Merging Precaution with Sound Science under the Endangered Species Act

Merging Precaution with Sound Science under the Endangered Species Act

DANIEL J. McGARVEY

Hypothesis tests, which aim to minimize type I errors (false positive results), are standard procedures in scientific research, but they are often inappropriate in Endangered Species Act (ESA) reviews, where the primary objective is to prevent type II errors (false negative results). Recognizing this disparity is particularly important when the best data available are sparse and therefore lack statistical power, because hypothesis tests that use data sets with low statistical power are likely to commit type II errors, thereby denying necessary protection to threatened and endangered species. Equivalence tests can alleviate this problem, and ensure that imperiled species receive the benefit of the doubt, by switching the null and alternative hypotheses. These points are illustrated by critiquing a recent review of ESA requirements for endangered fishes in Upper Klamath Lake (southern Oregon).

Keywords: hypothesis test, equivalence test, statistical power, burden of proof, type I and type II error

Hypothesis tests are integral components of conventional, peer-reviewed research, but they are frequently incompatible with the Endangered Species Act (ESA) for two reasons. First, hypothesis tests assume that type II error (failing to detect a significant effect) is preferable to type I (errantly claiming a significant effect). This assumption is prudent in laboratory settings, where the scientific community can duplicate experiments many times, and research outcomes do not involve distinct "winners" and "losers." Type I errors are likely to lead future research astray, whereas type II errors may entail little more than delays (Kuhn 1970, Shrader-Frechette and McCoy 1992). When dealing with threatened and endangered species, however, scientists can no longer assume that type II error, which often results in failure to provide necessary protection, and is therefore prone to facilitate extinction, is preferable to type I (figure 1); unlike other forms of environmental damage, which can sometimes be remedied after the fact, extinction constitutes an irreversible harm (NRC 1995, Ludwig et al. 2001, Kinzig et al. 2003).

Second, the data on hand in ESA reviews are often inadequate to perform rigorous hypothesis tests (see "Statistical power" below), and the ESA does not include an affirmative requirement to collect additional data (NRC 1995, Brennan et al. 2003, Doremus 2004). Rather, it specifies that all reviews must comply with predetermined schedules (e.g., listing reviews must be completed within 12 months of their initiation), using only the "best...data available" (16 U.S.C. 1533[b][1–3]; 16 U.S.C. 1536[a][2], [b][1]). These schedules ensure timely ESA reviews (and prevent scientists from "studying species to death"), but they also tend to violate the hypothesis test assumption that sufficient data can be obtained to estimate experimental parameters with a high level of confidence (Toft and Shea 1983, Peterman 1990).

These two caveats would effectively place the burden of proof on imperiled species and their advocates, without ensuring that those parties had a realistic opportunity to make their case, any time hypothesis tests were required to initiate protective measures (NRC 1995, Ludwig et al. 2001). Fortunately, the US Fish and Wildlife Service (USFWS) and the National Oceanic and Atmospheric Administration's National Marine Fisheries Service (NOAA Fisheries), which administer all ESA activity, are not obligated to incorporate hypothesis tests in ESA reviews. The ESA mandates a more precautionary approach, stipulating that the USFWS and NOAA Fisheries must "insure that any action...is not likely to jeopardize the continued existence of any endangered or threatened species" (16 U. S. C. 1536[a][2]). Furthermore, the federal courts are generally willing to uphold USFWS/NOAA Fisheries decisions that are based on professional discretion, rather than explicit hypothesis tests. So long as these agencies adhere to ESA procedures, taking into consideration all of the relevant data available, and offer rational explanations for why

Daniel J. McGarvey (e-mail: mcgar002@bama.ua.edu) is a PhD candidate in the Department of Biology, University of Alabama, Tuscaloosa, AL 35487. © 2007 American Institute of Biological Sciences.

Figure 1. The four possible outcomes of a hypothesis test, relative to an Endangered Species Act listing decision. "Endangered" status is the alternative hypothesis, and therefore carries the burden of proof, while "nonendangered" status is the null hypothesis, and therefore receives the benefit of the doubt. If the candidate species is not truly endangered, but the listing review is approved, a type I error will occur, and unnecessary regulations will be the result. However, if the candidate species is endangered, but the listing review is rejected, a type II error will occur, and the species will be denied necessary protection.

particular sources of information or differing conclusions are favored over others, their decisions tend to withstand judicial review (Sidle 1998, Brennan et al. 2003).

The ESA is not, however, a panacea for species conservation (Norris 2004). Critics are quick to point out that many ESA regulations have (so far) been marginally successful in promoting species' recoveries, and that the underlying science is rarely as comprehensive as the work presented in peer-reviewed journals (Pombo 2004, Buck et al. 2005). Indeed, their call for more "sound science," which has become a cornerstone of congressional attempts to reform the ESA (Brennan et al. 2003, Buck et al. 2005), does raise an interesting question: Given the financial burdens that are typically involved, is it prudent to enforce ESA regulations that have not been substantiated through a stringent peer-review process? It therefore behooves ESA supporters to communicate the risks that a more conservative ESA would entail, and to discuss possible alternatives.

This article demonstrates why hypothesis testing must be used cautiously in ESA science, and explores an alternative method, equivalence testing, that could be used to evaluate ESA studies in an equally quantitative, peer-reviewed fashion. Hypothesis testing and equivalence testing are similar statistical procedures, but they differ in how they assign the burden of proof. To make the comparison, I examine a recent review of ESA regulations in the Upper Klamath Lake region of southern Oregon. The Upper Klamath Lake review is an ideal case study because it is one of the few instances in which a strict hypothesis-testing approach has been used (NRC 2004). It has also been cited as a potential model for future ESA reviews (Manson 2002).

The Upper Klamath Lake case study

Upper Klamath Lake is the primary habitat of two federally endangered fishes: the Lost River sucker (*Deltistes luxatus*) and the shortnose sucker (*Chasmistes brevirostris*). The impaired status of these species is attributed to commercial and recreational harvest (now limited to a single tribal fishery), entrainment in irrigation facilities, habitat losses, predation from and competition with invasive species, and, most importantly, low dissolved oxygen levels (NRC 2004).

Dissolved oxygen depletion is the most immediate threat to endangered sucker survival, as evidenced by its causative role in three consecutive (1995, 1996, 1997) fish kills—kills that may have eliminated up to 50 percent of the adult Lost River sucker and shortnose sucker populations (NRC 2004). This depletion is the result of a persistent annual algal bloom, dominated by the blue-green alga *Aphanizomenon flos-aquae.* Most summers, a massive algal bloom, followed by the senescence and decay of superimposed algal tissue, drives the dissolved oxygen in Upper Klamath Lake to critically low levels (< 1 to 2 mg per L), creating a potentially lethal environment for both of the endangered fishes (NRC 2004).

In 2001, the USFWS recommended that specific minimum water levels (expressed as lake elevations, relative to mean sea level) be maintained in Upper Klamath Lake (USFWS 2001). These guidelines, which were intended to increase habitat quality (i.e., by mitigating algal blooms) and quantity (i.e., by inundating additional nearshore spawning and rearing habitat) for endangered suckers, were predicated on a series of logical assumptions regarding the dynamics of algal growth (USFWS 2001). For example, maintaining higher water levels might constrain algal densities through a dilution effect, or by inhibiting wind-driven phosphorus recruitment (i.e., upwelling) from the lake's benthic sediments. It is important to note, however, that the USFWS recommendations stemmed largely from general ecological principles and studies of analogous systems ("The following chain of causal relationships and mechanisms, which is supported by the scientific literature, is characteristic of hypereutrophic lake systems such as Upper Klamath Lake"; USFWS 2001, section III, part 2, p. 72), as empirical data from Upper Klamath Lake were not readily available.

The following year, a National Research Council (NRC) review of the USFWS recommendations was commissioned by the US Department of the Interior. To assess the strength of the evidence underlying the USFWS lake elevation prescriptions, the NRC review examined nine years ($n = 9$) of maximum chlorophyll *a* concentration (a surrogate measure of algal density) and Upper Klamath Lake elevation data (figure 2a). (Only nine years of empirical data were available when the review was conducted.) This was a test to determine whether algal density tended to decrease as lake elevation increased (NRC 2004). The NRC review concluded that "there is no scientific support for the proposition that higher water levels correspond to better water quality" (NRC 2004), as an inverse relationship between lake elevation and chlorophyll *a* was not readily apparent (figure 2a). (Complete

background information on endangered species management in Upper Klamath Lake is provided in USFWS [2001], NRC [2004], and references therein.)

The hypothesis-testing approach. Although the NRC review did not report formal statistical results, its analysis was, in effect, a hypothesis test. Specifically, it used a linear regression approach (a type of hypothesis test) to compare the following null (H_0) and alternative (H_A) hypotheses:

H_0: Slope of the chlorophyll *a* regression ≥ 0.

H_A: Slope of the chlorophyll *a* regression < 0.

Linear regression determines whether some variable of interest (the dependent variable) can be calculated as a function of a second variable (the independent, or predictor, variable). More precisely, it determines whether the relationship between two variables (assuming there is one) can be described with a straight line. It does so by determining the slope and position of the *least-squares line;* this line simultaneously minimizes the sum of the squared vertical distances between itself and each of the data points (i.e., it minimizes the average vertical offset; see figure 2a). Linear regression then uses two statistics to evaluate the fit and reliability of the least-squares line: the *coefficient of determination* and the *P-value.* The coefficient of determination (r^2), which ranges from 0 to 1, measures the degree of straight-line association between the independent and dependent variables. Generally speaking, large r^2 values indicate that the least-squares line is likely to be a good fit for the data in question. The *P*-value, which also ranges from 0 to 1, is a measure of the evidence in support of H_0. In hypothesis tests, small *P*-values are necessary to reject H_0 ($P \leq 0.05$ is the traditional cutoff value in peer-reviewed research; see Gotelli and Ellison [2004] for complete details on linear regression.)

Formal linear regression analysis (i.e., hypothesis testing) of the Upper Klamath Lake data corroborates the NRC review's conclusion that there is minimal evidence of an inverse relationship (a least-squares line with a negative slope) between lake elevation and chlorophyll *a*. Although the slope of the least-squares line is negative (–31.5), the r^2 value is relatively small (0.08), and the *P*-value is relatively large (0.76; figure 2a). (The linear model residuals are approximately normally distributed [$P > 0.15$], with constant variances.) These results indicate that the least-squares line does not fit the data particularly well, and that there is little reason to believe H_0 (no inverse relationship between lake elevation and chlorophyll *a*) is incorrect. In quantitative terms, this is why the NRC review disagreed with the USFWS lake elevation regulations; given the available data, it seemed that accepting H_A entailed a high risk of type I error, or unnecessary regulation.

The hypothesis-testing approach is common in peer-reviewed research, but it is inappropriate in the ESA context, for both of the reasons discussed earlier: It places the burden of proof on the USFWS to demonstrate that regulations are

necessary to protect listed species, and it does not take into consideration the practical limitations of small data sets. (Nine observations constitute a small data set by virtually any
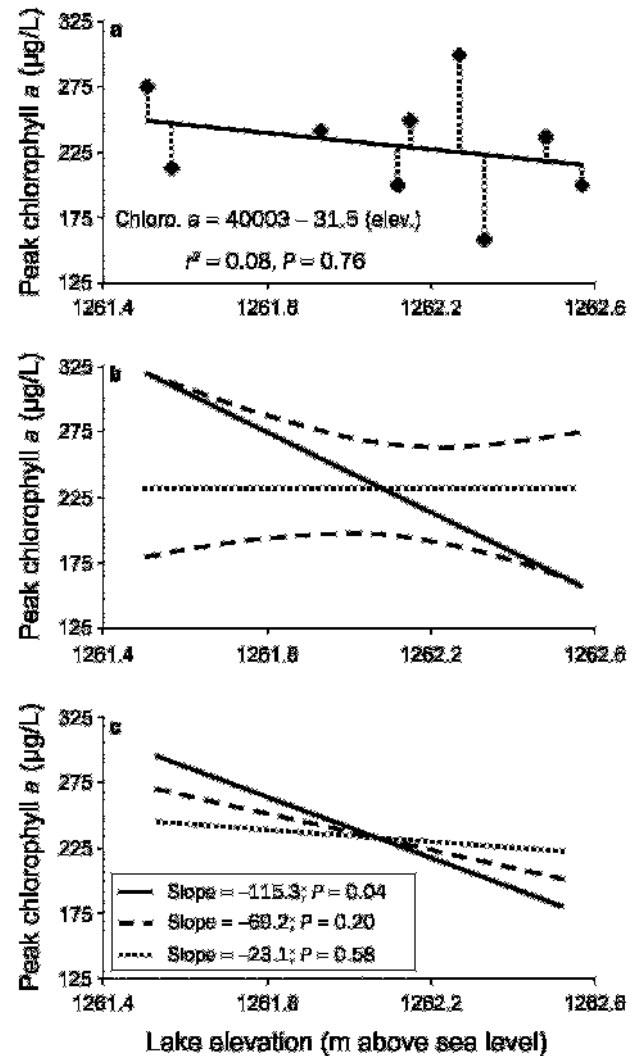

Figure 2. Results of the case study at Upper Klamath Lake, southern Oregon. (a) Maximum chlorophyll **a** *concentration and lake elevation data, as analyzed in the National Research Council review. The least-squares line (solid line) and regression results (linear model equation, coefficient of determination, and* **P***-value) are shown with the empirical data. Dashed vertical lines illustrate the residual distances between each data point and the least-squares line. (b) Ninety-five percent confidence bands (dashed, curved lines) are shown for the linear regression, as well as the zero-slope line (dotted line) and the minimum detectable slope (solid line). (c) Results of the three equivalence tests (*H_0*, slope ≤ –115.3;* H_0*, slope ≤ –69.2;* H_0*, slope ≤ –23.1). Each of the lines is centered at the mean chlorophyll* **a** *concentration (230.5 µg per L) and mean lake elevation (1262.1 m). All plots are shown at identical scales. Data are from Welch and Burke 2001, as shown in NRC 2004.*

statistical criterion.) How the burden of proof should be assigned in ESA reviews is a normative question (NRC 1995, Ludwig et al. 2001), and one that section 7(a)(2) of the ESA has largely answered. (See also H.R. Conference Report 96-697, 96 Cong., 1st. sess. 12 [1979], which explicitly directs the USFWS to "give the benefit of the doubt to the species.") Continuing debate on the meaning of "best data available" suggests (e.g., Brennan et al. 2003, Doremus 2004, Ruhl 2004), however, that the problems associated with small data sets are poorly understood, and therefore in need of further discussion.

Statistical power. Small data sets are problematic because they tend to lack *statistical power,* which is the likelihood of detecting a significant effect or relationship when it does, in fact, exist (i.e., the probability of rejecting H_0 when it is actually false). Statistical power is a function of three factors: the level of certainty one requires to reject H_0 (inversely proportional to statistical power), the number of samples one is working with (directly proportional to statistical power), and the size of the effect one is trying to detect (directly proportional to statistical power). Therefore, if a researcher wishes to be highly certain (e.g., 95 percent) not to commit a type I error, is working with a small data set, and is trying to detect a relatively small and inconspicuous effect, the results will have low statistical power and be prone to type II error. To increase statistical power, researchers must be willing to increase the risk of a type I error, obtain a larger data set, or refocus their search on a larger, more readily detected effect (Peterman 1990, Taylor and Gerrodette 1993).

The limitations of small data sets are intuitively illustrated with *confidence bands.* For example, figure 2b displays 95 percent confidence bands for the Upper Klamath Lake regression. These bands encompass the entire range of least-squares lines (centered at the mean lake elevation and mean chlorophyll *a* values) that *cannot* be distinguished from the zero-slope line (which depicts the absence of a linear relationship between lake elevation and chlorophyll *a*), given the available data and a 95 percent certainty criterion (i.e., the *P*-value cannot be greater than 0.05). Importantly, while the 95 percent confidence bands do include the zero-slope line (dotted, horizontal line in figure 2b), they also include many negative-slope lines (as well as a number of positive-slope lines), any of which might reflect important biological relationships. In this particular instance, it would be impossible to detect an inverse relationship between lake elevation and chlorophyll *a* (i.e., to reject H_0 in favor of H_A), with 95 percent confidence that a type I error would not be committed, unless the slope of the least-squares line was at least –153 (figure 2b). Thus, it is not compelling to note that there is currently no evidence of a negative relationship between lake elevation and chlorophyll *a* concentration in Upper Klamath Lake; the width of the confidence bands clearly indicates that the available data are insufficient (i.e., statistical power is too low) to detect anything less than a dramatic, highly improbable relationship (Hoenig and Heisey 2001).

Equivalence testing. One solution to the hypothesis-testing problem is the *equivalence test.* Equivalence tests switch the burden of proof by making the effect of concern (e.g., an inverse relationship between lake elevation and chlorophyll *a*) the null hypothesis, and making the "no effect" conclusion the alternative hypothesis (McBride 1999, Parkhurst 2001, Dixon and Pechmann 2005). In this way, imperiled species receive the benefit of the doubt, as opponents of regulation are required to prove that a proposed guideline is not necessary. Equivalence tests, which are otherwise analogous to standard hypothesis tests (Dixon and Pechmann 2005), also force researchers to focus on specific, explicitly defined effects. This allows equivalence test results to be integrated in ground-level policy more easily than general "no effect" hypothesis tests (Parkhurst 2001).

For example, the USFWS might wish to determine whether raising the elevation of Upper Klamath Lake by 1 m is likely to reduce chlorophyll *a* concentration by at least 10 percent. (Raising the elevation by 1 m would increase the total water volume substantially, because of the lake's gradual bathymetry and large surface area [approximately 360 km^2 when full; NRC 2004].) If the average values for chlorophyll *a* (230.5 µg per L) and lake elevation (1262.1 m) were used as a baseline, a 10 percent reduction in chlorophyll *a* (230.5 – [0.1 • 230.5] = 207.5 µg per L, at 1262.1 + 1.0 = 1263.1 m elevation) would amount to a linear relationship with a slope of –23.1 (figure 2c). This relationship and the nine available data points could then be used in a one-sided equivalence test (referred to as a "reverse test" by Parkhurst [2001]) of the following hypotheses (note that the proposed inverse relationship is now H_0):

H_0: Slope of the chlorophyll *a* regression $\leq$ –23.1.

H_A: Slope of the chlorophyll *a* regression > –23.1.

The *P*-value for this test is 0.58. (The equivalence test *P*-value is the one-tailed probability of observing a slope greater than –23.1, under a *t* distribution with *n* – 2 degrees of freedom; see Parkhurst [2001].) Hence the USFWS could assert that there is no reason to believe H_0 is false (i.e., *P* >> 0.05).

Additional equivalence tests could also be performed to help the USFWS evaluate a range of management options. For example, if 30 percent (H_0, slope $\leq$ –69.2) and 50 percent (H_0, slope $\leq$ –115.3) decreases in chlorophyll *a* (relative to the proposed 1-m lake elevation increase) were tested, the USFWS would obtain *P*-values of 0.20 and 0.04, respectively (figure 2c). It could then conclude that increasing lake elevation by 1 m is highly unlikely to reduce chlorophyll *a* by 50 percent, but could not rule out a 30 percent reduction (assuming that 95 percent certainty, or *P* $\leq$ 0.05, is required to reject H_0). These procedures would relieve the USFWS of the burden of proof, and would facilitate a more comprehensive review process by providing a rational basis for choosing among multiple objectives.

Precautionary science

As the scientific community has become increasingly aware of the risks and limitations of hypothesis testing (e.g., Yoccoz 1991, NRC 1995, Anderson et al. 2000, Ludwig et al. 2001), many of its members have chosen to endorse the "precautionary principle" (e.g., Carroll et al. 1996, SCB 2005). The precautionary principle recognizes that hypothesis tests, which are designed to prevent type I error, are frequently at odds with conservation objectives, which aim to minimize type II error (figure 1). The essence of the precautionary principle is the normative belief that scientific uncertainty should not constrain efforts to protect imperiled species (or human livelihood), particularly when the threat of irreversible damage exists (Kinzig et al. 2003).

The NRC review considered the precautionary principle, but ultimately rejected it on the grounds that "whether to apply the Precautionary Principle is a policy decision and as such is outside the present committee's scope of work, which pertains to 'whether the [USFWS recommendations were] consistent with the available scientific information'" (NRC 2004). Law professor J. B. Ruhl, who was a coauthor of the NRC review, also suggested that the precautionary principle is inherently nonscientific (Ruhl 2004):

> By demanding rigorous empirical testing and confirmation...the Scientific Method hopes to reduce Type I error in the form of unjustified protection of species. By calling for protective action *without undergoing the complete battery of Scientific Method tests,* the Precautionary Principle Method hopes to reduce Type II error in the form of underprotection of species. [Emphasis added]

These statements reflect a subtle but important misunderstanding of the precautionary principle. While it is true that the origins of the precautionary principle are more political than scientific (Foster et al. 2000), there is no epistemological reason why it cannot be employed as a legitimate standard in scientific research. So long as it is articulated with defensible, quantitative techniques, the precautionary principle is every bit as "scientific" as the more conservative hypothesis-testing approach.

The equivalence tests described above provide just one example of how science can adopt a more precautionary strategy without sacrificing analytical rigor. Similar examples are discussed in McBride (1999), Parkhurst (2001), Dixon and Pechmann (2005), and references therein. Bayesian techniques also provide quantitative alternatives to hypothesis testing, with the added benefit of incorporating results from similar but independent studies (Bayesian "prior probabilities"). (Bayesian methods are beyond the scope of this article, but excellent introductions are provided by Gotelli and Ellison [2004] and Ellison [2004].) Regardless of the precise analytical method that is chosen, it is critical for scientists and policymakers to realize that hypothesis testing is not the only way to achieve sound science.

Conclusions

A strict hypothesis-testing protocol that disregarded the likelihood and ramifications of type II error would be a logically, ethically, and legally unacceptable standard for ESA reviews (Ludwig et al. 2001, 16 U.S.C. 1536[a][2]). In ESA science, the consequence of a type II error (when H_0 amounts to the conclusion that protective action is not necessary) will fall somewhere between the failure to assist an imperiled species and the unintentional abetting of its extinction. But an indiscriminate tolerance for type I error would not be appropriate either, as type I errors are also likely to cause significant losses and should be avoided whenever possible (NRC 1995). An optimal strategy for ESA reviews will therefore require careful, case-by-case assessment of the available information, with the full understanding that hypothesis tests are only one of the many tools available to the scientific community. At a minimum, scientists must assess what is currently known, what can realistically be determined, and what is at stake (Ludwig et al. 2001, Kinzig et al. 2003). By addressing these questions in a more explicit, transparent manner, the scientific community can continue to provide ESA administrators with scientifically sound counsel, without limiting themselves to the hypothesis-testing methods and assumptions that work so well in more conventional research.

Acknowledgments

References cited

Anderson DR, Burnham KP, Thompson WL. 2000. Null hypothesis testing: Problems, prevalence, and an alternative. Journal of Wildlife Management 64: 912–923.

Brennan MJ, Roth DE, Feldman MD, Greene AR. 2003. Square pegs and round holes: Application of the "best scientific data available" standard in the Endangered Species Act. Tulane Environmental Law Journal 16: 387–444.

Buck EH, Corn ML, Sheikh P, Baldwin P, Meltz R. 2005. The Endangered Species Act (ESA) in the 109th Congress: Conflicting Values and Difficult Choices. Washington (DC): Library of Congress. (28 November 2006; *www.ncseonline.org/NLE/CRSreports/05apr/IB10144.pdf*)

Carroll R, Augspurger C, Dobson A, Franklin J, Orians G, Reid W, Tracy R, Wilcove D, Wilson J, Lubchenco J. 1996. Strengthening the use of science in achieving the goals of the Endangered Species Act: An assessment by the Ecological Society of America. Ecological Applications 6: 1–11.

Dixon PM, Pechmann JHK. 2005. A statistical test to show negligible trend. Ecology 86: 1751–1756.

Doremus H. 2004. The purposes, effects, and future of the Endangered Species Act's best available science mandate. Environmental Law 34: 397–450.

Ellison AM. 2004. Bayesian inference in ecology. Ecology Letters 7: 509–520.

Foster KR, Vecchia P, Repacholi MH. 2000. Science and the Precautionary Principle. Science 288: 979–981.

Gotelli NJ, Ellison AM. 2004. A Primer of Ecological Statistics. Sunderland (MA): Sinauer.

Hoenig JM, Heisey DM. 2001. The abuse of power: The pervasive fallacy of power calculations for data analysis. American Statistician 55: 19–24.

Kinzig A, et al. 2003. Coping with uncertainty: A call for a new science-policy forum. Ambio 32: 330–335.

Kuhn T. 1970. The Structure of Scientific Revolutions. Chicago: University of Chicago Press.

Ludwig D, Mangel M, Haddad B. 2001. Ecology, conservation, and public policy. Annual Review of Ecology and Systematics 32: 481–517.

Manson C. 2002. Testimony of Craig Manson, Assistant Secretary for Fish and Wildlife and Parks, U.S. Department of the Interior, before the House Resources Committee, regarding H.R. 4840. (9 December 2006; *www.fws.gov/laws/Testimony/107th/2002/2002june19.htm*)

McBride GB. 1999. Equivalence tests can enhance environmental science and management. Australian and New Zealand Journal of Statistics 41: 19–29.

Norris S. 2004. Only 30: A portrait of the Endangered Species Act as a young law. BioScience 54: 288–294.

[NRC] National Research Council. 1995. Science and the Endangered Species Act. Washington (DC): National Academies Press.

———. 2004. Endangered and Threatened Fishes in the Klamath River Basin: Causes of Decline and Strategies for Recovery. Washington (DC): National Academies Press.

Parkhurst DF. 2001. Statistical significance tests: Equivalence and reverse tests should reduce misinterpretation. BioScience 51: 1051–1057.

Peterman RM. 1990. Statistical power analysis can improve fisheries research and management. Canadian Journal of Fisheries and Aquatic Sciences 47: 2–15.

Pombo RW. 2004. The ESA at 30: Time for Congress to Update and Strengthen the Law. Committee Report. Washington (DC): US House of Representatives Committee on Resources. (29 November 2006; *http://resourcescommittee.house.gov/issues/more/esa/whitepaper.htm*)

Ruhl JB. 2004. The battle over Endangered Species Act methodology. Environmental Law 34: 555–603.

[SCB] Society for Conservation Biology. 2005. Society for Conservation Biology code of ethics. Conservation Biology 19: 3.

Shrader-Frechette KS, McCoy ED. 1992. Statistics, costs and rationality in ecological inference. Trends in Ecology and Evolution 7: 96–99.

Sidle JG. 1998. Arbitrary and capricious species conservation. Conservation Biology 12: 248–249.

Taylor BL, Gerrodette T. 1993. The uses of statistical power in conservation biology: The vaquita and northern spotted owl. Conservation Biology 7: 489–500.

Toft CA, Shea PJ. 1983. Detecting community-wide patterns: Estimating power strengthens statistical inference. American Naturalist 122: 618–625.

[USFWS] US Fish and Wildlife Service. 2001. Biological/Conference Opinion Regarding the Effects of Operation of the Bureau of Reclamation's Klamath Project on the Endangered Lost River Sucker (*Deltistes luxatus*), Endangered Shortnose Sucker (*Chasmistes brevirostris*), Threatened Bald Eagle (*Haliaeetus leucocephalus*) and Proposed Critical Habitat for the Lost River/Shortnose Suckers. Washington (DC): US Department of the Interior. (29 November 2006; *www.usbr.gov/mp/kbao/esa/34_final_sucker_bo_4_06_01.pdf*)

Welch EB, Burke T. 2001. Interim Summary Report: Relationship between Lake Elevation and Water Quality in Upper Klamath Lake, Oregon. Portland (OR): Bureau of Indian Affairs.

Yoccoz NG. 1991. Use, overuse, and misuse of significance tests in evolutionary biology and ecology. Bulletin of the Ecological Society of America 72: 106–111.