



Using Multiple Treatment Levels as a Means of Improving Inference in Wildlife Research

Todd D. Steury; Aaron J. Wirsing; Dennis L. Murray

The Journal of Wildlife Management, Vol. 66, No. 2. (Apr., 2002), pp. 292-299.

Stable URL:

<http://links.jstor.org/sici?sici=0022-541X%28200204%2966%3A2%3C292%3AUMTLAA%3E2.0.CO%3B2-V>

The Journal of Wildlife Management is currently published by Allen Press.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/acg.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

The JSTOR Archive is a trusted digital repository providing for long-term preservation and access to leading academic journals and scholarly literature from around the world. The Archive is supported by libraries, scholarly societies, publishers, and foundations. It is an initiative of JSTOR, a not-for-profit organization with a mission to help the scholarly community take advantage of advances in technology. For more information regarding JSTOR, please contact support@jstor.org.

USING MULTIPLE TREATMENT LEVELS AS A MEANS OF IMPROVING INFERENCE IN WILDLIFE RESEARCH

TODD D. STEURY,^{1,2} Department of Fish and Wildlife Resources, University of Idaho, Moscow, ID 83844, USA
AARON J. WIRSING,³ Department of Fish and Wildlife Resources, University of Idaho, Moscow, ID 83844, USA
DENNIS L. MURRAY, Department of Fish and Wildlife Resources, University of Idaho, Moscow, ID 83844, USA

Abstract: Under the most common experimental protocol, which uses comparison of control and experimental groups subject to a single level of manipulative treatment, conclusions regarding the influences of manipulation are restricted to the presence or absence of an effect. An alternative approach, in which the experimental variable is invoked across multiple treatment levels, should improve biological inference by further clarifying the relationship between the independent and dependent variables. Mathematical analyses comparing relative precision indicate that the experimental design with greater standard deviation in the independent variable will have greater precision. If the precision of 2 experimental designs is similar, or if the statistical power of a given design is high or low, power is changed minimally by modifying the number of treatment levels. Conversely, if the power of 1 experimental design is intermediate, then the alternative design may have substantially more or less power depending on the relative precision of the 2 models. In a review of 134 food supplementation experiments, we found that 95% of these studies were designed with a single treatment level, implying that most of these studies revealed little about the relationship between food abundance and dependent variables. Researchers should evaluate the relative merits of study designs with single or multiple treatment levels a priori, and apply the protocol that is most likely to provide statistical or inferential advantages.

JOURNAL OF WILDLIFE MANAGEMENT 66(2):292–299

Key words: experimental design, precision, regression, replication, statistical power, treatment selection.

The means by which ecological experiments are designed and analyzed recently have been the subject of notable attention (Peters 1983, 1991; Murphy and Noon 1991; Yoccoz 1991; Petraitis 1998; Johnson 1999). Criticism has focused largely on the failure of ecological research to produce reliable knowledge that can be used to understand theory or prescribe effective management protocols (Romesburg 1981, 1991; Nudds and Morrison 1991; Goldberg and Scheiner 1993). The problem is especially prevalent in wildlife research, where the scale and variability associated with the study of free-ranging animals usually hinders the formulation of definitive conclusions.

The questionable record of accomplishment of many wildlife studies may be partially a function of experimental design. The most widely used experimental design involves comparing control and experimental groups with respect to a single level of treatment or element of natural variation. This experimental protocol, however, produces results that pertain only to the fixed treatment

level and consequently have limited value in the context of ecological theory, management, or conservation, due to a lack of relevance to other populations or systems (Peters 1991, Goldberg and Scheiner 1993, Potvin 1993).

For example, if a researcher wishes to determine the effect of commercial timber harvest on the survival of a species occupying forested habitats, the researcher typically compares the survival rates of multiple populations (replicates) occupying harvested and nonharvested (control) study areas (Fig. 1A). This method may indicate whether or not the treatment effect is statistically and biologically significant. However, in this example, no numerical information regarding the manipulation is provided; the level of treatment is simply a categorical variable: harvested or not harvested. Thus, the results obtained are of limited value because they may apply only to the level of treatment under investigation.

One relatively simple alternative to this approach is to distribute the levels of treatment across a gradient rather than replicate at the single level of manipulation (see Goldberg and Scheiner 1993). This method should be especially enlightening when data are analyzed using linear or non-linear regression techniques. Returning to the timber harvest example, the researcher could disperse the levels of treatment across a range of

¹ Present address: Department of Life Sciences, Indiana State University, Terre Haute, IN 47809, USA.

² E-mail: lssteury@scifac.indstate.edu

³ Present address: Behavioural Ecology Research Group, Department of Biological Sciences, Simon Fraser University, Burnaby, BC V5A 1S6, Canada.

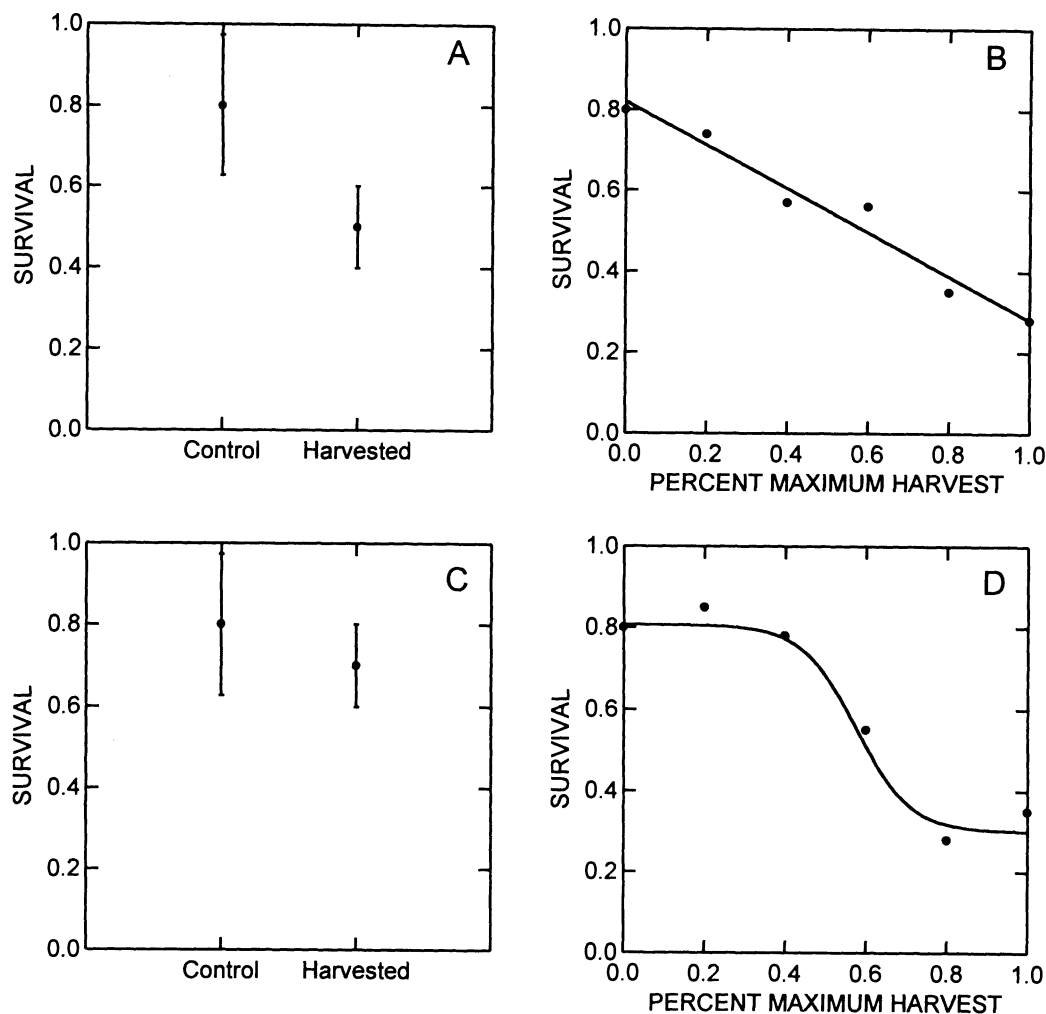


Fig. 1. Hypothetical relationships between 1 demographic variable of interest (survival) and level of timber harvest under experimental designs with single and multiple levels of treatment. The same total number of samples was used in each hypothetical experiment. (A) Results obtained using an experimental design with 1 level of treatment. The control is no harvest. Error bars are 95% CI. (B) Results obtained using an experimental design with multiple levels of treatment. (C) Results obtained using a single-treatment-level design with a low level of experimental treatment (20% maximum harvest). (D) Results obtained using a multiple-treatment-level design in which a nonlinear relationship between treatment level and the effect on survival is indicated.

harvest levels, including 1 or more control population(s) that is/are not harvested. The researcher could then use linear (or nonlinear) regression to relate survival rates to the range of harvest levels (Fig. 1B). In this case, the independent variable consists of absolute measures of the extent of harvesting. Thus, when applicable, experimental designs with multiple treatment levels should be superior to those with only 1 treatment level because they better describe the relationship between the independent and dependent variables, and thereby provide predictive ability.

BENEFITS OF MULTIPLE TREATMENT LEVELS

In addition to greater inference and predictive ability, other important benefits are associated with using multiple treatment levels. First, experiments using multiple treatment levels have a greater ability to address levels of treatment that result in relatively small effects. Returning to the example of timber harvest, the researcher may want to know the effect of light commercial harvest (10–20% of maximum harvest) on survival of

a species. From the experiment with multiple treatment levels, such an effect can be interpolated rather easily without the explicit need for a new experiment conducted at the desired harvest level (e.g., Fig. 1B). In contrast, experiments with only 1 level of treatment (or multiple-treatment-level experiments analyzed with analysis of variance [ANOVA]) would not offer similar predictive power and would require a new experiment specifically at the desired level of harvest. In addition, with only 1 level of treatment, the ability of a researcher to detect the effect of small changes in an independent variable may be limited to experiments with large sample sizes or detectable effect sizes. Fig. 1C demonstrates 1 possible result from such a manipulation with 1 level of treatment. In this case, inferential statistics would reveal no statistically significant difference between survival of populations occupying harvested and nonharvested stands, although such an effect may truly exist and be biologically significant.

Second, experiments using multiple treatment levels and analyzed with regression give the researcher the potential to detect nonlinear effects. Experiments with only 1 level of treatment (or experiments with multiple treatment levels but analyzed with ANOVA) reveal nothing about the relationship between the variable being tested and the effect it has on populations or individuals, only whether or not a relationship might exist. For example, in situations where there is a threshold level of commercial timber harvest beyond which survival is affected dramatically (Fig. 1D), results obtained using 1 level of treatment would be dependent on the level of harvest used in the experiment. If the level of harvest was below the threshold, no effect would be demonstrated, whereas above the threshold, the influence of harvest would be detected. However, the existence of the threshold itself would not be revealed.

Finally, use of multiple treatment levels effectively eliminates the need to justify a particular level of manipulation. Because of the often arbitrary nature of treatment-level selection, conclusions drawn from experiments using only 1 level of treatment may not be indicative of natural systems if the level of manipulation was inappropriate (Goldberg and Scheiner 1993). Indeed, researchers designing experiments with 1 treatment level often choose an extreme level of treatment to maximize the probability of detecting an effect. For example, food often is supplemented to an experimental population *ad libitum* to test for the presence of food limitation (see Boutin

1990). The differences between control and experimental groups demonstrated by these experiments reveal little about how the populations under investigation respond to natural, relatively minor changes in food abundance. Thus, any general conclusions drawn from these experiments about the effects of natural food variation are open to criticism. With regard to experiments using multiple treatment levels, criticisms about experimental conclusions primarily are limited to instances when one must generalize beyond the range covered by the regression equation or when the range of treatment levels does not represent natural variation (see Petraitis 1998).

LITERATURE REVIEW

Although the benefits of experimental designs with multiple treatment levels may seem obvious and intuitive to many researchers, such designs are far from pervasive in wildlife research. We categorized the study design protocols of published field experiments involving food supplementation to free-ranging terrestrial vertebrates ($n = 134$ published experiments). We chose food supplementation experiments because of their frequency in the wildlife literature, the ease with which food quality or quantity can be manipulated at single or multiple levels, and the availability of a recently published review (Boutin 1990). In addition to surveying the major reference on this topic (Boutin 1990), we also reviewed recently (i.e., 1989–2000) published literature dealing with food supplementation experiments in free-ranging terrestrial vertebrates.

Our review revealed that only 6 (4.5%) studies used multiple levels of food supplementation. Of these, 4 studies used analysis of variance (ANOVA) to compare the effects of different treatment levels when a regression-type approach may have been more revealing. Thus, only a small minority of food supplementation studies have manipulated food at multiple levels and used regression to examine the relationship between dependent and independent variables.

The apparent paucity of experimental designs with multiple treatment levels may be due to apprehension researchers might have about such designs. Researchers may believe that such experiments require a greater number of samples or otherwise have lower statistical power and precision. Alternatively, researchers may believe that replication is of paramount importance in experiments; a theme that tends to be reinforced in many statistics courses. Finally, researchers may

have difficulty in deciding how many treatment levels to use in an experiment. Thus, we (1) discuss how the number of treatment levels can influence statistical power and precision, (2) examine the importance of replication to both single- and multiple-treatment-level experimental designs, and (3) propose a conceptual framework by which researchers can decide the number of treatment levels to use in their experiments. Although many alternatives exist to the classic, 1-treatment-level experiment (see Kamil 1988, Eberhardt and Thomas 1991, Ratti and Garton 1994 for examples), here we only discuss the use of multiple treatment levels as a means of improving inference in wildlife research because of the rather large benefit that can be gained from a relatively small and simple change in experimental design.

PRECISION AND POWER

We used linear regression analysis to demonstrate how the number and placement of treatment levels can influence precision in the estimate of an effect. In an experiment with only 1 treatment level (plus a control), significance tests typically are conducted using *t*-tests or ANOVA. However, regression analysis (i.e., the general linear model) also may be used to test for significant effects in experiments designed with only 1 level of treatment. Thus, we compared the relative precision of experiments utilizing 1 level of treatment plus control with those utilizing multiple levels of treatment using linear regression. We assumed that the values of the independent variable are known, rather than coded, and that the slopes and standard deviation in error would not be influenced by the number of levels of treatment. Furthermore, we assumed that an equal total number of samples would be used in either design. Finally, we assumed that the relationship between dependent and independent variables was linear.

In linear regression, precision is related to the standard error (SE) of the estimate of the slope of the regression line fitted to the data. The standard error of the slope is described by

$$SE = \frac{s_e}{\sqrt{\sum_i (x_i - \bar{x})^2}} \quad (1)$$

where s_e is the standard deviation of the error, x_i are the independent values of the study samples (i.e., levels of treatment or control), and \bar{x} is the mean value of the independent variable (i.e.,

mean level of treatment including controls). The denominator in equation 1 is equal to the product of the standard deviation in the independent variable (s_x) and the square root of the number of samples minus 1. Thus, equation 1 can be simplified to

$$SE = \frac{s_e}{s_x \sqrt{n-1}} \quad (2)$$

Because the standard deviation of the error should be identical regardless of the number of treatment levels used, and we assumed the total number of samples are the same, the relative precision of a given experimental design is a function of the deviation in the independent variables (i.e., treatment and control levels).

Over a given range in the level of manipulation, an experimental design that uses 1 treatment level (plus control) at the extremity of that range will maximize the standard deviation in the independent variable and thus minimize the standard error of the estimate of the slope of the regression line (i.e., maximizing precision). Thus, for example, an experiment that examines the demographic influence of timber harvest at 0% and 100% removal rates, each replicated twice ($s_x = 0.58$), should have greater precision than an experiment that compares the demographic influence of timber harvest at 0%, 33%, 67%, and 100% removal ($s_x = 0.43$). However, if the range in the independent variable can be increased, then the standard deviation in x -values will also increase (as will model precision). A timber harvest experiment with 0%, 25%, 50%, and 75% removal rates ($s_x = 0.32$) will have slightly greater precision than the same experiment with 0% and 50% removal rates each replicated twice ($s_x = 0.29$). Thus, for a fixed number of study units, an experiment using only 1 level of treatment will have the greatest precision over a given range. However, an experiment involving multiple treatment levels may have greater precision if the range in treatment levels exceeds that used in an identical experiment involving only a single level of treatment.

In many ecological experiments, however, there are biological and logistical limits to the levels of manipulation (e.g., timber harvest cannot exceed 100%, food never can be effectively supplemented beyond ad libitum, one never can remove more than all the predators, parasites, competitors). In these systems, designing an experiment with multiple treatment levels that

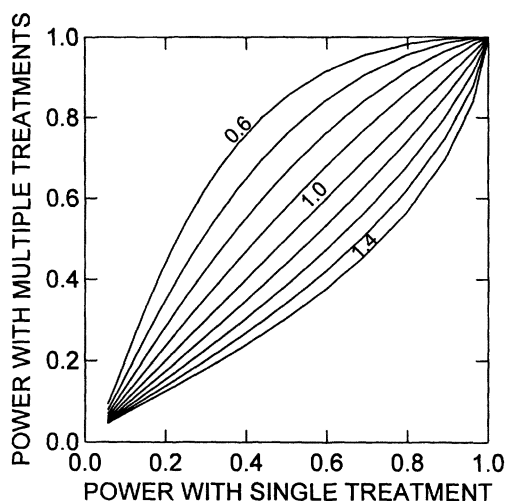


Fig. 2. Relationship between the power of 2 experimental designs, 1 with a single level of treatment and 1 with multiple levels of treatment. The values in the center of the graph represent the relative precision of the 2 models (i.e., SE of slope of design with >1 level of treatment/SE of slope of design with 1 level of treatment).

matches or exceeds the expected precision of the same experiment with only 1 level of treatment may be impossible. Nevertheless, the benefits associated with multiple-treatment-level designs may outweigh the expected loss in precision if the design has sufficient power to detect a statistically significant effect.

The power of a statistical test is the probability that the test will correctly reject the null hypothesis when it is false (Zar 1996). Power and precision are both functions of sample size and variance in the dependent variable. Power, however, also is a function of the detectable effect size, and as such is influenced by the randomness of data sampling. If the precision of 2 experimental designs (1 with a single level of treatment, the other with multiple levels of treatment) is the same, then the power of the 2 designs will be similar (Fig. 2). As the relative precision of the 2 experimental designs diverges from parity, the relative power of the 2 designs generally diverges as well (Fig. 2). If either design has very high power (close to 1.0), or very low power (close to 0.0), then limited power is lost (or gained) in switching to the other design (Fig. 2). Alternatively, if either design has moderate power, then a substantial amount of power may be lost (or gained) by switching to the other model, depending on the relative precision of the 2 designs.

The above conclusions regarding precision and power are dependent on an assumption of linearity. If the true relationship between the dependent and independent variables deviates from linearity, and linear regression is used to analyze data from an experiment with multiple treatment levels, the estimates of effect (slope) will be subject to lower precision, and more importantly, will be biased. Power and precision in a single-treatment-level design (plus control) are unaffected by the true shape of the relationship. Similarly, if a multiple-treatment-level design is used, power and precision will be unaffected by the shape of the relationship if the data are analyzed using ANOVA, which makes no assumption about the relationship between dependent and independent variables, and thus is unaffected by nonlinearity.

Thus, when using multiple treatment levels, if the relationship between the dependent and independent variables appears to be nonlinear, essentially 3 options exist for analyzing the data: (1) fit a linear equation to the data anyway, despite the expected bias and decrease in precision (not recommended); (2) attempt to determine the nonlinear equation that best describes the relationship between the dependent and independent variables; or (3) analyze the data using ANOVA. However, in experiments with multiple treatment levels, ANOVA typically will have equal or less power and precision than linear and nonlinear regression techniques due to the greater number of degrees of freedom required by the statistical model. Furthermore, an ANOVA only indicates if the effect of manipulation is different among 2 or more treatment levels, rather than providing information about the relationship between the independent and dependent variables.

REPLICATION AND LACK OF FIT

As mentioned previously, for a fixed number of study units and given range of manipulation, a researcher can maximize precision in estimates of effect by replicating units at 1 treatment level (at the extremity of that range), rather than distributing study units across multiple treatment levels. Replication also is important in demonstrating that observed differences between treatment levels are attributable to the effect of treatment rather than random variation. Indeed, data cannot be analyzed using *t*-tests or ANOVA without some replication. Regression techniques do not require replication because the random variation (error) is estimated from the dependent value of

each sample and the expected value of that sample as estimated using the curve-fitting technique. Thus, multiple-treatment-level designs may be used even when the number of samples being analyzed is extremely small (e.g., 4 study areas). The benefit of replication in multiple-treatment-level designs is that replication allows for a lack-of-fit test (Draper and Smith 1998).

A lack-of-fit test examines whether the chosen regression equation adequately explains the relationship between the dependent and independent variables. Replication at particular levels of treatment or control provides information about random variation in the dependent variable due to experimental error or natural variation. The total error in any regression model is the combination of this pure error and the lack of fit of the regression equation to the data. A lack-of-fit test simply examines the ratio of the amount of error due to lack of fit to the pure error (i.e., the ratio is the mean squared error [MSE] due to lack of fit compared to the MSE due to pure error). Under the null hypothesis that the regression equation adequately describes the relationship between the 2 variables, the expected ratio follows an *F* distribution with degrees of freedom described by equations 3 and 4:

$$df_1 = n - (k + 1) - \sum_i (n_i - 1) \quad (3)$$

$$df_2 = \sum_i (n_i - 1) \quad (4)$$

where *k* is the number of parameters in the chosen regression equation (not including the *Y*-intercept), and *n* is the total number of samples or the number of samples at treatment level *i*. Note that to perform a lack-of-fit test, not only must the number of independent values for which data have been collected be greater than the number of parameters in the regression equation, but some replication must exist within treatment levels.

CONCEPTUAL FRAMEWORK

Given the dual importance of replication within treatment levels and the number of treatment levels, and the impact sample distribution has on precision and inferential ability, how do researchers designing experiments decide to distribute samples between replicates and levels of treatment? Here, we propose a conceptual framework that may assist experimenters with this decision (Fig. 3).

If the independent variable is not continuous (e.g., vaccinated or not vaccinated) or manipu-

lating the independent variable at multiple levels is logistically difficult (e.g., predator or parasite-removal experiments), then an experimental design with only 1 level of treatment is best. Similarly, experiments involving only 1 level of treatment may be best if the researcher only desires to determine if an effect of manipulation exists. Researchers may only want to demonstrate the existence of an effect when they need preliminary results, when they are testing new theories such that the null hypothesis (no effect) is believed to be true, or when the level of manipulation is well established (e.g., food supplementation experiments typically provide ad libitum food [see Boutin 1990]). In these situations, experimental designs involving only 1 level of treatment may be desirable, especially if they are followed by further experiments using multiple treatment levels after the detection of an effect. Researchers are reminded, however, that the statistical detection of an effect is highly dependent on sample size and arbitrary *P*-values, and subject to problems associated with assumptions about the null hypothesis (Cherry 1998, Johnson 1999). Thus, researchers are encouraged to report instead the magnitude of the effect of manipulation along with measures of precision (Cherry 1998, Johnson 1999).

If the researcher wishes to determine the relationship between independent and dependent

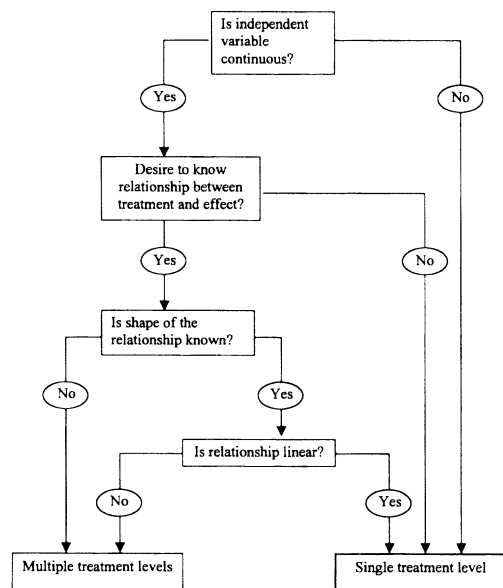


Fig. 3. Conceptual framework researchers might employ to help them choose the number of treatment levels to use in an experimental design.

variables, then the number of treatment levels to use should be dependent upon any a priori knowledge of that relationship. If the relationship between the independent and dependent variables is linear and the researcher has knowledge of this fact a priori, the researcher need only determine the equation of the line. Under this scenario, the researcher would be best served by replicating at only 2 independent values (i.e., 1 level of treatment and control) at the extremes in the natural range (Draper and Smith 1998, Petraitis 1998). Such a design maximizes the precision in the estimate of the effect. Similarly, if the relationship between independent and dependent variables is known a priori to be characterized by a particular nonlinear equation, then the researcher need only parameterize the equation. Thus, the researcher would be best served to replicate at k levels of treatment (including replicated controls) where k is the number of parameters in the nonlinear equation. Likewise, if the researcher is only interested in determining the effect of manipulation at a few important levels of treatment, the researcher could replicate at those levels of treatment and use ANOVA to analyze the data.

More often than not, a researcher does not have any a priori knowledge about the relationship between dependent and independent variables. In such cases, an experimental design involving multiple treatment levels should provide the greatest inferential ability. The researcher should determine the most complex model that may describe the relationship between the independent and dependent variables, and then set the number of treatment levels (including control) to 1 or 2 greater than the number of parameters in that model. This number of treatment levels allows the researcher to compare possible curve-fits using the Akaike Information Criterion (Anderson et al. 2000), Mallows' C_p statistic (Mallows 1973), or some other method for determining the regression equation that provides the best fit to the data. This number of treatment levels also allows the researcher to perform a lack-of-fit test on the selected model to reaffirm that it is the most appropriate descriptor of the relationship. If there is an expectation of linearity in the relationship, then the precision in the estimate of effect (slope) can be maximized by placing most replicates at the extremities in the range of manipulation. Note, however, that precision in nonlinear equations may not be similarly increased by such actions.

CONCLUSIONS

Peters (1983, 1991) asserts that the field of ecology has reached a crossroad, arguing that ecologists too often have produced study results that lack explanatory and predictive ability. This argument represents a clear mandate for a new experimental paradigm, which Peters refers to as predictive ecology, where the goal of research becomes the production of mathematical and/or conceptual constructs capable of testing, and therefore more adequately explaining ecological theory (Peters 1983, 1991). The same principle should guide research related to wildlife ecology, management, and conservation.

Experimental designs with multiple treatment levels fit nicely within the predictive paradigm advocated by Peters. Using parameter estimates furnished by such a continuous approach, biologists should be better able to address theoretical research hypotheses pertaining to natural systems, and thus begin to address the issue of unreliable knowledge (see Romesburg 1981). Similarly, the range of information provided by this type of experimental design should better enable professionals to set conservation and management objectives at appropriate levels.

ACKNOWLEDGMENTS

Funding during the writing of this paper was provided by the Idaho Department of Fish and Game and the U.S. Forest Service. We thank B. Dennis, K. Steinhorst, C. Williams, and 2 anonymous reviewers for statistical advice and valuable comments on earlier drafts of the manuscript.

LITERATURE CITED

- 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.
- BOUTIN, S. 1990. Food supplementation experiments with terrestrial vertebrates: patterns, problems, and the future. *Canadian Journal of Zoology* 68:203-220.
- CHERRY, S. 1998. Statistical tests in publications of The Wildlife Society. *Wildlife Society Bulletin* 26:947-953.
- DRAPER, N. R., AND H. SMITH. 1998. *Applied regression analysis*. John Wiley & Sons, New York, USA.
- EBERHARDT, L. L., AND J. M. THOMAS. 1991. *Designing environmental field studies*. Ecological Monographs 61:53-73.
- GOLDBERG, D. E., AND S. M. SCHEINER. 1993. ANOVA and ANCOVA: field competition experiments. Pages 69-93 in S. M. Scheiner and J. Gurevitch, editors. *Design and analysis of ecological experiments*. Chapman & Hall, New York, USA.
- JOHNSON, D. H. 1999. The insignificance of statistical significance testing. *Journal of Wildlife Management* 63:763-772.

- KAMIL, A. C. 1988. Experimental design in ornithology. Pages 313–346 in R. F. Johnston, editor. Current ornithology. Volume 5. Plenum Press, New York, USA.
- MALLOWS, C. L. 1973. Some comments on C_p . *Technometrics* 15:661–675.
- MURPHY, D. D., AND B. D. NOON. 1991. Coping with uncertainty in wildlife biology. *Journal of Wildlife Management* 55:773–782.
- NUDDS, T. D., AND M. L. MORRISON. 1991. Ten years after “reliable knowledge:” are we gaining? *Journal of Wildlife Management* 55:757–760.
- PETERS, R. H. 1983. The ecological implications of body size. Cambridge University Press, Cambridge, United Kingdom.
- . 1991. A critique for ecology. Cambridge University Press, Cambridge, United Kingdom.
- PETRAITIS, P. S. 1998. How can we compare the importance of ecological processes if we never ask, “Compared to what?” Pages 183–201 in W. Resetarits and J. Bernardo, editors. *Experimental ecology, issues and perspectives*. Oxford University Press, Oxford, United Kingdom.
- POTVIN, C. 1993. ANOVA: experiments in controlled environments. Pages 46–68 in S. M. Scheiner and J. Gurevitch, editors. *Design and analysis of ecological experiments*. Chapman & Hall, New York, USA.
- RATTI, J. T., AND E. O. GARTON. 1994. Research and experimental design. Pages 1–23 in T. A. Bookhout, editor. *Research and management techniques for wildlife and habitats*. Fifth edition. The Wildlife Society, Bethesda, Maryland, USA.
- ROMESBURG, H. C. 1981. Wildlife science: gaining reliable knowledge. *Journal of Wildlife Management* 45:293–313.
- . 1991. On improving the natural resources and environmental sciences. *Journal of Wildlife Management* 55:744–756.
- YOCOZ, N. G. 1991. Use, overuse, and misuse of significance tests in evolutionary biology and ecology. *Bulletin of the Ecological Society of America* 72:106–111.
- ZAR, J. H. 1996. *Biostatistical analysis*. Prentice-Hall, Upper Saddle River, New Jersey, USA.

Received 3 March 2001.

Accepted 26 October 2001.

Associate Editor: Morrison.