further efforts are needed to improve practices.

The medical field recognized deficiencies in NHST and, in response, institutionalized confidence-interval reporting (Altman et al. 2000; Fidler et al. 2004*b*). By the mid 1980s, most major medical journals (including the *New England Journal of Medicine*, *British Medical Journal*, and *Journal of the American Medical Association*) had policies warning against over-reliance on *p* values (ICMJE 1988). Surveys of medical journals before and after statements of editorial policy suggest these were largely effective at changing statistical-reporting practices (e.g., Seldrup 1997; Fidler et al. 2004*b*). In psychology, *p* values still dominate, despite decades of advocacy of alternative methods. Recently the American Psychological Association (APA) Publication Manual, which sets editorial standards for over 1000 journals across many disciplines, acknowledged the null-hypothesis-test debate and recommended confidence intervals as "in general, the best reporting strategy" (APA 2001). It remains to be seen what effect this encouragement will have. Currently, 23 journals in psychology have policies warning of pitfalls of NHST and/or recommending alternatives, usually effect sizes, confidence intervals, and clinical significance (Hill & Thompson 2004). In economics the theme is repeated.

**Table 2.  Percentage of articles (and 95% CIs) in *Conservation Biology* and *Biological Conservation* with statistical significance tests that also reported, or omitted, an effect size measure,[a] variance measure (SD or SE), or sample size (*n* or df).**

| | Good practice?[b] | Conservation Biology *and* Biological Conservation *(2000 and 2001)* | | Conservation Biology *and* Biological Conservation *(2005)* | |
|---|---|---|---|---|---|
| | | % *(of 92)* | *95% CI (%)* | % *(of 78)* | *95% CI (%)* |
| At least one effect size | √ | 87 | 79–92 | 89 | 80–94 |
| Missing at least one effect size | × | 43 | 34–54 | 58 | 47–68 |
| At least one SD or SE | √ | 48 | 38–58 | 47 | 37–58 |
| Missing at least one SD or SE | × | 67 | 57–76 | 85 | 75–91 |
| At least one *n* or df | √ | 76 | 66–84 | 77 | 66–85 |
| Missing at least one *n* or df | × | 36 | 27–46 | 51 | 40–62 |

[a] *We classified the following measures as an effect size: mean (or percent or proportion) difference; any relevant standardized measure (e.g., Cohen's* d*); b or use of beta($\beta$); variance-accounted-for measures such as* $r^2$ *(for regression) or* $\eta^2$ *(for ANOVA); correlation coefficients and other unit-free measures, such as odds ratios.*
[b] *Key: x, not recommended; √, recommended; √√, highly recommended; √√√, very highly recommended.*

NHST still dominates but criticisms are increasing (Altman 2004; Ziliak & McCloskey 2004).

Bayesian and information theoretic methods offer ways out of the current undesirable situation in statistical reporting in ecology and conservation biology. As their computational difficulties become less daunting, with more capable computers and better-developed software, these methods will increase in popularity. Nevertheless, they are yet to be incorporated in most undergraduate curricula and mainstream training of ecologists, so it is unreasonable to expect a rapid uptake. Because current NHST practices have the potential to cause serious damage to the progress of sciences and to the subjects of its study (e.g., people, species, and the environment), immediate action is desirable. A switch from null-hypothesis testing and *p* values to effect sizes and estimates of uncertainty (confidence intervals) will go a long way toward preventing such damage. Confidence intervals have the advantage of familiarity and wide acceptance and could immediately be implemented by many researchers as a step toward

remediating current problems (Altman et al. 2000; Cumming & Finch 2001, 2005; Smithson 2002; Kline 2004).

Any recommendations for improved statistical-reporting practice are unlikely to be effective without editorial support. There is ample evidence from other disciplines to support this claim (Sedlmeier & Gigerenzer 1989; Fidler et al. 2004*a*). In ecology there have been a few attempts to improve practice through editorial policy. The *Journal of Wildlife Management* (1995; Otis 1995) encouraged the use of statistical power and confidence intervals. The Ecological Society of America (ESA) publishes six journals including *Ecology*, *Ecological Applications*, *and Ecological Monographs*. Their "Guidelines for Statistical Analysis and Data Presentation" (http://www.esapubs.org/esapubs/Statistics.htm) also go some way toward encouraging reform. For example, they point out that "effect size and biological importance must not be confused with statistical significance." They also recommend that reported information include a measure of precision, such as a standard error or confidence

**Table 3.  Percentage of articles (and 95% CIs) in *Ecology* and *Journal of Ecology* with statistical significance tests, confidence intervals, and figures.**

| | Good practice?* | Ecology *(2005)* | | Journal of Ecology *(2005)* | |
|---|---|---|---|---|---|
| | | *articles,% (*n*)* | *95% CI (%)* | *articles,% (*n*)* | *95% CI (%)* |
| Any statistical significance test | | 92 (46/50) | 81–97 | 86 (43/50) | 74–93 |
|   nil null hypothesis | × | 98 (45/46) | 89–100 | 100 (43/43) | 92–100 |
|   ambiguous use of *significant* | × | 76 (35/46) | 62–86 | 81 (35/43) | 67–90 |
|   exact *p* value | √ | 72 (33/46) | 58–83 | 72 (31/43) | 57–83 |
|   *p* value asterisks (i.e., *, **, ***) | × | 43 (20/46) | 30–58 | 44 (19/43) | 30–59 |
|   nonsignificant result | | 83 (38/46) | 69–91 | 79 (34/43) | 65–89 |
|     statistical power | √√ | 5 (2/38) | 2–17 | 0 (0/34) | 0–10 |
|     indirect reference to power | √ | 8 (3/38) | 3–21 | 21 (9/34) | 15–43 |
|     interpret as "no effect" | × | 74 (28/38) | 58–85 | 79 (34/43) | 65–89 |
| Any confidence interval | √√ | 28 (14/50) | 18–42 | 18 (9/50) | 10–31 |
|   interpret confidence interval | √√√ | 50 (7/14) | 27–73 | 22 (4/18) | 1–45 |
| Any figure with data | √√ | 90 (45/50) | 79–96 | 96 (48/50) | 87–99 |
|   error bars on figure | √√√ | 67 (30/45) | 52–79 | 73 (35/48) | 59–83 |

*Key: x, not recommended; √, recommended; √√, highly recommended; √√√, very highly recommended.*

interval, and they encourage graphical presentation of results. Nevertheless, they confusingly refer to confidence intervals as "descriptive statistics" when in fact they are inferential in nature and describe power as only "occasionally" useful. Perhaps more problematically, there are no examples of how to report alternatives to the very familiar null-hypothesis tests.

We recommend that journal editors consider a collaborative double-pronged approach to improving statistical practices. First, require authors right now to avoid null-hypothesis significance testing, especially dichotomous decision making, wherever possible, and instead use estimation (i.e., effect sizes and standard errors and/or confidence intervals) wherever appropriate; and, second, encourage authors to continue to explore a range of better alternatives to statistical significance testing.

## Acknowledgment

## Literature Cited

Altman, M. 2004. Introduction. Special issue on statistical significance. Journal of Socio Economics **33:**615–630.

Altman, D. G., D. Machin, T. N. Bryant, and M. J. Gardner. 2000. Statistics with confidence: confidence intervals and statistical guidelines. 2nd edition. British Medical Journal Books, London.

APA (American Psychological Association). 2001. Publication Manual of the American Psychological Association. 5th edition. American Psychological Association, Washington, D.C.

Anderson, D. R., K. P. Burnham, and W. L. Thompson. 2000. Null hypothesis testing: problems, prevalence and an alternative. Journal of Wildlife Management **64:**912–923.

Burnham, K. P., and D. R. Anderson. 2002. Model selection and multimodel inference: a practical information-theoretic approach. Springer-Verlag, New York.

Cohen, J. 1994. The earth is round ($p < .05$). American Psychologist **49:**997–1003.

Cumming, G., and S. Finch. 2001. A primer on the understanding, use, and calculation of confidence intervals that are based on central and non-central distributions. Educational and Psychological Measurement **61:**532–574.

Cumming, G., and S. Finch. 2005. Inference by eye: confidence intervals, and how to read pictures of data. American Psychologist **60:**170–180.

DiStefano, J. 2003. A confidence interval approach to data analysis. Forest Ecology and Management **187:**173–183.

Ellison, A. M. 2004. Bayesian inference in ecology. Ecology Letters **7:**509–520.

Fidler, F., G. Cumming, M. Burgman, and N. Thomason. 2004*a*. Statistical reform in medicine, psychology and ecology. Journal of Socio Economics **33:**615–630.

Fidler, F., N. Thomason, G. Cumming, S. Finch, and J. Leeman. 2004*b*. Editors can lead researchers to confidence intervals but they can't make them think: statistical reform lessons from medicine. Psychological Science **15:**119–126.

Fidler, F., G. Cumming, N. Thomason, D. Pannuzzo, J. Smith, P. Fyffe, H. Edmonds, C. Harrington, and R. Schmitt. 2005. Evaluating the effectiveness of editorial policy to improve statistical practice: the case of the Journal of Consulting and Clinical Psychology. Journal of Consulting and Clinical Psychology **73:**136–143.

Finch, S., G. Cumming, J. Williams, L. Palmer, E. Griffith, C. Alders, J. Anderson, and O. Goodman. 2004. Reform of statistical inference in psychology: the case of memory and cognition. Behavior Research Methods, Instruments, and Computers. **36:**312–324.

Hill, C. R., and B. Thompson. 2004. Computing and interpreting effect sizes. Pages 175–196 in J. C. Smart, editor. Higher education: handbook of theory and research. Volume 19. Kluwer, New York.

Hoenig, J. M., and D. M. Heisey. 2001. The abuse of power: the pervasive fallacy of power calculations for data analysis. American Statistician **55:**19–24.

ICMJE (International Committee of Medical Journal Editors). 1988. Uniform requirements for manuscripts submitted to biomedical journals. Annals of Internal Medicine **108:**258–265.

Johnson, D. H. 1999. The insignificance of statistical significance testing. Journal of Wildlife Management **63:**763–772.

Johnson, D. H. 2002. The role of hypothesis testing in wildlife science. Journal of Wildlife Management **66:**272–276.

Kline, R. B. 2004. Beyond significance testing: reforming data analysis methods in behavioral research. American Psychological Association, Washington, D.C.

McCarthy, M. A., and K. M. Parris. 2004. Clarifying the effect of toe clipping on frogs with Bayesian statistics. Journal of Applied Ecology **41:**780–786.

McCloskey, D. N. 1995. The insignificance of statistical significance. Scientific American **272:**104–105.

Meehl, P. E. 1978. Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. Journal of Consulting and Clinical Psychology **4:**806–843.

Newcombe, R. G., and D. G. Altman. 2000. Proportions and their differences. Pages 39–50 in D.G. Altman, D. Machin, T.N. Bryant, and M.J. Gardner, editors. Statistics with confidence: confidence intervals and statistical guidelines. 2$^{nd}$ edition. British Medical Journal Books, London.

Nickerson, R. S. 2000. Null hypothesis significance testing: a review of an old and continuing controversy. Psychological Methods **5:**241–301.

Otis, D. 1995. Journal news. Journal of Wildlife Management **59:**630.

Parris, K. M., and M. A. McCarthy. 2001. Identifying effects of toe clipping on anuran return rates: the importance of statistical power. Amphibia Repilia **22:**275–289.

Reed, J. M., and A. R. Blaunstein. 1995. Assessment of "nondeclining" amphibian populations using power analysis. Conservation Biology **9:**1299–1300.

Sedlmeier P., and G. Gigerenzer. 1989. Do studies of statistical power have an effect on the power of studies? Psychological Bulletin **105:**309–315.

Seldrup, J. 1997. Whatever happened to the *t* test? Drug Information Journal **31:**745–750.

Smithson, M. 2002. Statistics with confidence. Sage, Thousand Oaks, California.

Spiegelhalter, D. J., L. S. Freedman, and M. K. B. Parmar. 1994. Bayesian approaches to randomised trials. Journal of the Royal Statistical Society Association **157:**357.

Taylor, B. L., and T. Gerrodette. 1993. The use of statistical power in conservation biology: the vaquita and Northern Spotted Owl. Conservation Biology **7:**489–500.

Thomas, L., and F. Juanes. 1996. The importance of statistical power analysis: an example from Animal Behaviour. Animal Behaviour **52:**856–859.

Wade, P. R. 2000. Bayesian methods in conservation biology. Conservation Biology **14:**1308–1316.

Ziliak, S., and D. McCloskey. 2004. Size matters: the standard error of regressions in the American Economic Review. Journal of Socio Economics **33:**527–546.