Journal of Wildlife Diseases, 45(3), 2009, pp. 700–712

MAKING RELIABLE DECISIONS IN THE STUDY OF WILDLIFE DISEASES: USING HYPOTHESIS TESTS, STATISTICAL POWER, AND OBSERVED EFFECTS

Chris O'Brien,[1,4] Charles van Riper III,[2] and Donald E. Myers[3]

[1] School of Natural Resources, University of Arizona, 1311 E. 4th Street, 125 Biological Sciences East, Tucson, Arizona 85721, USA
[2] US Geological Survey, Southwest Biological Science Center, Sonoran Desert Research Station, and School of Natural Resources, University of Arizona, 1311 E. 4th Street, 125 Biological Sciences East, Tucson, Arizona 85721, USA
[3] Department of Mathematics, University of Arizona, 1311 E. 4th Street, 125 Biological Sciences East, Tucson, Arizona 85721, USA
[4] Corresponding author (email: obrienc@email.arizona.edu)

ABSTRACT: The increasing importance of wildlife diseases in conservation efforts places an additional importance on research study design, data analysis, and interpretation. In this paper, we explore the design and analysis of wildlife disease data with regard to hypothesis testing, statistical power, sample sizes, the relative costs of type I versus type II errors, and effect size. To illustrate these ideas, we conducted a literature review of the *Journal of Wildlife Diseases* (JWD), ran computer simulations that estimate type II error rates for statistical techniques commonly used in JWD, and reanalyzed previously published data on disease prevalence. Many studies published in JWD used chi-squared analysis on prevalence data, but only 19% reported estimates of the observed effect size. Furthermore, 10% of studies had pooled sample sizes ≤40, and many had potentially high costs of type II relative to type I errors. Our computer simulations suggest that many articles published in JWD lack sufficient statistical power, and this, coupled with our findings that many studies often ignore high costs of type II errors, argues for increased attention to statistical power. Finally, our data reanalysis shows how the presentation of observed effect sizes could allow a better assessment of the biologic significance of findings reported in JWD. We conclude with some general guidelines to assist wildlife disease researchers in the design of future studies and the statistical analysis of their data.

Key words: Abundance, negative binomial, precautionary principle, prevalence, statistical analyses, statistical power, type I and type II error.

INTRODUCTION

Disease is increasingly recognized as an important, and perhaps crucial, element in the management and conservation of wildlife species (Tompkins and Wilson, 1998; Deem et al., 2001). In the rapidly changing world of wildlife disease, researchers are being called upon to measure the effects of disease on small, often endangered, wildlife populations. To assess patterns of disease in wildlife populations, scientists often quantify the rate or degree of disease or parasitic infection and analyze these data using classic techniques of statistical hypothesis testing.

Two common measures of wildlife disease reported in the literature are prevalence and abundance. Prevalence is the proportion of individuals in a sample that are infected with a disease (Bush et al., 1997), and in its raw form it constitutes a dichotomous response variable (infected/not infected). Given a random sample of hosts, prevalence can be representative of the disease status of a host population. Alternatively, abundance is a count of the number of parasites or disease units that are found in a single host (Bush et al., 1997), taking an ordinal value of zero or greater. Averaged across all individuals in a random sample, abundance estimates the mean number of parasites or disease units carried by a single animal within a population.

Ideally, random sampling can be used to estimate the prevalence or parasite abundance within animal populations, and statistical analyses can be used to assess how these measures of disease vary with different factors (e.g., sex, age, time of year, levels of human impact). The

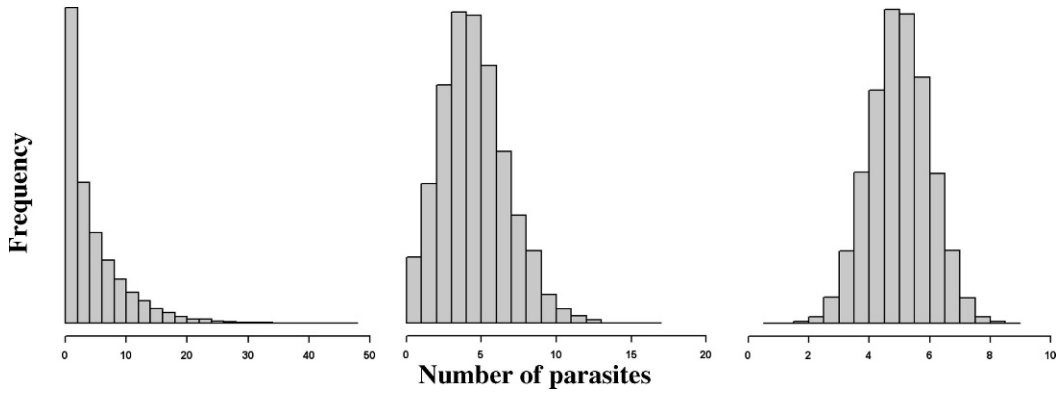

FIGURE 1. Examples of three distributions: the negative binomial (A), Poisson (B), and normal (C). For all three distributions, mean=5. For negative binomial, variance=30, k=1; for Poisson, variance=5; for normal, variance=1.0.

analyses of these data are complicated because prevalence and abundance measures violate assumptions of standard linear models such as analysis of variance (ANOVA) and linear regression. Prevalence is bounded by 0 and 1, requiring a transformation (the arc-sin and logit transformations are common), contingency table analysis (e.g., chi-squared or log-linear models), logistic regression, or the use of nonparametric methods. Parasite abundance is usually strongly positively skewed (Fig. 1A), and it can be described by a negative binomial that incorporates an overdispersion term (k) that accounts for the degree to which the variance exceeds the mean (Crofton, 1971). A Poisson model, in which the variance is equal to the mean, has also been used to model parasite abundance (Wilson et al., 1996; Fig. 1B). In a review of previously published studies, 268 of 269 scientific papers reported that the variance in parasite abundance exceeded the mean (Shaw and Dobson, 1995), suggesting a widespread pattern of parasite aggregation in hosts. Because of this pattern, the logarithmic transformation has historically been employed to normalize abundance data before the application of linear models. Alternatively, generalized linear models (GLM) allow the user to explicitly model negative-binomial abundance data, offering a more robust alternative to traditional linear model analysis (Wilson and Grenfell, 1997).

Type I and type II errors

Because the material presented herein will make repeated reference to statistical concepts of error, we first review definitions of type I and type II error. The probability of a type I error (α) is the probability of rejecting the null hypothesis when the null hypothesis is true. This error type has traditionally been of primary interest to biologists. An accepted benchmark used for determining statistical significance is $\alpha<0.05$, a standard convention, which, while popular, is somewhat arbitrary (Johnson, 1999).

In contrast, a type II error (β) is the probability of accepting the null hypothesis when the null hypothesis is false, and statistical power ($1-\beta$) is the probability of correctly rejecting the null hypothesis. For example, if anthropogenic effects truly increase the prevalence of disease in a species, but the study fails to detect that effect, a type II error has been committed. A standard convention is $\beta\leq0.20$; however, there are practical arguments for minimizing the probability of β well below 0.2 (Di Stefano, 2003; see following).

Statistical theory dictates an inverse

relationship between type I and type II error; a decrease in acceptable levels of one error type increases the probability of making an error of the other type. Biologists have traditionally sought to minimize type I errors at the expense of type II errors. However, in many conservation applications, the consequences of a type II error may actually outweigh those of a type I error (Dayton, 1998)—this concept is known as the precautionary principle (Peterman and M'Gonigle, 1992; Kriebel et al., 2001). According to the precautionary principle, type II errors should be minimized if the cost of failing to find an effect is high, thus risking continued harm. This has implications when studying an organism that is locally rare or threatened, and we argue that consideration of the relative costs of type I and type II errors is important when planning a study of wildlife disease. Therefore, when planning studies or conducting statistical inference, it is important to consider both type I and type II errors (Cohen, 1977), and the appropriate error to minimize should depend upon the situation.

There is growing concern about the reliance on hypothesis testing in the biologic sciences (e.g., Johnson, 1999), and many alternatives have been proposed (e.g., Fidler et al., 2006). Despite this ongoing debate, science still overwhelmingly embraces statistical hypothesis testing (for a specific example, see Fidler et al., 2006). While not covered here, methods such as information theoretic approaches (Burnham and Anderson, 2002), Bayesian statistics (Johnson, 1999), and equivalence testing (Hoenig and Heisey, 2001) are three proposed alternatives to traditional hypothesis testing that may be preferable in the analysis of certain wildlife disease data. Researchers are urged to familiarize themselves with these methods and, when appropriate, apply them in their research.

In this article, we report the results of several exercises that demonstrate the ways in which statistical hypothesis testing can be used more effectively in wildlife disease research: 1) *Literature review*. To understand how data are currently analyzed and presented in the wildlife disease field, we reviewed the *Journal of Wildlife Diseases* (JWD) for common methods used in the analysis of prevalence and abundance data, sample sizes used in those studies, observed effect size, and the relative costs of type I and type II errors. 2) *Computer simulation*. To assess common statistical methods used in JWD, we conducted computer simulation to estimate power and the probability of type II errors associated with a variety of common techniques used on prevalence and abundance data. 3) *Data reanalysis*. Finally, to demonstrate the ways in which observed effect size and associated confidence intervals are important for understanding biologic significance, we analyzed previously published data on the prevalence of blood parasites in different populations of birds (Super and van Riper, 1995).

MATERIALS AND METHODS

Literature review

We reviewed five years (2000–04) of papers published in JWD that tested hypotheses about differences in either the prevalence or abundance of wildlife diseases and parasites. We included studies that measured the prevalence and/or the abundance of macroparasites (e.g., helminths, ectoparasites) determined from visual counts, or microparasites determined through seroprevalence tests or counts of parasites per unit volume (e.g., viruses, bacteria, blood hematazoa). For each study identified in our review, we tabulated the types of analyses conducted, whether the magnitude of the observed effects was estimated, and the total number of sampled animals (pooled sample size, n). We also scored each study for the relative costs of type I to type II errors, according to the philosophy of the precautionary principle (Peterman and M'Gonigle, 1992; Kriebel et al., 2001), where we subjectively perceived that there was a *possibility* that costs of type II errors could exceed costs of type I errors (i.e., if a type II error could result in the authors failing to

detect a true harm to a species). We chose to include this assessment in order to raise awareness about the possibility of the high costs of type II errors in the study of wildlife diseases.

Computer simulation

Because computer simulation provides a robust method for estimating type II error rates, especially for non-normally distributed data (Crawley, 2002), we used simulations to estimate β associated with different statistical techniques of both prevalence and abundance data. To accomplish this, we selected equal random samples from two hypothetical populations with different mean prevalence or abundance and used different statistical methods to test the null hypothesis of no difference ($\alpha<0.05$) between the two populations. For prevalence data sets, we estimated type II error rates for the chi-squared test for independence (Ramsey and Shafer, 2002, section 19.3), Fisher's exact test (Ramsey and Shafer, 2002, section 19.4), log-linear regression (Ramsey and Shafer, 2002, chapter 22; Nelder, 2000), and logistic regression (Ramsey and Shafer, 2002, chapter 20). We conducted tests over a range of pooled samples sizes ($n=20$–1,000) and observed differences between the two simulated populations, (range=0.01–0.4). Because the power of hypothesis tests on prevalence data depends on the location of the proportion relative to 0.5 (Cohen, 1977), we completed two sets of comparisons: one in which the base proportion was 0.5, and another in which the base proportion was 0.1.

For abundance data, we determined type II error rates for *t*-tests (Ramsey and Shafer, 2002, chapter 2), *t*-tests after log-transformation (Ramsey and Shafer, 2002, chapter 3), nonparametric Wilcoxin tests (Ramsey and Shafer, 2002, chapter 4), and negative binomial regression (a GLM with negative binomial errors and a log link function; Venables and Ripley 2002, section 7.4). These analyses used the same sample sizes as above, and differences in mean abundance ranged from 1 to 500. Because type II error rates vary with k (the aggregation parameter of the negative binomial), we conducted comparisons for three different values of k (0.3, 1, 1.5), which span the range observed in most studies of parasites in wildlife hosts (Shaw and Dobson, 1995; Shaw et al., 1998). In each simulation, we generated a pair of random samples with different mean prevalence or abundance and conducted a hypothesis test, repeating this 10,000 times. We calculated the type II error rate as the proportion of the 10,000 tests in which a type II error was made ($P\geq0.05$).

To compare type II error rates for testing hypotheses of differences in prevalence versus differences in mean abundance, we took random samples from two negative binomial distributions with means of 1 and 2, respectively, for three values of k (0.3, 1.0, 1.5). We then conducted a statistical test for difference in mean abundance between the two samples, converted abundance values to prevalence, and conducted a statistical test for difference in prevalence between the two samples using log-linear regression. These simulations were done for a range of sample sizes with a balanced design ($n=20$–1,000), and the probability of a type II error was computed as above.

Data reanalysis

We reanalyzed data from a previously published JWD paper to demonstrate the advantages of estimating observed effect size. Super and van Riper (1995) used chi-squared contingency tables to test if the prevalence of avian hematazoan parasites was different on coastal islands than on the California mainland, and between resident and migratory bird communities. In contrast to the methods of Super and van Riper (1995), we used log-linear models with "infected/not infected" as a response (Nelder, 2000), and "island/mainland" and "migratory/resident" as independent variables, and we tested the null hypothesis of independence between the response and the independent variables based on a chi-squared distribution for one degree of freedom (Crawley, 2002). We estimated the effect size using the odds ratio of the two factor terms (Nelder, 2000) and determined significance of the effect with a Wald's chi-squared test.

Our method, in addition to being more statistically powerful than chi-squared tests, was chosen primarily because it provides a parameter estimate that allows the wildlife disease researcher to infer the magnitude of the observed effect that factors have upon the prevalence of disease, but we could have used other statistical methods that also estimate observed effects. For example, with 2×2 tables, a log-linear model is identical to logistic regression (Nelder, 2000). We could have also estimated the odds ratio from the 2×2 table and computed confidence intervals using the binomial distribution, although this method tests a hypothesis of homogeneity rather than independence, and sampling schemes can dictate the appropriate analysis (Ramsey and Shafer, 2002, section 19.2).

TABLE 1. Summary of analysis type and statistical techniques of 70 papers that conducted hypothesis tests about difference in mean prevalence or abundance in the *Journal of Wildlife Diseases* from 2000 to 2004.

	Count	% of total
Analysis type		
Prevalence	67	96
Abundance	17	24
Both	14	20
Total	70	
Prevalence		
Two-sample *t*-test/ANOVA	4	5.7
G-test/log-linear	5	7.1
Nonparametric rank test[a]	2	2.8
Chi-squared/Fisher's exact test	44	63
Logistic regression	15	21
Total	70[c]	
Abundance		
Two-sample *t*-test/ANOVA	4	24
GLM–negative binomial	1	5.9
Nonlinear regression	1	5.9
Nonparametric rank test[b]	7	41
None	4	24
Total	17	

[a] Kruskal-Wallis.

[b] Kruskal-Wallis, Wilcoxin-Mann-Whitney.

[c] This number is greater than total number of studies that analyzed prevalence because some studies used more than one methodology to analyze prevalence data.

All simulations and statistical analyses were conducted with the R package for Statistical Computing (R Development Core Team, 2007), and we used the *rnegbin* and *glm.nb* functions from the MASS library to take random samples and test hypotheses regarding the negative binomial distribution (Venables and Ripley, 2002).

RESULTS

Literature review

From 2000–04, 591 articles were published in the *Journal of Wildlife Diseases* (JWD). Of these, 70 papers tested hypotheses regarding differences in mean abundance or prevalence. The majority of studies (96%) reported prevalence, while 20% reported both prevalence and abundance (Table 1). Differences in mean prevalence were most commonly tested using chi-squared contingency tables (63%),

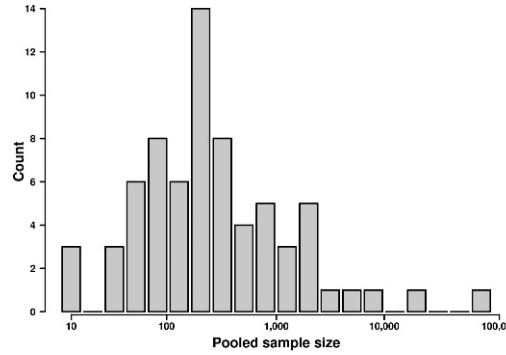

FIGURE 2. Histogram of sample sizes from 70 studies from the *Journal of Wildlife Diseases*. Values represent pooled sample sizes, not those for each grouping variable. We adopted this convention because designs were rarely balanced, and the number of design factors varied. *X*-axis is plotted on a logarithmic scale.

while nonparametric tests (Kruskal-Wallis, Wilcoxin-Mann-Whitney) were the most common method used (41%) for testing differences in mean abundance (Table 1).

The number of factors and factor levels in these 70 studies varied widely, and designs were rarely balanced; pooled sample sizes in the studies ranged from 12 to 63,451. The distribution of sample sizes was strongly right skewed, and 28% of studies had pooled sample sizes (n) of <100; 10% of studies had $n \leq 40$ (Fig. 2). The median sample size was 216.5. Authors of JWD articles did not always provide estimates of the magnitude of the observed effect in their studies, even when their methods generated these results. Only 19% of studies reported estimates of observed effect sizes from linear models (or GLM) or odds ratios. We also found that the potential cost of type II errors exceeded the cost of type I errors in 30% of the studies, suggesting that added attention should be given to the power of statistical tests and in balancing the probability of type I and type II errors relative to their potential costs.

Computer simulation

Our simulations of prevalence data generated high probabilities of type II

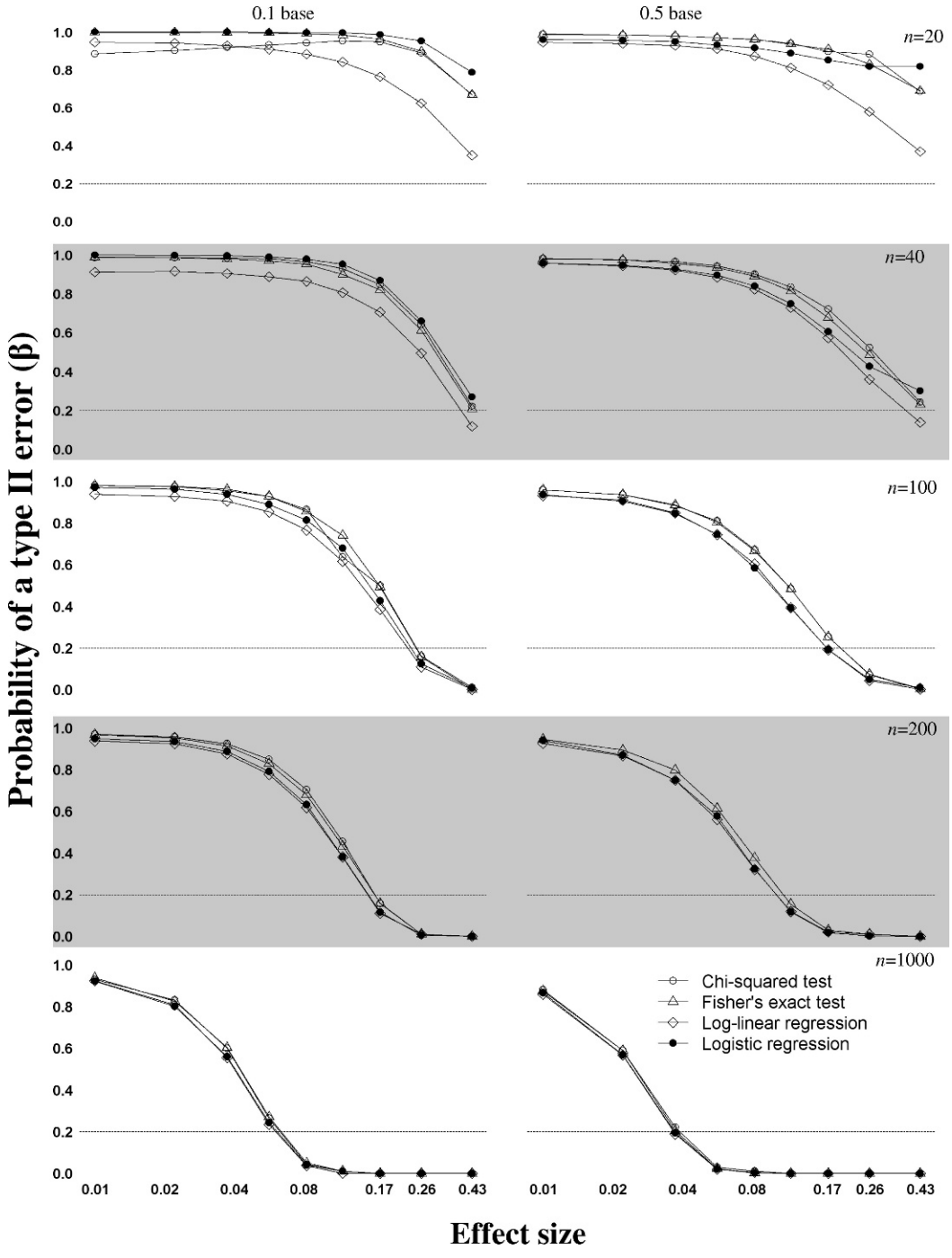

FIGURE 3. Probability of type II errors (y-axis) plotted for increasing sample sizes (vertically) and increasing effect sizes (x-axis) for two different base proportions for five different statistical methods. The effect size represents the difference in mean prevalence between the two populations, and it is plotted on a logarithmic scale. Horizontal lines indicate $\beta=0.20$, the generally accepted upper limit of beta. Sample sizes given are pooled n, for a balanced design comparing two groups (i.e., $n=200$ corresponds to a test comparing two samples, each of size 100).

errors, especially for small sample sizes and small effect sizes (Fig. 3). For small sample sizes ($n=20$), log-linear regression produced the lowest type II error rates. The probability of type II errors also decreased when comparing two proportions that were both closer to 0.5 (Fig. 3; right) than when comparing two proportions that were far from 0.5 (Fig. 3; left). Large sample sizes are important when comparing groups using prevalence; $n<200$ produced type II errors >0.20, except when effect size was ≥ 0.17. When prevalence values were close to 0.5, the type II error was ≤ 0.20 for the raw effect size ≥ 0.13.

Error rates for abundance data also produced high type II error rates, especially for small sample sizes, small effect sizes, and small values of the aggregation parameter k (Fig. 4). Type II errors decreased as the aggregation parameter (k), sample size, and effect size increased (Fig. 4). Highly aggregated samples ($k=0.3$) produced type II errors rates >0.2, except when pooled sample size was $\geq 1{,}000$. For more moderate values of k (1.0 and 1.5), $n=200$ resulted in low values of β. Effect size had less of an impact on type II error rates in the analysis of abundance than in the analysis of prevalence (Figs. 3, 4) and was more pronounced for $k>0.3$ with abundance data. A GLM with negative binomial errors produced the lowest type II error rates with pooled sample size <100. Our simulation of statistical power for prevalence and abundance from the same data showed that analysis of abundance is always more powerful than analysis of prevalence (Fig. 5), at least when $n<1{,}000$.

Data reanalysis

Super and van Riper (1995) used chi-squared tests of independence to reject the null hypotheses of no difference for hematozoan prevalence between passerine birds found in island versus continental coastal scrub communities, and no significant differences in hematozoan prevalence between resident breeding versus migratory nonbreeding birds in California coastal scrub communities.

Our analysis produced the same conclusion reached by these authors. However, we were also able to estimate the effect of geographic location and migratory status on hematozoan parasite prevalence, something that could not be accomplished with chi-squared analyses. Like Super and van Riper (1995), we found that birds on the mainland site of Palomarin, California, were more likely to be infected than were birds from San Miguel Island ($P<0.001$, deviance=126.2, df=1). However, we were able to estimate that the odds of infection with hematozoan parasites were 9.9 (95% confidence interval [C.I.]=6.1–17.1) times greater at Palomarin than at San Miguel island ($P<0.001$, $z=8.754$, from a Wald's test). When we restricted our analysis to resident breeding species only, as did Super and van Riper (1995), we found the same effect ($P<0.001$, deviance=193.7, df=1), but additionally we were able to determine that for breeding birds only, the odds of infection at Palomarin were 57 (95% C.I.=23.3–184.7) times greater than on San Miguel Island ($P<0.001$, $z=8.754$).

We also compared the hematozoan prevalence for resident versus migratory birds at the two different study sites. Like Super and van Riper (1995), we found that the odds of infection for migratory birds varied by migration status at the island site ($P<0.001$, deviance=24.11, df=1); in addition, the odds of infection for migrant birds were 13.3 (95% C.I.=4.6–48.4) times greater than for resident species ($P<0.001$, $z=4.42$). On the mainland, we found that there was also a difference in prevalence between migratory and resident birds ($P<0.001$, deviance=50.66, df=1), but the pattern was reversed; the odds of infection for migrants was 3.4 (95% C.I.=2.4–4.8) times lower than the odds of infection for resident species. In summary, using more informative statistical methods than those used by Super and

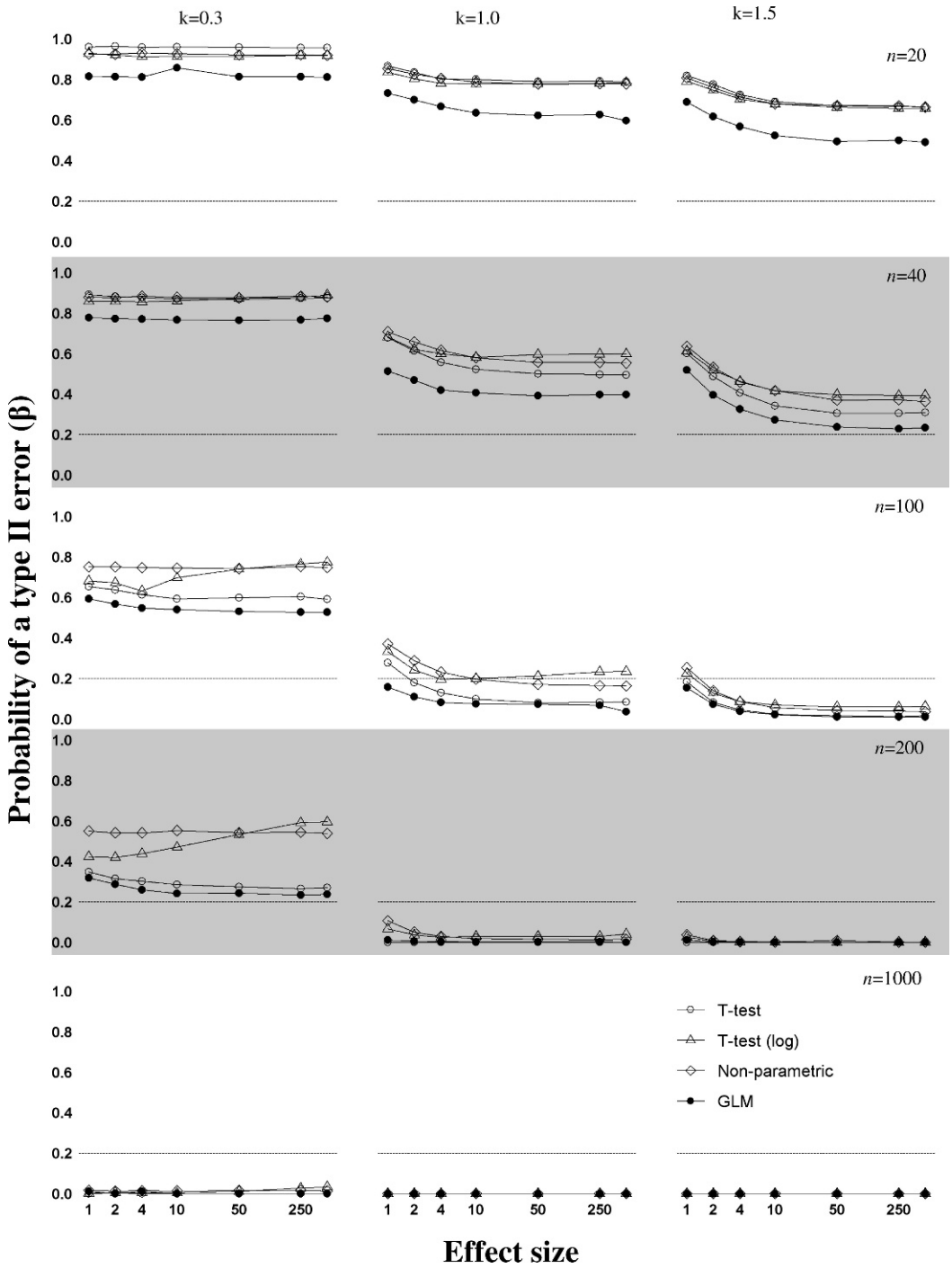

FIGURE 4. Probability of type II errors (*y*-axis) by effect size (*x*-axis), plotted for four different statistical techniques for a range of pooled sample sizes and aggregation parameters (*k*). Horizontal lines indicate $\beta=0.20$, the generally accepted upper limit. The *x*-axis (plotted on a logarithmic scale) represents the difference in mean abundance (effect size) between two populations. Sample sizes given are pooled *n*, for a balanced design, comparing two groups.

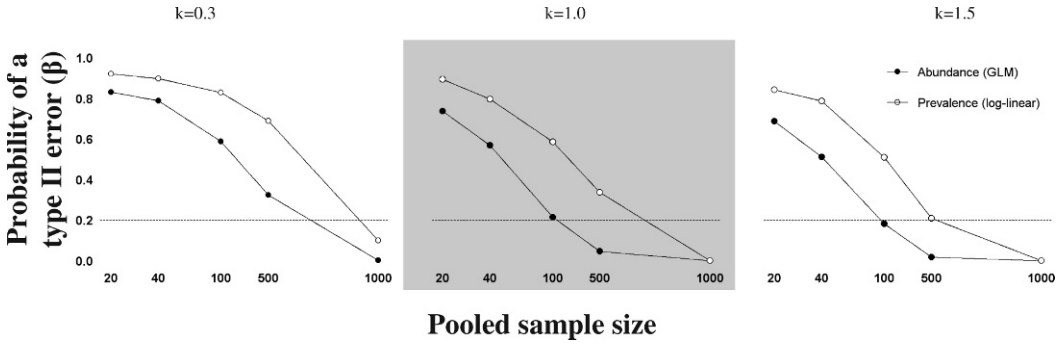

FIGURE 5. Probability of type II errors (y-axis) for increasing sample sizes (x-axis) for the analysis of abundance data using a GLM and of prevalence data using log-linear regression. For all three simulations, samples were randomly drawn from two populations with means=1 and 2, for k=0.3 (left), k=1.0 (middle), and k=1.5 (right). The corresponding prevalence values were 0.36 and 0.46 (left), 0.49 and 0.66 (middle), and 0.53 and 0.72 (right). Horizontal line indicates β=0.2. Sample sizes shown are pooled n, for a balanced design, comparing two groups.

van Riper (1995), we were able to estimate effect sizes, an important step toward understanding statistical results in a biologic context.

DISCUSSION

Literature review

Articles published in the *Journal of Wildlife Diseases* most commonly used Pearson's chi-squared test of independence for contingency tables when analyzing count data for disease prevalence. We argue that other techniques may be more useful because chi-squared tests are one of the least informative statistical tests due to the lack of an estimated parameter that allows the user to describe the degree of dependence between the variables of interest (Ramsey and Shafer, 2002). Chi-squared tests are also limited by their ability to only determine independence between sets of variables and homogeneity of proportions. In the study of wildlife disease, the researcher is often interested in measuring infection as a response that is a function of one or more explanatory variables. Alternative statistical methods, such as logistic regression and log-linear regression, allow the user to explicitly model the probability of infection given one or a number of explanatory variables, and associated parameter estimates can provide insight into the magnitude of those effects. Also, we have shown here that log-linear regression has greater statistical power than other techniques, and wildlife disease studies with small sample sizes should consider the use of this technique.

In our review of *Journal of Wildlife Diseases* papers, we found that some articles report data for studies in which pooled sample sizes were very small (e.g., 3 of the 70 reviewed articles had pooled sample sizes $\leq$15). At these sample sizes, for small to intermediate effect sizes, the probability of type II errors approaches 100%. Under these situations, statistical hypothesis testing becomes meaningless, especially if there is any cost to committing a type II error. We argue that when sample sizes are very small, it may be preferable to simply report descriptive statistics with associated confidence intervals, use methods more suited to such small sample sizes (for example, bootstrapping methods; Efron and Tibshirani, 1993), or wait to publish results until larger sample sizes become available. When wildlife disease researchers are dealing with critically endangered species, small sample sizes are a reality. We hope

the findings presented here demonstrate that hypothesis testing may not be the best way to understand such limited data sets.

The balance between statistical errors and the practical costs of type I and type II errors is often ignored in the scientific literature, and statistical methods commonly arbitrarily reduce the probability of a type I error at the expense of increasing type II errors (Di Stefano, 2003). Our literature review revealed that these concerns are also largely overlooked in wildlife disease research, and that in at least a portion of the studies that we investigated, the potential costs of making type II errors could equal or outweigh the cost of type I errors. When focusing on a species of conservation concern, wildlife disease researchers should make every effort in attempting to reduce the probability of making a type II error.

We recognize that our efforts to accurately assess the relative costs of type II and type I errors in the work of others may be imperfect; the individual researcher is eminently more suited to evaluate these relativities in her or his own work. However, we hope that by addressing this issue here, researchers will recognize the importance of considering the relative costs of type I and type II errors during the planning stage of future studies.

Computer simulation

Our results have produced some general guidelines for sample sizes in the analysis of prevalence and abundance data. Our computer simulation revealed that below $n=200$, analysis of prevalence data lacks statistical power, except for the largest effect sizes. For abundance data, there is also low power below $n=100$. These findings, combined with results from our literature review, suggest that at least some of the articles published in the *Journal of Wildlife Diseases* lack sufficient statistical power. This observation may be conservative for two reasons. First, our simulations used balanced sample sizes with one 2-level factor in the design. Studies reviewed in JWD usually had unbalanced designs that are inherently less powerful. Also, for a given pooled sample size, as the number of factors increases from one with more than two levels, statistical power also decreases. For these reasons, lack of statistical power may be more common than our study demonstrates. By utilizing the sample size guidelines presented in this paper, researchers can conduct their own prospective power analyses before a study design is implemented.

Our simulation results demonstrate that an analysis of parasite count data is always more powerful than an analysis of prevalence data, at least for $n<1{,}000$. When given the choice, abundance data should always be analyzed using appropriate methods, for example, when these counts are possible and feasible, such as in the study of macroparasites. Furthermore, independent analyses of abundance and prevalence data from a given data set are useful because they describe the disease dynamics of host wildlife populations in different ways.

We found that, for abundance data, negative binomial regression is more powerful than some alternative methods; this finding was also reported by Wilson et al. (1996). However, this technique does have shortcomings and should not be considered a panacea. For example, negative binomial regression may not be suited for models in which a single dispersion parameter is fit to multiple combinations of terms in a complex model. When models are simple and sample sizes are large, recommended alternatives include more complex nonlinear maximum likelihood methods and bootstrapping (Wilson and Grenfell, 1997; Newey et al., 2005). An analysis of dispersion also allows the user to account for variation in the dispersion parameter between combinations of model terms (Shaw et al., 1998).

Data reanalysis

Our analysis of the contingency tables in Super and van Riper (1995) came to the same general conclusion made in that paper. The goal of our data reanalysis was

not to find fault with the authors of the original paper, but to show that alternative methods could provide further insight into their research findings. We reanalyzed their data to make the point that chi-squared tests may not always be the most informative statistical tool, and that estimations of the observed effect size should be presented when possible. We argue that our use of an alternative method allows for a more informative exploration of their data. Instead of simply answering the question "Does the prevalence of blood hematazoa depend upon geographic location," we feel that our additional analyses addressed a more complex and perhaps more biologically meaningful question: "To what degree does the prevalence of blood hematazoa depend upon geographic location?" If, instead of geographic effects, we were interested in the role of an anthropogenic effect on the prevalence of disease in an endangered species, we believe that it would be important to know not only if an effect exists, but also how large that effect is. This could be done simply by estimating the difference between an anthropogenic treatment and control, and computing a confidence interval of the difference.

An understanding of the biologic importance of a statistically significant finding is important, as is the interpretation of findings that fail to reject the null hypothesis. The use of parameter estimates and associated confidence intervals can help in both cases (Steidl et al., 1997, 2000). In the case of our data reanalysis, our estimates of the differences in the prevalence of disease between different areas allowed us to interpret the magnitude of the observed effect, which can then lead to a discussion of the biologic importance of this effect. Because wildlife disease workers collect data that contain biologic information, we should use our data to come to biologic, not simply statistical, conclusions (Steidl et al., 2000). Furthermore, when statistical inference fails to detect a difference, issues of statistical power come into play. Because assessing the power of a test retrospectively can be problematic (Gerard et al., 1998), confidence intervals should be used to guide inferences when researchers fail to reject their null hypotheses (Steidl et al., 1997; Gerard et al., 1998).

Finally, it should be noted that unreliable or biased numbers work just as well as reliable ones when conducting statistical hypothesis tests. Many newer methods are being developed to better increase the diagnostic reliability of estimation and analysis measures such as prevalence (e.g., Senar and Conroy, 2004; Heisey et al., 2006; Jennelle et al., 2007). While these methods have not been addressed here, wildlife disease researchers need to explore and, where appropriate, implement new and emerging statistical techniques as the study of wildlife diseases plays an ever-increasing role in the conservation of wildlife species.

In summary, we offer several suggestions regarding the analysis of wildlife disease data: 1) It is important to consider statistical power when designing and analyzing data in wildlife disease studies. If the costs of type II errors are potentially high, researchers should ensure that it is feasible to collect enough data to adequately answer the question at hand, and then use statistical tests that have the most power. In determining necessary sample sizes, the general guidelines provided here may be used, or prospective power analysis may be used to estimate the power of the proposed study. 2) If possible, data should be collected and analyzed that will allow analysis of parasite abundance. Analysis of abundance is not only more powerful for a given sample size than the analysis of prevalence, but it also allows one to describe disease dynamics in an alternative way. 3) When possible, statistical techniques should be used that provide parameter estimates of the effect size observed in the study. Parameter estimates should be reported along with confidence intervals, which will allow both researchers and readers to assess the biologic significance of reported findings.

Wildlife disease workers are faced with a wide array of statistical options for analyzing their data. Additionally, many of us strive to develop research programs that have relevance in a management setting. With this increasing complexity, there comes the added responsibility to use appropriate statistical techniques and to maximize information transfer between the scientist and the user of scientific information. We hope that our suggestions and comparison of statistical techniques in this article will provide food for thought in the design of future wildlife disease studies and in eventual data analysis and interpretation.

ACKNOWLEDGMENTS

Paul Super generously shared the data used in Super and van Riper (1995); that research was supported by Point Reyes Bird Observatory and Channel Islands National Park. Bob Steidl provided thoughtful discussion, and A. Flesch, S. M. Samuel, C. Jennelle, and an anonymous reviewer provided comments to the manuscript, as did F. La Sorte, who also provided technical suggestions for the analysis. This study was supported financially by the US Geological Survey, and it was completed in partial fulfillment of a Ph.D. in Wildlife and Fisheries Science at the University of Arizona by C.O.

LITERATURE CITED

Burnham, K. P., and D. R. Anderson. 2002. Model selection and multimodel inference: A practical information-theoretic approach. 2nd Edition. Springer-Verlag, New York, New York, 488 pp.

Bush, A. O., K. D. Lafferty, J. M. Lotz, and A. W. Shostak. 1997. Parasitology meets ecology on its own terms: Margolis et al. revisited. Journal of Parasitology 83: 575–583.

Cohen, J. 1977. Statistical power analysis for the behavioral sciences, revised edition. Academic Press, New York, 474 pp.

Crawley, M. J. 2002. Statistical computing: An introduction to data analysis using S-Plus. John Wiley and Sons Ltd., West Sussex, UK, 761 pp.

Crofton, H. D. 1971. A quantitative approach to parasitism. Parasitology 62: 179–193.

Dayton, P. K. 1998. Reversal of the burden of proof in fisheries management. Science 279: 821–822.

Deem, S. L., W. B. Karesh, and W. Weisman. 2001. Putting theory into practice: Wildlife health in conservation. Conservation Biology 15: 1224–1233.

Di Stefano, J. 2003. How much power is enough? Against the development of an arbitrary convention for statistical power calculations. Functional Ecology 17: 707–709.

Efron, B., and R. J. Tibshirani. 1993. An introduction to the bootstrap. Chapman & Hall, Cambridge, Massachusetts, 436 pp.

Fidler, F., M. A. Burgman, G. Cumming, R. Buttrose, and N. Thomason. 2006. Impact of criticism of null-hypothesis significance testing on statistical reporting practices in conservation biology. Conservation Biology 20: 1539–1544.

Gerard, P. D., D. R. Smith, and G. Weerakkody. 1998. Limits of retrospective power analysis. Journal of Wildlife Management 62: 801–807.

Heisey, D. M., D. O. Joly, and F. Messier. 2006. The fitting of general force-of-infection models to wildlife disease prevalence data. Ecology 87: 2356–2365.

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

Jennelle, C. S., E. G. Cooch, M. J. Conroy, and J. C. Senar. 2007. State-dependant detection probabilities and disease prevalence. Ecological Applications 17: 154–167.

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

Kriebel, D., J. Tickner, P. Epstein, J. Lemons, R. Levins, E. L. Loechler, M. Quinn, R. Rudel, T. Schettler, and M. Stoto. 2001. The precautionary principle in environmental science. Environmental Health Perspectives 109: 871–875.

Nelder, J. A. 2000. The analysis of contingency tables with one factor as the response: Round two. The Statistician 49: 383–388.

Newey, S., D. J. Shaw, A. Kirby, P. Montieth, P. J. Hudson, and S. J. Thirgood. 2005. Prevalence, intensity and aggregation of intestinal parasites in mountain hares and their potential impact on population dynamics. International Journal for Parasitology 35: 367–373.

Peterman, R. M., and M. M'Gonigle. 1992. Statistical power analysis and the precautionary principle. Marine Pollution Bulletin 24: 231–234.

R Development Core Team. 2007. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org. Accessed February 2007.

Ramsey, F. L., and D. W. Shafer. 2002. The statistical sleuth: A course in methods of data analysis. Duxbury Press, Pacific Grove, California, 742 pp.

Senar, J. C., and M. J. Conroy. 2004. Multi-state analysis of the impacts of avian pox on a population of Serins (*Serins serinus*): The importance of estimating recapture rates. Animal Biodiversity and Conservation 27: 133–146.

Shaw, D. J., and A. P. Dobson. 1995. Patterns of

macroparasite abundance and aggregation in wildlife populations: A quantitative review. Parasitology 111: S111–S133.

———, B. T. GRENFELL, AND A. P. DOBSON. 1998. Patterns of macroparasite aggregation in wildlife host populations. Parasitology 117: 597–610.

STEIDL, R. J., J. P. HAYES, AND E. SCHAUBER. 1997. Statistical power analysis in wildlife research. Journal of Wildlife Management 61: 270–279.

———, S. DESTEFANO, AND W. J. MATTER. 2000. On increasing the quality, reliability, and rigor of wildlife science. Wildlife Society Bulletin 28: 518–521.

SUPER, P. E., AND C. VAN RIPER III. 1995. A comparison of avian hematozoan epizootology in two California coastal scrub communities. Journal of Wildlife Diseases 31: 447–461.

TOMPKINS, D. M., AND K. WILSON. 1998. Wildlife disease ecology: From theory to policy. Trends in Ecology and Evolution 13: 476–477.

VENABLES, W. N., AND B. D. RIPLEY. 2002. Modern applied statistics with S. 4th Edition. Springer, New York, New York, 495 pp.

WILSON, K., AND B. T. GRENFELL. 1997. Generalized linear modeling for parasitologists. Parasitology Today 13: 33–38.

———, ———, AND D. J. SHAW. 1996. Analysis of aggregated parasite distributions: A comparison of methods. Functional Ecology 10: 592–601.

Received for publication 26 February 2007.