

Generalized Linear Models and Point Count Data: Statistical Considerations for the Design and Analysis of Monitoring Studies

Nathaniel E. Seavy,^{1,2,3} Suhel Quader,^{1,4} John D. Alexander,² and C. John Ralph⁵

Abstract

The success of avian monitoring programs to effectively guide management decisions requires that studies be efficiently designed and data be properly analyzed. A complicating factor is that point count surveys often generate data with non-normal distributional properties. In this paper we review methods of dealing with deviations from normal assumptions, and we focus on the application of generalized linear models (GLMs). We also discuss problems associated with overdispersion (more variation than expected). In order to evaluate the statistical power of these models to detect differences in bird abundance, it is necessary for biologists to identify the effect size they believe is biologically significant in their system. We illustrate one solution to this challenge by discussing the design of a monitoring program intended to detect changes in bird abundance as a result of Western juniper (*Juniperus occidentalis*) reduction projects in central Oregon. We estimate biologically significant effect sizes by using GLMs to describe variation in bird abundance relative to natural variation in juniper cover. These analyses suggest that for species typically positively associated with juniper cover, a 60-80 percent decrease in abundance may be expected as a result of juniper reduction projects. With these estimates of expected effect size and preliminary data on bird abundance, we use computer simulations to investigate the power of GLMs. Our simulations demonstrate that when data are not overdispersed and sample sizes are relatively large, the statistical power of GLMs is approximated well by formulas that are currently available in the bird literature for other statistical techniques. When data are overdispersed, as may be the case with most point count data, power is reduced.

Key words: Generalized linear models, juniper removal, monitoring, overdispersion, point count, Poisson.

Introduction

Measuring changes in bird abundance over time and in response to habitat management is widely recognized as an important aspect of ecological monitoring (Greenwood et al. 1993). The use of long-term avian monitoring programs (e.g., the Breeding Bird Survey) to identify population trends is a powerful tool for bird conservation (Sauer and Droege 1990, Sauer and Link 2002). Short-term studies comparing bird abundance in treated and untreated areas are also important because they can identify changes in bird responses to specific management actions (Nichols 1999). However, the ability of avian monitoring programs to effectively guide management decisions requires that studies be designed to detect differences that are biologically meaningful.

The statistical ability to detect differences in abundance between two treatments (e.g., habitats or management practices) is described by statistical power. To evaluate power one must answer two questions. (1) What statistical models are appropriate for the distributional properties of the data? Probably the most widely used method for monitoring bird abundance is point count surveys (Ralph et al. 1995). However, the properties of data produced from point counts may violate assumptions of commonly used statistical techniques. Count data often show non-normal distributions, especially when abundances are low and there are many zeros in the data. One approach to analyzing these data is to use generalized linear models. With such models, one may specify non-normal distributions and results can be interpreted in a manner similar to the familiar analysis of variance framework. (2) What is a biologically meaningful effect size? That is, what difference in abundance can be considered biologically, not just statistically, significant? Once these two questions have been answered, power analyses can be carried out to identify the appropriate sampling effort needed for a rigorous study design (Nur et al. 1999, Foster 2001).

¹Department of Zoology, University of Florida, Gainesville, FL 32611-8525

²Klamath Bird Observatory, Box 758, Ashland, OR 97520

³Corresponding author: e-mail: nes@klamathbird.org

⁴Present address: Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, United Kingdom

⁵USDA Forest Service, Redwood Sciences Laboratory, Pacific Southwest Research Station, 1700, Bayview Drive, Arcata, CA 95521

In this paper we discuss some properties of count data with an emphasis on how they can be analyzed with generalized linear models. We then use these models to analyze point count data from central Oregon and evaluate variation in bird abundance as a function of natural variation in Western juniper (*Juniperus occidentalis*) cover. We use the results to define a biologically significant effect of juniper removal and conduct computer simulations to investigate the power of generalized linear models to detect changes in bird abundance resulting from reduction of juniper cover. Our objective is to illustrate how the design of monitoring projects based on point count data can be improved by the definition of *a priori* effect sizes, the application of generalized linear models, and attention to the distributional properties of the data.

Distributional Properties of Count Data

The fixed-radius point count method of measuring bird abundance generates count data (i.e., the number of birds at a station). The distribution of such data is bounded by zero because one cannot detect a negative number of birds. If stations are visited once, the number of individuals detected per station will always be either zero or a positive integer. When bird abundances are low, the frequency distribution of detections is likely to be highly non-normal (right-skewed, with a majority of observations at or near zero). If this is the case, then the use of statistical tests that assume normal distributions (e.g., t-tests, least-squares regressions) may be inappropriate. Perhaps the simplest approach to comparing patterns of abundance measured by point counts is to use tests that are more flexible about the distributional properties of the data. Such approaches include non-parametric tests (Sokal and Rohlf 1995, Zar 1999) or resampling tests that use permutations of the data to make statistical inferences (Manly 1991, Crowley 1992).

Another approach is to transform the data such that they meet assumptions of statistical methods based on the normal distribution. Such transformations must fulfill two objectives: they must standardize the shape of the distribution to the normal curve and decouple any relationship between the mean and variance. Often, square-root or log transformations are suggested for Poisson distributed count data, in which the variance increases with the mean (Sokal and Rohlf 1995, Kleinbaum et al. 1998, Zar 1999). When counts contain zeros, a constant must be added prior to log transformation, and is often suggested for square root transformation. Values of 0.5 or 3/8 for this constant are suggested (Sokal and Rohlf 1995, Zar 1999), but should be used with the recognition that the choice of constants may have implications for statistical estimation (Thomas and Martin 1996).

An alternative to nonparametric statistics or data transformations is the use of statistical techniques in which non-normal distributions can be specified. Generalized linear models (GLMs) are a flexible and widely used class of such models that can accommodate continuous, count, proportion, and presence/absence response variables (McCullagh and Nelder 1989, Agresti 1996, Crawley 1997, Dobson 2002). Many statistical packages (e.g., SAS, SPSS) implement these models for logistic, Poisson, and negative binomial regression. Although the application of GLMs to point count data is not new (Link and Sauer 1998, Brand and George 2001, Robinson et al. 2001), we review these models here to provide the context for our estimates of effect size and power.

Generalized Linear Models

The models with which biologists are most familiar are linear models of the form

$$y = b_0 + b_1x_1 \dots + b_jx_j + e$$

where the first part of the right-hand side of the equation ($b_0 + b_1x_1 \dots + b_jx_j$) specifies the expected value of y given $x_1 \dots x_j$. This is called the 'mean' or 'systematic' part of the equation and contains a linear combination of x variables. The random component of the model (denoted by e), describes the deviation of the observed y values from the expected, and is assumed to be drawn from a normal distribution with constant variance. Generalized linear models (GLMs) extend simple linear models by allowing transformations to linearity in the mean part of the model, as well as non-normal random components (Agresti 1996). If μ is the expected value of y , then we can model not μ itself, but some function $f(\mu)$ of the expected value

$$f(\mu) = b_0 + b_1x_1 \dots + b_jx_j$$

The function $f(\mu)$ is called the link function. For the simple linear model above, the link function is $f(\mu) = \mu$. An alternative for loglinear models is to use a log link function, $f(\mu) = \log(\mu)$. Our model now becomes

$$\log(\mu) = b_0 + b_1x_1 \dots + b_jx_j$$

What about the random component of such a model? For count data, the random component is often likely to follow a Poisson distribution (Zar 1999), a distribution of integers between zero and infinity. The best-fit model is found by maximum likelihood techniques rather than minimizing the sums-of-squares. For a GLM with Poisson error, the usual link function is

$\log(\mu)$ as above. Note that when the parameter estimates $b_0 \dots b_j$ are used to calculate a predicted value for y , the formula is

$$\mu = \exp(b_0 + b_1x_1 \dots + b_jx_j)$$

A key property of the Poisson distribution is that the variance equals the mean. This means that, for any given value of μ , the distribution of observed y should have a variance of μ . That is, $\text{Variance}(y) = \mu$. If, instead, we observe that $\text{Variance}(y) > \mu$, the data are overdispersed. Conversely, if $\text{Variance}(y) < \mu$, the data are underdispersed.

Overdispersion is often caused by differences among data not accounted for by the model. Suppose the number of individuals of a particular species at a count site depended on both the treatment (burned vs. unburned) and the topography (slope) of the site. Unaware of this, we simply model the dependence of the count on the treatment alone. The remaining variation among sites accounted for by slope is thus ignored, and leads to greater variance in the data than expected in our Poisson GLM. The major problem associated with overdispersion is an inflated risk of Type I error (see below). When faced with such a problem, one solution is to calculate an overdispersion parameter, and multiply it with the estimated standard errors - thus increasing the 95 percent confidence intervals around the estimated parameters, $b_0 - b_j$, describing the treatment effects (Agresti 1996). Alternatively, one may specify a distribution that can incorporate the extra variance that cannot be accommodated by a Poisson distribution. One of the most common distributions for this application is the negative binomial (Crawley 1997). The negative binomial distribution has an extra parameter (k) that describes dispersion. As k approaches 0, there is less overdispersion and the negative binomial distribution approaches a Poisson distribution.

Apart from affecting Type I error, the distributional properties of count data have important implications for understanding our statistical ability (i.e., power) to detect differences in bird abundance of two areas that are compared.

Statistical Power: A Brief Review

Although the topic of statistical power as it applies to wildlife ecology has been addressed by others (e.g., Steidl et al. 1997 and Nur et al. 1999), we briefly review the concept to provide the reader with definitions of terms that we use throughout this paper. The backbone of conventional (i.e., frequentist) hypothesis testing is the evaluation of null hypotheses. If the null is rejected, then we conclude, for example, that there is a true difference between groups. Conversely, while failure to reject the null means that there is insufficient

evidence of a true difference, it does not imply that the treatments are the same (Johnson 1999). Statistical tests based on frequentist theory provide a tool to estimate the probability, if the null hypothesis was true, of obtaining data as or more extreme than that observed (this probability is called the P-value). Such tests are concerned with two possible errors: Type I error and Type II error.

Type I error is the probability of rejecting the null when the null is true. This probability is usually represented by alpha (α), and for most tests the criteria for statistical significance is, by convention, $\alpha = 0.05$. Traditionally, there has been a strong emphasis on statistical hypothesis testing and the 0.05 critical value, but the arbitrary nature of critical values has been criticized (Johnson 1999, Anderson et al. 2000). Type II error represents the probability of failing to reject the null when the null is truly false, and is represented by beta (β). Beta is usually discussed in the context of power, defined as $1-\beta$, the probability of detecting a difference when one in fact exists. For most studies, power of >80 percent is considered acceptable (Nur et al. 1999), but like the 0.05 value for alpha, this threshold is arbitrary (Di Stefano 2003). Ideally, one should minimize both α and β , but because these are inversely related, a compromise must be reached. Alpha and beta are determined by distributional characteristics of the data, including the dispersion of the data, magnitude of difference we want to detect (i.e., effect size), and sample size.

A fundamental question that must be asked when calculating power concerns the magnitude of effect to be detected (Nur et al. 1999, Lenth 2001). In the context of point counts and management, an effect size would be the difference in bird abundance between managed and unmanaged habitats. If the hypothesized effect is large, then high power can be achieved with a relatively low sample size; alternatively, if the effect is small, then a larger sample size will be needed to achieve the same power. The difficulty comes in establishing *a priori* what effect size is important (Cohen 1988, Lenth 2001). Although a number of authors have addressed the issues of statistical power for detecting differences among treatments surveyed by point counts (Dawson 1981, Barker and Sauer 1995, Aigner et al. 1997, Nur et al. 1999), there has been little discussion of what should constitute a biologically significant change in abundance. Unfortunately, there is no simple rule of thumb that can be applied to all studies or all species. When population changes are of concern, modeling may be useful for generating estimates of change that would threaten the ability to meet management objectives. For example, population viability analyses may identify a threshold density below which the probability of extirpation increases. When

such a threshold is available, then a biologically significant effect could be defined as one that would drop the population density below this threshold. In other cases, the goal of monitoring may be to evaluate whether or not a management treatment has achieved the desired ecological conditions. In such cases a pilot study describing the response variable as a function of natural variation in the habitat component of management interest may provide valuable information on the expected effect of biological interest. In the sections below we have illustrated such an approach.

To design an effective monitoring study, then, one must first determine the magnitude of the effect we consider important, and then find the sample size necessary to detect such an effect size at a given level of statistical power.

GLMs and the Design of Monitoring Studies

To illustrate the application of generalized linear models to the design of bird monitoring, we present an analysis of data collected at point count stations in the Upper Klamath basin of Oregon. In this area, the Oregon Bureau of Land Management (BLM) is concerned with Western juniper encroachment into sagebrush habitat and has implemented juniper control through mechanical removal of juniper trees. The goal is to restore pre-encroachment conditions. In conjunction with BLM, the Klamath Bird Observatory is designing monitoring programs to evaluate ecological effects of juniper removal treatments based on changes in abundance of bird species that serve as indicators of desired habitat conditions. Here we illustrate the use of generalized linear models to (1) identify which species show patterns of abundance that are associated with natural variation in juniper cover, (2) generate an estimate of the expected effect of juniper removal based on this response, and (3) estimate the required sample size to detect this difference between treated and untreated stations.

Methods and Results

Estimating Expected Effect Size

Data Collection and Analysis

In May and June 2002, bird abundance was measured using standardized point count survey methodologies (Ralph et al. 1993) at 78 stations that had not received juniper reduction treatments. Five-minute counts were conducted between sunrise and 1000 PDT at each station; all landbirds seen or heard within 50 m of the station were recorded. Vegetation data were collected at each station using a relevé method (Ralph et al.

1993). These stations had no tree cover other than *Juniperus occidentalis*. Juniper cover was classified into two categories: high (>25 percent cover) and low (0-25 percent cover). Previous studies have shown changes in bird community composition occur when juniper cover is >30 percent (Holmes and Barton 2002), thus we considered this a biologically relevant classification. We restricted our analyses to the five most abundant bird species which had detections at >5 stations in both high and low juniper areas.

To evaluate the ability of juniper cover (our variable of biological interest) to explain variation in bird abundance, we fit GLMs with Poisson distributions and log links with juniper cover (high and low) as a categorical predictor variable. Such an analysis is often referred to as "Poisson regression" and is discussed at length in both GLM texts (McCullagh and Nelder 1989, Dobson 2002) and more general texts on regression (Kleinbaum et al. 1998). We used the Pearson chi-squared statistic to evaluate overdispersion (Agresti 1996). When we found evidence of overdispersion, we refit the models with a negative binomial distribution and log link. All statistical tests were conducted with SAS (PROC GENMOD).

GLM Results

Model results for the five species are presented in *table 1*. The first step in interpreting GLM results is to review the output for evidence of overdispersion. In the absence of overdispersion the Pearson chi-square statistic should be approximately equal to the residual degrees of freedom (number of observations minus number of model parameters; Agresti 1996). Statistical evidence of overdispersion can be evaluated by computing the probability of obtaining the observed Pearson chi-square statistic from a chi-square distribution with an expected mean equal to the residual degrees of freedom; small P-values provide evidence of poor model fit (Agresti 1996). For our data, most models using juniper cover as an explanatory variable were overdispersed; the only species for which the P-value of the Pearson chi-square statistic was >0.05 was the American Robin (*for scientific names see table 1*). Therefore, caution is warranted when interpreting the significance of parameter estimates.

Because counts of American Robins were not significantly overdispersed, the effect of habitat can be interpreted from the Poisson model. For this species, the habitat parameter was significantly different from zero ($P = 0.002$), evidence of a difference between the mean number of individuals detected at station of low and high juniper cover. For the remaining four species, we refit the models using a negative binomial distribution. Three fit the negative binomial model well (*table 1*). Only the Mountain Chickadee still showed a

Pearson statistic with a low P-value ($P = 0.055$) suggesting a lack of fit. The mean parameter estimates for the Poisson and negative binomial models are the same, but the confidence intervals are re-scaled to account for the additional variance. Using the negative binomial model, only one of these four species that all showed a significant effect of juniper cover with the Poisson model still showed a significant effect at the $\alpha = 0.05$ level (table 1). The only species that retained a significant effect of juniper cover was the Chipping Sparrow (table 1), although the estimates for Mountain Chickadees and Gray Flycatcher approached the 0.05 criterion (table 1).

The parameter estimates of the model can be used to calculate the expected (mean) number of birds per

station in high and low juniper cover (table 2). For example, the predictive equation for the number of American Robins detected per point is:

$$\mu = \exp(-2.30 + 1.61x)$$

where x is 0 when juniper cover is ≤ 25 percent and 1 when juniper cover is > 25 percent. Thus, there was an average of 0.10 individuals per station in areas with low juniper cover, and five times that number (0.5 individuals per station) at stations with high juniper cover.

Table 1– Generalized linear model (specifying Poisson or negative binomial distributions and log link) results for the five most commonly detected bird species at 78 stations surveyed in 2002 in the Lakeview District of the Oregon Bureau of Land Management. The variable “Habitat” is coded 0 for stations with low (0-25%) juniper cover and 1 for stations with high (>25%) juniper cover. Thus $\exp(\text{Intercept})$ is the mean abundance at low cover and $\exp(\text{Intercept} + \text{Habitat})$ is mean abundance at high cover.

Species	Parameter	Poisson			Negative binomial		
		Estimate	SE	P-value	Estimate	SE	P-value
Gray Flycatcher (<i>Empidonax wrightii</i>)	Intercept	-1.61	0.29	<0.001	-1.61	0.42	<0.001
	Habitat	1.12	0.42	0.008	1.12	0.61	0.094
	Dispersion (k)	na	na	na	3.76	2.13	
	Pearson chi-square	161.18		<0.001	83.09		0.270
	Df	76			76		
Mourning Dove (<i>Zenaida macroura</i>)	Intercept	-2.02	0.35	<0.001	-2.02	0.52	<0.001
	Habitat	1.20	0.50	0.016	1.20	0.94	0.200
	Dispersion (k)	na	na	na	8.67	5.28	
	Pearson chi-square	188.00		0.000	75.34		0.500
	DF	76			76.00		
Mountain Chickadee (<i>Poecile gamboli</i>)	Intercept	-2.30	0.41	<0.001	-2.30	0.69	<0.001
	Habitat	1.61	0.53	0.002	1.61	0.89	0.069
	Dispersion (k)	na	na	na	9.43	6.11	
	Pearson chi-square	215.00		0.000	96.71		0.055
	DF	76			76		
American Robin (<i>Turdus migratorius</i>)	Intercept	-2.30	0.41	<0.001	Not needed		
	Habitat	1.61	0.53	0.002			
	Dispersion (k)	na	na	na			
	Pearson chi-square	95.00		0.069			
	DF	76					
Chipping Sparrow (<i>Spizella passerina</i>)	Intercept	-2.15	0.38	<0.001	-2.15	0.44	<0.001
	Habitat	1.34	0.52	0.010	1.34	0.60	0.030
	Dispersion (k)	na	na	na	2.89	2.32	
	Pearson chi-square	102.64		0.023	66.65		0.770
	DF	76			76		

Table 2—Distributional properties and effect sizes used in the power analyses for the four species with a significant ($P < 0.10$) difference in abundance between areas of low and high juniper cover. Mean abundance was predicted from GLM results (table 1). Standard deviations were calculated from raw data in each cover class. Percent difference is calculated as: $[(\mu_{\text{high juniper}} - \mu_{\text{low juniper}}) / \mu_{\text{high juniper}}] \times 100$. Dispersion parameters were estimated with a generalized linear model specifying a log-link and negative binomial error distribution.

Species	Predicted mean abundance (SD)		Percent difference	Dispersion parameter (<i>k</i>)
	Low juniper	High juniper		
Gray Flycatcher	0.20 (0.66)	0.61 (1.10)	67%	3.76
Mountain Chickadee	0.10 (0.54)	0.50 (1.10)	80%	9.43
American Robin	0.10 (0.35)	0.50 (0.78)	80%	na
Chipping Sparrow	0.12 (0.37)	0.44 (0.92)	74%	2.89

Estimating Required Sample Size: Power Analysis

Given our estimates of the magnitude of the effect that juniper removal might have on bird populations, what sampling intensity must be used to detect whether these effects actually occur? Power analysis software is now widely available, but nearly all are based on statistical designs that assume normality (e.g., ANOVA, t-test, linear regression). We were not aware of a widely available software package that could be used to perform power analysis for Poisson regression. Thus, given the distributional properties of the pilot data, we used a simple Monte Carlo simulation to investigate how the statistical power to detect our expected effects varied with sample size.

Power Analysis Methods

For our power analysis, we assumed that the difference between bird abundance in areas of high and low juniper cover reflects the magnitude of change we would expect to see if juniper removal treatments across a similar range of cover classes were indeed having the desired ecological effects on bird abundance. We defined this effect size as:

$$\text{Effect size} = (\mu_{\text{high juniper}} - \mu_{\text{low juniper}}) / \mu_{\text{high juniper}}$$

This effect size can be expressed as a percent difference by multiplying it by 100. Calculated effect sizes for the four species with significant habitat parameters are presented in table 2. For example, American Robin abundance is predicted to be 80 percent lower in areas of low juniper cover compared to where juniper cover is high. Thus, we might consider this an estimate of the predicted effect on American Robins if management activities generate a similar change in juniper cover.

We compare three methods of calculating power. First, we used an analytical equation for calculating the power for detecting a difference between means of

point count data from two treatments presented by Dawson (1981):

$$\text{number of stations} > (b \times 20000) / (d^2 \times m)$$

where b is a value that corresponds to the desired power (3.84 for 50 percent, 6.15 for 70 percent, 7.84 for 80 percent, and 10.50 for 90 percent; Nur et al. 1999), and d is the percent difference between the groups, defined as:

$$d = 100 \times [(m_1 - m_2) / m]$$

where m_1 and m_2 are the means in each treatment category and m is the overall mean for both areas. This formula assumes that the variance equals the mean (as in a Poisson distribution), but generates the estimate of power based on Z scores.

We compared this estimate with two Monte Carlo simulations. The first assumed that bird detections were normally distributed and compared populations with a t-test. Although this approach ignores the distributional properties of count data, it offers an approximation of power calculations that would be available in most power analysis software that are based on statistical designs that assume normality. In these simulations, "control" data were generated by drawing values from a normal distribution with the estimated mean and standard deviation of bird detections at stations with high juniper cover (table 2). "Treatment" data were generated from a normal distribution with a mean and standard deviation that were reduced by the effect size generated by our earlier analysis (table 2; we assume the standard deviation scales linearly with the mean). To statistically test for differences between our simulated control and treatment data, we used a two-tailed t-test. We rejected the null hypothesis of no difference in means between high and low juniper areas when $P < 0.05$ for the habitat parameter.

Our second Monte Carlo procedure provided a more realistic simulation of the generation and analysis of count data. Data were generated by drawing from

either a Poisson distribution (American Robin) with a single parameter representing the mean, lambda (λ), or a negative binomial distribution (Chipping Sparrow, Mountain Chickadee, and Gray Flycatcher) with mean parameter (p) and a scale parameter (k) to simulate overdispersion. In both cases, the distributional parameters (λ or p and k) were estimated from the generalized linear model analysis (*table 1*) and are presented in *table 2*. To statistically test for differences between our simulated control and treatment data, we used a generalized linear model specifying a log link and Poisson or negative binomial distribution. Again, we rejected the null hypothesis of no difference in means between high and low juniper areas when $P < 0.05$ for the habitat parameter.

In both cases these simulations were repeated 100 times for each sample size (varied in increments of 1) across the range of necessary sample sizes suggested by the Dawson (1981) equation. Power was calculated as the percent of trials in which the null hypothesis was rejected. All simulations were performed in R (R Development Core Team 2003) and code is available upon request from the authors.

Power Analysis Results

For the American Robin, we investigated the sample size required to detect an 80 percent reduction in abundance from the estimated mean of 0.5 individuals per station at high juniper cover sites. The estimates of the Dawson equation and the approximation based on an assumption of normality provided similar results (*Fig. 1*). When the sample sizes were large and power >80 percent, our Poisson GLM simulations gave similar estimates to the other two methods. However, when sample sizes were small, the Dawson equation and the assumption of normality both overestimated power compared with our simulated trials with the Poisson GLM (*fig. 1*). When data were overdispersed, the Dawson equation consistently overestimated power at small sample sizes, but usually converged with the GLM estimations as sample size increased. In contrast, simulations based on an assumption of normality typically produced estimates of power that were similar to the GLM estimates when sample sizes were small, but consistently underestimated power at larger sample sizes (*fig. 1*).

We found that in order to achieve 60-80 percent power to detect a change in the bird species we examined one would require surveys of 30 to 50 stations each in control and treatment areas. Overdispersion reduced power. A comparison is provided by American Robins and Mountain Chickadees, which shared the same abundance in treated and untreated areas, but Mountain Chickadees were overdispersed (*table 2*). Accordingly,

the power to detect a difference for Mountain Chickadee was consistently lower than for American Robin (*fig. 1*). This difference was not accommodated by the Dawson equation (*fig. 1*).

Discussion

Fixed-radius point counts generate count data that, even after transformation, are often non-normally distributed. As a result, using statistical tests that assume normality may be inappropriate. Here, we have illustrated the application of generalized linear models to designing monitoring studies and analyzing point count data. These models are an effective alternative to non-parametric statistics because they generate estimates of means with Poisson or negative binomial distributions. These distributions can be used as the basis of simulations for modeling or power analysis. However, overdispersion is common and its statistical consequences are important to consider. If ignored within the context of Poisson regression, overdispersion inflates the risk of Type I error. Using Poisson regression results without considering overdispersion, we would have concluded that all six species showed a significant effect of juniper cover at the $\alpha = 0.05$ level (*table 1*). After correcting for overdispersion, only two showed an effect of juniper cover at the $\alpha = 0.05$ level.

Overdispersion will also affect power. Failing to account for overdispersion leads to artificially high estimates of power. As recognized in the original presentation (Dawson 1981) and reiterated more recently (Nur et al. 1999), the Dawson equation assumes that data are Poisson distributed. Our approach differs from previous estimates of power for treatment comparisons using point counts (e.g., Dawson 1981, Aigner et al. 1997) because we used computer simulations to generate data sets from Poisson or negative binomial distributions that more closely mimicked actual data generated by single-visit point counts. Using this approach, we have shown that the equation of Dawson (1981) tends to overestimate power relative to our simulated estimates of power for Poisson or negative binomial data analyzed with GLMs. This problem is most extreme when sample sizes are small; we conclude that the Dawson equation should not be used to estimate power of GLMs when the sample size will be small. The Dawson equation does not account for overdispersion, thus it will always overestimate power when data are overdispersed. Similarly, using power analyses that assume normal distributions may also produce erroneous results. One of the advantages of GLMs is that they provide a means of describing the dispersion of data and evaluating power with distributions that are appropriate for count data.

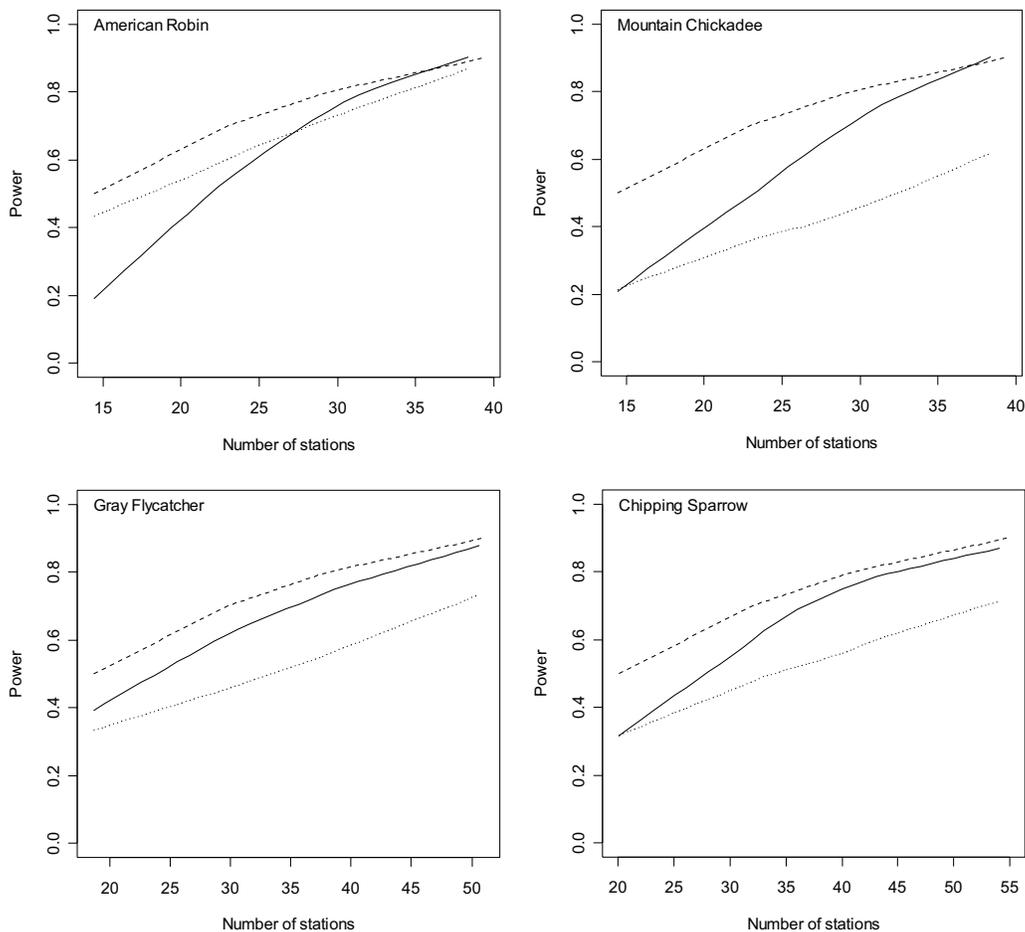


Figure 1—The effect of sample size (number of stations per treatment) on the estimated power to detect the difference in bird abundance between areas with and without juniper removal. Simulations increased sample size in increments of 1. Lines represent smoothed curves; simulations were run 100 times for each sample size. Dashed lines are estimates generated by the Dawson (1981) equation, dotted lines are based on simulations assuming normal distributions, and solid lines are based on simulations using Poisson or negative binomial distributions. Differences between treated and untreated areas and the normal, Poisson, and negative binomial parameters used to generate the simulations are presented in *table 2*.

When monitoring programs are designed with hypothesis testing as one of the goals, statistical power must be considered. Without consideration of power, monitoring programs risk inconclusive results that are unable to guide management decisions. To calculate statistical power for a given sample size, one must have a hypothesized or expected effect size in mind. We encourage ornithologists and managers to devote more effort to carefully considering the magnitude of effects that may have important biological implications. As one approach to estimating a biologically significant effect size, we have analyzed patterns of bird abundance relative to natural variation in juniper cover. This analysis suggests that the most common species in these habitats may exhibit relatively large changes in abundance (60-80 percent). Such information is valuable because it can be used to evaluate the statistical efficacy of monitoring designs to meet their

objectives. Indeed, this information should be used to design monitoring projects before they begin.

Even with the modest sample size of our pilot data, we detected significant variation in the abundance of American Robins and Chipping Sparrows associated with natural variation in juniper cover. This difference suggests that these species will serve as useful indicators of the ability of juniper removal management activities to achieve the desired habitat conditions in these ecosystems. By generating an understanding of the strength with which these species are associated with natural variation in habitat conditions, these studies provide valuable information that can be used to generate and test hypotheses about the effect of habitat management activities.

Avian monitoring has the potential to be a useful tool for land managers throughout the world. However, for

this potential to be realized in an efficient manner, monitoring programs that aim to compare avian abundance in different habitats or treatments must be designed in such a way that they achieve high power for statistical inference. When used and interpreted correctly, generalized linear models provide a powerful tool for detecting changes in avian abundance.

Acknowledgments

This project was funded by the Joint Fire Sciences Project and the Lakeview District of the Oregon Bureau of Land Management. Support and assistance for the project from P. Buetner was critical to its success. Field work was conducted by R. P. Chapman, D. P. Haupt, K. P. Gluckert, and D. Van den Broek. Comments from J. Bart, M. Chase, N. Nur, T. Okuyama and K. Sieving's lab group greatly improved the paper.

Literature Cited

- Agresti, A. 1996. **An introduction to categorical data analysis.** John Wiley and Sons, New York, NY.
- Aigner, P. A., W. M. Block, and M. L. Morrison. 1997. **Design recommendations for point counts of birds in California oak-pine woodlands: Power, sample size, and count stations versus visits.** In: Pillsbury, N., J. Verner, and W. D. Tietje [Tech. Coords.] 1997. Proceedings of a symposium on oak woodlands: ecology, management, and urban interface issues: 19-22 March 1996. Gen. Tech. Rep. PSW-GTR-160.
- Anderson, D. R., K. P. Burnham, and W. L. Thompson. 2000. **Null hypothesis testing: problems, prevalence, and an alternative.** *J. Wildlife Management* 64:912-923.
- Barker, R. J., and J. R. Sauer. 1995. **Statistical aspects of point count sampling.** In C. J. Ralph, J. R. Sauer, S. Droege, (Eds.). *Monitoring bird populations by point counts.* Gen. Tech. Rep. PSW-GTR-149. Albany, CA; 124-133.
- Brand, L. A., and T. L. George. 2001. **Response of passerine birds to forest edge in coast redwood forest fragments.** *Auk* 118:678-686.
- Cohen, J. 1988. **Statistical Power Analysis for the Behavioral Sciences.** Academic Press, New York.
- Crawley, M. J. 1997. **GLIM for ecologists.** Blackwell Science, Oxford, U. K.
- Crowley, P. H. 1992. **Resampling methods for computation-intensive data analysis in ecology and evolution.** *Annual Review of Ecology and Systematics* 23: 405-447.
- Dawson D. G. 1981. **Experimental design when counting birds.** In C.J. Ralph and J.M Scott (Eds.) *Estimating numbers of terrestrial birds.* *Studies in Avian Biology* 6:12-16.
- 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.
- Dobson, A. J. 2002. **An introduction to generalized linear models.** Chapman and Hall, London.
- Foster, J. 2001. **Statistical power in forest monitoring.** *Forest Ecology and Management* 151:211-222.
- Greenwood, J. J. D., S. R. Baillie, H. Q. P. Crick, J. H. Marchant, and W. J. Peach. 1993. **Integrated population monitoring: detecting the effects of diverse changes.** In: *Birds as monitors of environmental change.* R. W. Furness and J. J. D. Greenwood (Eds.). Chapman and Hall, London; 267-342.
- Holmes, A. L., and D. Barton. 2002. **Determinants of bird abundance and distribution in sagebrush habitats of southeastern Oregon and eastern Washington.** Final report to the Bureau of Land Management. Point Reyes Bird Observatory, Stinson Beach, CA.
- Johnson, D. H. 1999. **The insignificance of statistical significance testing.** *J. Wildlife Management* 63:763-772.
- Kleinbaum, D. G., L. L. Kupper, K. E. Muller, and A. Nizam. 1998. **Applied regression analysis and other multivariable methods.** Duxbury Press, Pacific Grove, CA.
- Lenth, R. V. 2001. **Some practical guidelines for effective sample size determination.** *The American Statistician* 55:187-193.
- Link W. A., and Sauer J. R. 1998. **Estimating population change from count data: Application to the North American Breeding Bird Survey.** *Ecological Applications* 8:258-268.
- Manly, B. F. J. 1991. **Randomization and Monte Carlo methods in biology.** Chapman and Hall, New York, NY.
- McCullagh, P., and J. A. Nelder. 1989. **Generalized Linear Models.** Chapman and Hall, New York, NY.
- Nichols, J. D. 1991. **Extensive monitoring programs viewed as long-term population studies: The case of North American waterfowl.** *Ibis* 133 suppl. 1:89-98.
- Nichols, J. D. 1999. **Monitoring is not enough: on the need for a model-based approach to migratory bird management.** In: Bonney, R. D. N. Pashley, R. J. Cooper, and L. Niles, (Eds.) 1999. *Strategies for Bird Conservation: The Partners in Flight Planning Process.* Cornell Lab of Ornithology. Unpublished report available at <http://birds.cornell.edu/pifcapemay>.
- Nur, N., G. R. Guepel, and S. Jones. 1999. **A statistical handbook to data analysis of standardized avian population monitoring programs.** U.S. Department of Interior, Fish and Wildlife Service, BTP-R6001-1999, Washington, D.C.
- R Development Core Team. 2003. **R: A language and environment for statistical computing.** R Foundation for Statistical Computing, Vienna, Austria. Available at <http://www.R-project.org>.

- Ralph, C. J., G. R. Guepel, P. Pyle, T. E. Martin, and D. F. DeSante. 1993. **Handbook of field methods for monitoring landbirds**. Gen. Tech. Rep. PSW-GTR-144.
- Robinson R. A., Wilson J. D., Crick H. Q. P. 2001. **The importance of arable habitat for farmland birds in grassland landscapes**. Journal of Applied Ecology 38:1059-1069.
- Sauer, J. R., and S. Droege, eds. 1990. **Survey designs and statistical methods for the estimation of avian population trends**. Biological Report 90(1). Washington, D.C.: Fish and Wildlife Service, U.S. Department of the Interior; 166 p.
- Sauer, J. R., and W. A. Link. 2002. **Hierarchical modeling of population stability and species group attributes from survey data**. Ecology 86:1743-1751.
- Sokal, R. R., and F. J. Rohlf. 1995. **Biometry**. W. H. Freeman and Co., New York, NY.
- Steidl, R. J., J. P. Hayes, and E. Schaubert. 1997. **Statistical power in wildlife research**. Journal of Wildlife Management 61:270-279.
- Thomas, L., and K. Martin. 1996. **The importance of analysis method for breeding bird survey population estimates**. Conservation Biology 10:479-490.
- Zar, J.H. 1999. **Biostatistical analysis**. Prentice-Hall, Upper Saddle River, NJ.