

VOLUME 15, ISSUE 1, ARTICLE 9

Blomberg, E. J., and C. A. Hagen. 2020. How many leks does it take? Minimum samples sizes for measuring local-scale conservation outcomes in Greater Sage-Grouse. *Avian Conservation and Ecology* 15(1):9. https://doi.org/10.5751/ACE-01517-150109 Copyright © 2020 by the author(s). Published here under license by the Resilience Alliance.

Methodology

How many leks does it take? Minimum samples sizes for measuring local-scale conservation outcomes in Greater Sage-Grouse

Erik J. Blomberg¹ and Christian A. Hagen²

¹University of Maine, Department of Wildlife, Fisheries, and Conservation Biology, ²Oregon State University, Department of Fisheries and Wildlife

ABSTRACT. Monitoring population response to conservation actions, such as habitat management, is critical to evaluate conservation outcomes. Greater Sage-Grouse (Centrocercus urophasianus) has been the recipient of substantial recent conservation efforts in North America. Sage-Grouse are often surveyed using counts of males displaying on breeding leks, and these lek counts offer a practical method for monitoring Sage-Grouse population trends. Although substantial work has assessed the utility of lek count data for largescale population monitoring, there has been comparably little effort focused on the use of lek counts to evaluate local-scale management. We used Greater Sage-Grouse lek count data from Oregon, USA, combined with simulation, to evaluate the sample sizes (number of leks, years of monitoring) required to detect a positive outcome of habitat management on population growth. We further assessed assumptions associated with male detection, and compared analyses that both did (N-mixture models) and did not (Poisson regression) account for detection probability. We found that when treatments produced a 5% increase in annual population growth, and leks were monitored for at least 10 years, lek counts produced unbiased and detectable estimates of treatment effects with as few as seven treatment and seven control leks. Using an unbalanced design with a greater number of control leks (n = 16) permitted inference from even fewer treatment leks (n = 4), however, we found no scenarios where use of more control leks permitted detection of smaller treatment effects or allowed shorter duration studies. We found that N-mixture models and Poisson regression of the maximum of three repeated counts produced equivalent results when detection probability was constant, but at the small sample sizes we evaluated, confounding between detection probability and habitat management compromised the accuracy of all analysis methods. Our results show that lek counts hold promise for efficient monitoring of local-scale conservation, but further work is needed to understand the mechanisms that affect male detection during lek surveys.

Combien de leks sont nécessaires? Taille d'échantillon minimum requise pour mesurer les retombées de conservation à l'échelle locale chez le Tétras des armoises

RÉSUMÉ. À la suite de l'instauration d'activités de conservation, telles que l'aménagement d'un habitat, le suivi des populations est essentiel afin qu'on puisse évaluer les retombées sur la conservation. Le Tétras des armoises (Centrocercus urophasianus) a fait l'objet d'efforts considérables récents de conservation en Amérique du Nord. Le dénombrement des mâles nicheurs qui paradent aux leks constitue une méthode d'inventaire fréquente chez le tétras, et ces comptes aux leks s'avèrent pratiques pour suivre les tendances de population. Bien qu'il y ait eu de nombreux travaux attestant de l'utilité des comptes aux leks pour le suivi de population à grande échelle, peu d'efforts ont ciblé l'utilisation de comptes aux leks pour évaluer la gestion à petite échelle. Nous avons utilisé des données de comptes de Tétras des armoises sur des leks en Oregon, É.-U., conjointement à des simulations, afin d'évaluer la taille d'échantillons (nombre de leks, nombre d'années de suivi) requise pour détecter un effet positif de l'aménagement d'un habitat sur la croissance de population. Nous avons ensuite évalué les prémisses relatives à la détection de mâles et avons comparé des analyses qui tenaient compte (modèles N-mélange) ou non (régression de Poisson) de la probabilité de détection. Nous avons établique lorsqu'un traitement entrainait une hausse de la croissance annuelle de population de 5 % et que les leks étaient suivis durant au moins 10 ans, les comptes aux leks permettaient d'obtenir des estimations des effets des traitements non biaisées et détectables avec aussi peu que sept leks traitements et sept leks témoins. L'utilisation d'un échantillonnage non équilibré avec un nombre plus élevé de leks témoins (n = 16) nous a même permis de tirer des conclusions à partir de moins de leks traitements (n = 4); toutefois, nous n'avons trouvé aucun scénario dans lequel l'utilisation d'un plus grand nombre de leks témoins permettait de détecter des effets plus faibles du traitement ou de faire des études de plus courte durée. Nous avons observé que les modèles N-mélange ou la régression de Poisson sur un maximum de trois comptes répétés produisaient des résultats semblables lorsque la probabilité de détection était constante, mais considérant les petits échantillons que nous avons analysés, la confusion possible entre la probabilité de détection et l'aménagement d'un habitat compromet la précision de toutes les méthodes d'analyse. Nos résultats indiquent que les comptes aux leks sont prometteurs pour suivre efficacement les retombées d'activités de conservation à l'échelle locale, mais d'autres travaux sont nécessaires si on veut comprendre les mécanismes qui affectent la détection des mâles durant les comptes aux leks.

Key Words: Greater Sage-Grouse; habitat management; lek counts; N-mixture models; population index

INTRODUCTION

Global shifts in climate and continued conversion of native vegetation to agriculture or other human development has contributed to marked loss of biodiversity (Gaston et al. 2003, Scott et al. 2005). Commensurate with the overall loss of biodiversity has been an increased threat of extinction to "conservation-reliant" species: those that require continued management to persist into the future regardless of legal protections, such as that afforded by the U.S. Endangered Species Act (Scott et al. 2005). The recognition of conservation-reliant species has in some cases led to unprecedented efforts to stabilize or reverse population declines (Miller et al. 2017), involving broad-based coalitions of partners to maintain the implementation of conservation actions (Scott et al. 2010). Three prominent, species-specific examples from North American avifauna include Greater Sage-Grouse (Centrocercus urophasianus; Stiver, Apa, Bohne, et al. 2006, unpublished manuscript), Lesser Prairie-Chicken (Tympanuchus pallidicinctus; Van Pelt et al. 2013), and Northern Bobwhite (Colinus virginianus: Palmer et al. 2011). Because of anthropogenic actions that have disrupted normal ecosystem functions, e.g., historic fire suppression, each of these species requires sustained human intervention, typically in the form of habitat management, to maintain the vegetation communities to which they are adapted. Without such management, vegetation composition transitions to alternate stable states that are often unsuitable for the species, resulting in population decline and local extirpations. Although broad-scale initiatives to benefit these species range in size from regional to international, specific conservation actions that fall under their umbrella, such as habitat management, are often implemented in practice at much more local scales. Understanding how a species responds to these local conservation efforts is paramount to ensure cost-effective methods are actually benefiting species by enabling managers to adapt according to the effectiveness of various actions.

Monitoring conservation outcomes often involves marking individual birds to track changes in behavior (i.e., space use) or demographic rates (i.e., survival, reproduction) in response to management interventions (e.g., Hagen et al. 2011, Gibson et al. 2018). Indeed, such studies are often critical to demonstrate individual responses and evaluate the effectiveness of alternative management techniques (e.g., Coates and Delehanty 2004, McNew et al. 2015, Severson et al. 2017a, b). Detailed study of marked individuals may become cost prohibitive, however, if implemented every time a conservation project is completed. Furthermore, the size of the area affected by management may be too small for individual marking studies if insufficient numbers of animals are available for capture relative to the data demands of particular quantitative methods. Together these limitations may preclude the ability for detailed study of demographic responses to management outcomes at local-scales; yet demonstrating effectiveness at these scales, and addressing shortcomings when goals are not met, are central to adaptive management. More passive, less cost-intensive methods, such as those based on counts of unmarked animals (e.g., Royle 2004), are therefore often required to monitor population response to management at local scales.

Not all animals are encountered during count-based surveys, an almost ubiquitous fact that introduces an inherent confounding

in raw count data between the true population state, e.g., abundance, and the probability of detection (Nichols et al. 2009). A variety of analytical methods exist to estimate detection probability and detection-corrected abundance or density under various survey designs (e.g., Royle 2004, Dail and Madsen 2011). The probability of detecting an individual animal during a given survey period is itself a complex process; animals may be temporarily absent from the survey area, may fail to display during surveys or otherwise be present but not functionally available, or observers may fail to observe and record an animal despite its presence and availability (Nichols et al. 2009). Variation in any of these processes may introduce heterogeneity in detection probability, which can complicate analysis and interpretation of both count-based indices and detection-corrected abundance estimates (Monroe et al. 2019).

The Greater Sage-Grouse (hereafter sage-grouse) is a conservation-reliant species that occupies 668,412 km² of western North America. Despite a recent decision not to protect it under the Endangered Species Act in the United States, significant coordinated conservation efforts continue to work toward reducing threats faced by the species (USFWS 2015, Miller et al. 2017). Although state wildlife agencies have established population goals, linking the conservation actions with population outcomes can be problematic when there is a scale mismatch between large-scale planning objectives and local-scale implementation. In Oregon, for example, conservation actions to benefit sage-grouse, such as removal of coniferous trees from shrub ecosystems (Severson et al. 2017*a*, *b*) are usually < 1000 ha in size and encompass a small number of local breeding sites (typically < 10 breeding leks), while the management areas within which populations are monitored and trends quantified, i.e., priority areas of conservation (PACs), are substantially larger, on average 135,821 ha (SD = 80,054) and containing 46.9 (SD = 34) breeding leks. Significant resources have been devoted to studying detailed behavioral and demographic response of individual sagegrouse to habitat management treatments (e.g., Cook et al. 2017, Severson et al. 2017a, b), and such studies are certainly fundamental to establish best practices and guide future management. But it is impractical to think that such detailed, individual-based research can be used to document the effectiveness of every management action, and as such there is a need for guidance on how to more efficiently monitor sage-grouse population response to management at local scales.

Sage-grouse are a lekking, i.e., communal breeding ground, species and their conspicuous displays each spring provides a convenient method for counting males during courtship. High fidelity to leks and surrounding nesting sites are well documented in sage-grouse, which provides the basis for use of lek counts as an index of population status (Connelly et al. 2011). Leks typically occur in the same vicinity each year with documented rates of continual use exceeding 85 years (Connelly et al. 2011). For all of these reasons, lek counts have been widely used by resource agencies to monitor trends in sage-grouse populations across long terms and large scales (Walsh et al. 2004, Connelly et al. 2011, McCaffery et al. 2016), and their potential for tracking population response to local-scale management may be high. There is a growing body of literature examining various aspects of lek counting methodology, including assessments of both field, e.g., male sightability, and sampling, e.g., repeated vs single count,

considerations (Fedy and Aldridge 2011, Monroe et al. 2016, 2019, Baumgardt et al. 2017, Wann et al. 2019) as well as the frequency of male lek attendance and its relevance to availability for detection (e.g., Blomberg et al. 2013, Fremgen et al. 2019, Monroe et al. 2019, Wann et al. 2019). However, these studies largely addressed sampling questions relevant to large-scale monitoring, such as statewide or subregions within a state, and there is comparably little research addressing sampling concerns surrounding local population monitoring at the relatively small scale of individual habitat treatments. Fundamental concerns for local monitoring, such as the minimum number of leks required to reliably detect a change in sage-grouse abundance following a particular treatment, remain untested.

In this paper, we evaluated sampling considerations for monitoring response of sage-grouse populations to local-scale management. We specifically developed simulations, informed by analysis of existing lek count data, to explore how sample size (number of leks), study duration (number of years), and conservation outcome (management influence on population growth) affect the ability to quantify effects of local management actions on sage-grouse abundance. We approached our evaluation under three alternate methods for data analysis: an index based on a single visit to each lek each year, an index derived from the maximum count of multiple repeated visits to each lek each year, and an *N*-mixture analysis of repeated counts that accounted for imperfect detection probability.

METHODS

Lek count analysis

We first used existing lek count data, fit to N-mixture models (Royle 2004), to derive baseline values and parameterize subsequent simulations. We used data from the Warners PAC in southern Oregon, which was collected as part of the Oregon Department of Fish and Wildlife's sage-grouse lek monitoring program. The Warners PAC was in Lake County, Oregon (Appendix 1). Elevation ranged from 1200 to 2200 m with an average of approximately 1700 m. Although the majority of the study area was dominated by uplands characterized by sagebrushbunchgrass plant associations, mesic areas such as wet meadows, irrigated fields, riparian areas, seeps, and high elevation sagebrush-dominated areas were also available in the study area. This dataset included count data from 56 leks collected between 1975 and 2016 using conventional field methods for sage-grouse lek counts. Lek counts consist of early morning visits to known leks during the spring breeding season (March, April, and May) that are generally repeated ≥ 2 times at each lek within each year. During lek counts observations are made from a stationary vehicle, ground blind, or distant observation point, and observers attempt to count all sage-grouse (Connelly et al. 2003). Data included counts of males, females, and birds of unclassified sex; we focus this analysis on males only because males are typically used for population monitoring in sage-grouse (Connelly et al. 2003, Walsh et al. 2004, Johnson and Rowland 2007, Blomberg et al. 2013). We first subset the data to consider only years from 2000 to 2016 because this time period reflected the range of years where the number of leks counted each year was relatively consistent compared to pre-2000. We excluded counts from the analysis that occurred after 15 May, because attendance rate declines dramatically and counts became more irregular after this date, and we also excluded data where an explicit count of males was not recorded. The resulting dataset comprised 56 leks counted a total of 804 times, representing 359 lek-year combinations, with a range of 1 to 7 counts of any particular lek during a given year (mean = 2.24 counts/lek/year). The mean number of years a single lek was monitored was 6.4 (range = 1-17 years).

We used the "pcount" function within the R package unmarked (Fiske and Chandler 2011) to fit *N*-mixture models as described by Royle (2004). Under this formulation of the *N*-mixture analysis, annual dynamics of the population can be assessed by fitting a categorical year effect (individual beta coefficient for each year), while multiyear trends can be assessed using a continuous trend effect (single beta coefficient for the effect of year as a continuous variable). This later approach approximates a basic exponential model of population growth,

$$N_t = N_0 e^r \tag{1}$$

where r reflects the instantaneous rate of population growth. Under a generalized linear modeling framework, an equivalent model of abundance is given as

$$log(N_t) = \beta_0 + \beta_1 Year_t$$
(2)

where the slope coefficient of the continuous year effect (β_1) is equivalent to the population growth rate *r*, in that it describes the mean rate of change of the population as a function of a linear effect of time, and the model intercept (β_0) reflects the log of mean initial abundance.

Three alternative distributions can be used in modeling abundance under the N-mixture framework of unmarked: negative binomial, Poisson, and zero-inflated Poisson. We first evaluated the best approximating distribution by fitting three competing models of identical structure that differed only in use of these three distributions. Each model allowed full annual variation in abundance (year as a categorical effect) and full annual variation in detection probability, as well as an ordinal date, i.e., day of year, term as a continuous effect to allow for some within-year variation in counts. We used AIC_c to select among these three competing models; the model fit under a negative binomial distribution was 2572 AIC_c better supported than the next best supported model, and we used it in all subsequent analyses. We next ran three additional models that included only a year effect, only a date effect, and no effect (intercept only null) on detection probability, in each case allowing full annual variation in mean abundance via a categorical year effect. We used the resulting parameter estimates from the best-fit model as input parameters for our simulations, as described below.

Data simulation

Basic structure

Simulations generally followed code provided in Kéry and Royle (2016), with modifications as needed to match our particular objectives. Our approach was also generally consistent with earlier simulations of sage-grouse lek count data (McCaffery et al. 2016, Monroe et al. 2019). We first simulated the true underlying state process. The initial abundance (N) of males at every j lek was Poisson distributed as defined by mean abundance (λ) and lek-level random variability (l_j)

$$N_i = Pois(\lambda + l_i) \tag{3}$$

We used the intercept of the abundance model fit to the Warner PAC data to define λ , and allowed l_j to vary among leks at a rate of ± 0.10 under a uniform distribution. We then simulated change in abundance at each lek for each subsequent year as

$$N_{(t+1)j} = N_t e^{(T_j + r_t)}$$
(4)

where for each lek/year combination, a lek-specific trend in the population growth rate (T_j) and annual (r_j) random variability influenced change from the previous year's abundance based on scenario-specific values described below. Our approach implied annual variation was consistent among leks, i.e., the population at large experienced the same underlying temporal processes, and produced simulated time series that were comparable to estimates from the Oregon lek count data.

From the abundances we further simulated counts of males on leks. These were again informed by results of the lek count analysis, such that simulated counts would mimic the typical structure of counts collected in the field. Here, the count (C) obtained during every *k* survey at each *j* lek and *t* year combination was drawn from a binomial distribution

$$C_{tik} = binom(N_{ti}, p_{ik}) \tag{5}$$

where counts were constrained by the known simulated abundance at each lek during each year, as well as lek-, year-, and count-specific variation in detection, defined on a logit-scale as

$$logit(p_{tjk}) = \alpha_0 + t_t + v_{tjk}$$
(6)

where α_0 defined the mean detection probability based on the intercept of the detection model fit to the Warner PAC data, t_i allowed random annual variation in detection probability that approximated the annual variation we observed in the Warner PAC results, and v_{ijk} was drawn from a normal distribution with a mean = 0.0 and SD = 0.05 to allow for count-level variability at each lek that reflected within-season variation in *p*. Collectively, t_i and v_{ijk} reflect among- and within-year variation in the detection process, respectively, which could arise from some combination of variable lek attendance, male sightability, or observer abilities during lek counts. We did not attempt to distinguish among distinct biological processes affecting variation in detection probability (but see Monroe et al. 2019).

Simulating and quantifying habitat manipulation effects

To simulate response of the population to effects of habitat manipulation, we defined simulated leks as either treatment (received habitat management) or control (no management) leks, and applied a different lek-level population growth rate (T_j) using equation 4 above. This approach mimicked a control-impact study design, where we assumed habitat manipulation, e.g. conifer removal, would illicit a difference in subsequent population trend between leks affected by treatments versus those not affected. Building on equation 2, the appropriate generalized linear model to test for evidence of a treatment effect is

$$\log(N_{tj}) = \beta_0 + \beta_1 Year_t + \beta_2 Treat_j + \beta_3 Year_t Treat_j$$
(7)

where the presence (*Treat* = 1) or absence (*Treat* = 0) of habitat manipulation treatments at a given lek changes the expected population growth rate as a function of the interaction between treatment and year, expressed by the β_3 term. When fitting this model, the estimated value of β_3 provides the distance between baseline growth rate (as defined by β_1) and the growth rate associated with treatment leks. As such, β_3 provides the estimated effect of habitat manipulations on population growth. For example, given a 5% increase in the annual growth rate for treatment leks, we would expect a resulting estimate of $\beta_3 = 0.05$ from a model that accurately captures the treatment effect. In practice, we implemented simulations assuming stable population growth (i.e. r = 0.0) at control leks, and assessed varying levels of effects for treatment leks (described below).

Executing simulations and assessment of results

We replicated simulations under a number of different scenarios that we designed to address specific questions about lek count utility for tracking effects of local-scale habitat treatment. In doing so we varied a number of different simulation variables (summarized in Table 1), including the number of leks monitored, annual variation in population growth (r_i) , the number of years in the dataset, and the magnitude of the treatment effect. To reduce the overall volume of scenarios we addressed individual questions in discrete stages, rather than attempting to run all possible combinations of all variables. First we evaluated the minimum sample size (number of treatment and control leks) required to detect a treatment effect across a 10-year period based on samples of 8, 14, and 20 leks. Here we assumed a balanced sampling design with 4, 7, and 10 treatment leks, and an equal number of control leks. The minimum sample size required to detect a treatment effect is clearly sensitive to the strength of the effect, so we replicated each of the three sample size levels using effect sizes, i.e., change in r; T_{i} , of 0.01, 0.025, 0.05, 0.075, and 0.10. Furthermore, we considered that background variation in the population growth rate not associated with a treatment effect, e.g., random annual variation in r; r_r , could produce noise in count data, where a sufficient degree of annual variation may obscure treatment effects. We replicated our initial scenarios accordingly by drawing r, from a normal distributing with mean = 0, and altering the SD of r_i using values of 0.05, 0.10, 0.15, and 0.20. Our initial scenarios varying sample size, treatment effect, and annual variation in growth rate resulted in 60 total scenarios.

We next considered that in certain circumstances it may be feasible to include a greater number of control leks relative to treatment leks, which would increase the overall sample size and could possibly improve model estimation and facilitate detection of treatment effects that would not be possible under a balanced design. Here, we used results from our initial assessment to conduct a series of more focused simulations that asked whether increasing the number of control leks permitted detection of treatment effects under a smaller effect size, shorter study duration, or with fewer treatment leks. We chose a representative simulation from earlier runs, increased the number of control leks by factors of 4 and 8 times the number of treatment leks, and evaluated changes in results that occurred between the original run and those with an inflated control sample size.

We also considered that nonrandom patterns in detection of male sage-grouse on leks could bias assessments of management **Table 1**. Summary of objectives and descriptions of scenarios that were used to conduct simulations designed to explore the utility of Greater Sage-Grouse (*Centrocercus urophasianus*) lek counts for detecting localized effects of habitat management on population growth rates.

Objective	Changes to baseline simulation ^{\dagger}			
Minimum number of leks	The number of leks varied by 8, 14, and 20 with an even sampling design			
	(equal treatment and control leks); ran for all effect sizes.			
Minimum treatment effect size	Treatment effect increased annual population growth by 0.01, 0.025, 0.05,			
	0.075, and 0.10; ran for all numbers of leks.			
Annual variance in population trend	Varied SD of annual growth rate (r) by values of 0.05, 0.10, 0.15, and 0.20.			
Study duration	Total number of years leks were monitored varied by 5 and 10.			
Increased number of control leks (effect size)	Changed ratio of treatment:control leks from 1:1 to 1:4 and 1:8; effect size			
	of 0.025.			
Increased number of control leks (number of leks)	Changed ratio of treatment:control leks from 1:1 to 1:2 and 1:4; number of			
	treatment leks $= 8$.			
Increased number of control leks (study years)	Changed ratio of treatment:control leks from 1:1 to 1:4 and 1:8; 5-year			
	study length.			
Temporal trend in detection	Detection probability declined at a rate of -0.05 per year, but the trend was			
<u>p</u>	similar among all leks			
Confounding trend in detection	Detection probability declined at a rate of -0.05 per year for only the			
comounding trend in detection	treatment leks was constant for control leks			
[†] The baseline model specifications were as follows: numb	per of leks: 14 (7 treatment, 7 control); repeated visits per lek: 3; years: 10;			

The baseline model specifications were as follows: number of leks: 14 (7 treatment, 7 control); repeated visits per lek: 3; years: 10; treatment effect (change in trend on treatment leks): 0.05; annual variance in r: 0.05; trend in detection probability: 0.0. If not noted explicitly above all simulations follow these parameter values.

effectiveness, particularly if detection probability exhibited systematic change that covaried with trends in abundance. Population trend assessments that are based on indices, i.e., raw counts, should be particularly sensitive to this source of error, however the ability of N-mixture models to disentangle covariance among p and N has also received recent attention (Dennis et al. 2015, Barker et al. 2018, Monroe et al. 2019). We felt this might be particularly relevant at the relatively small sample sizes inherent to our specific objectives. We evaluated two different scenarios, one where we introduced a negative trend in p for all leks, and a second where we included the same negative trend but only for treatment leks. We speculated that such trends in p could exist if changing environmental conditions altered sagegrouse breeding behavior systematically through time. For example, if daily lek attendance is influenced by rates of competition among males, then density-dependence in detection probability, i.e., a covariance between p and N, may be likely. Following this logic, trends in p that are confounded with habitat management could be expected in a scenario where treatments exhibit a positive response in abundance, that in turn elicits density-dependent effects on detection at only the treatment leks. Motivated by these concerns and based on our initial assessments, we explored a series of additional scenarios where we altered the presence or absence of trends (both general and interacting with treatment) in both p and N, and explored whether models could correctly identify the specified trends under each scenario. We used AIC_c to evaluate relative support for model structures that included combinations of general trends in p, general trends in N, and treatment*year interactions in either p or N. Specifically, we asked whether model selection was consistently able to identify confounding trends in p when they existed, if p-trend models would ever be selected when no true confounding existed, and whether estimates of treatment effects exhibited bias (described below) in situations where there were general or confounding trends on *p*.

For all scenarios, we evaluated estimates of treatment effects (interaction β) from the *N*-mixture analysis in two ways. We first assessed bias in the estimated treatment effect by comparing it to the true underlying trend in abundance, which we derived using a Poisson regression on the "known" abundances for each lek (N_{ii}) , fit using a model equivalent to equation 7. We used this approach rather than assessing bias based solely on the specified effect in the simulation (e.g. 0.025, 0.05, etc.) because the stochastic components of our simulations meant that the true underlying trend in the data regularly deviated from the specified mean. We calculated bias for each iteration as the difference between the interaction effect from the N-mixture model and the interaction effect from the true abundance data, and summarized bias based on the mean and range of bias values observed for all iterations within a particular simulation. We also evaluated the ability of any given analysis to identify statistical support for a treatment effect, given that it existed, by evaluating whether the 85% confidence intervals of the beta coefficient of the treatment effect overlapped 0.0 (no effect) for each iteration (Arnold 2010). Within a particular simulation we summarized the proportion of iterations where treatment effects could be distinguished.

For all simulations, we assessed whether *N*-mixture models provided increased inference relative to a simpler index-based approach that only used the raw lek-count data. We approached this assessment using two alternative indices; the maximum value of all counts obtained from a particular lek during a given year (hereafter max count), and a single count selected at random from within a given year (hereafter random count). In each case, we fit a generalized linear model with a Poisson distribution to the count index using a model structure equivalent to equation 7. We assessed bias and precision in the resulting estimates of the treatment effect as described above for the *N*-mixture analysis.

We initially chose to use the single season model of Royale (2004) over the alternative dynamic *N*-mixture formulation of Dail and Madsen (2011). We based this decision on the assumption that



Fig. 1. Annual estimates of (A) detection probability and (B) mean male abundance, derived from *N*-mixture analysis of Greater Sage-Grouse (*Centrocercus urophasianus*) lek count data, collected in the Warners Priority Area of Conservation (PAC), southern Oregon, 2000–2012.

the more complex structure of the dynamic models may not perform well under the low sample sizes inherent to our objectives, and because single season models can capture multiyear dynamics, i.e., equation 2. To evaluate these assumptions, we conducted a single simulation with 100 iterations based on a constant structure, ran equivalent single-year and dynamic Nmixture models for each iteration using appropriate functions in the unmarked package (pcount and pcountOpen, respectively), and compared both bias and 85% confidence intervals of treatment effects, as described above, between the two model types. We chose to use 85% confidence intervals largely for its congruence with the common criteria of 2.0 Δ AIC used during AIC-based model selection (Arnold 2010)

RESULTS

The best supported *N*-mixture model of male sage-grouse abundance in southern Oregon included effects of survey date ($\beta = 0.01 \pm 0.001$ SE) and a categorical year effect on detection probability, as well as a categorical year effect on abundance (Table 2). Year-specific predictions from this model demonstrate a large degree of annual variation in both detection probability and abundance (Fig. 1). Annual estimates of detection probability ranged from $p = 0.25 \pm 0.07$ SE to $p = 0.50 \pm 0.03$, while estimates of mean annual abundance (males/lek) ranged from $\lambda = 21.4 \pm 7.3$ SE to $\lambda = 69.4 \pm 37.4$ SE. **Table 2**. Model selection results from an initial analysis of Greater Sage-Grouse (*Centrocercus urophasianus*) lek count data collected in the Warners Priority Area of Conservation (PAC), southern Oregon, 2000–2012. Models were fit in an *N*-mixture framework using the R package unmarked.

Model [†]	AICc	∆AICc	w _i	K
p(date + year) N(year)	6177.85	0.00	1.00	36
p(date) N(year)	6265.41	87.56	0.00	20
p(year) N(year)	6296.28	118.43	0.00	35
p(.) N(year)	6413.59	235.74	0.00	19
1.				

 † date = ordinal day of year; year = categorical year effect; (.) = intercept only null model.

Mean treatment effects derived from *N*-mixture models were generally unbiased across a wide-range of sample sizes (numbers of leks), treatment effect sizes, and annual variation in population growth rate (Fig. 2A), however there was variation among individual model runs with maximum biases approaching ± 0.025 . For example, if treatments elicited a 0.05 increase on population growth at treatment leks, *N*-mixture models would on average capture that effect accurately, but would also occasionally return estimates as low as 0.025 or as high as 0.075. The range of absolute bias was generally narrow across all treatment effect sizes (Fig. 2A), which implies that proportional bias declined as effect size increased. We also found that *N*-mixture estimates offered only marginal improvement in performance over indices based on the maximum count of males during a given year. Max count models demonstrated a small (arguably trivial) amount of mean negative bias and a similar range in bias among all iterations when compared to *N*-mixture estimates (Fig. 2D). Both *N*-mixture and max count indices were superior to random count indices, which were unbiased on average but showed a larger range of variation in bias among iterations compared to the other approaches (Fig 2D).

Fig. 2. Collective results of 60 simulations designed to explore the utility of Greater Sage-Grouse (Centrocercus urophasianus) lek counts for detecting localized effects of habitat management on population growth rates. We compared the estimated effect against the known difference in population growth rate between treatment and control leks for (A) N-mixture models, (B) a population index based on a single random count, and (C) a population index based on the maximum of three repeated counts, where the dashed line reflects a perfect 1:1 correlation. We evaluated (D) error for all three analyses as the distance between the true and estimated effect. Simulations included varying levels of treatment effect size (ranges from 1 to 10%) difference in trend between treatment and control leks), sample size (either 8, 14, or 20 leks with even distribution of treatment and control), and annual variation in the population growth rate (0.05, 0.10, 0.15, and 0.20; depicted in panel D). Other simulation parameters are provided in Table 1.

The ability of *N*-mixture models to consistently detect a treatment effect over a 10-year period required a minimum sample of 14 leks (7 treatment and 7 control) under a balanced sampling design when treatments resulted in a 0.05 increase in population growth

rate (Fig. 3). At this level of effect and sample sizes, the 85% confidence intervals of the treatment effect on population growth did not overlap 0.0 in 98% of iterations, meaning that *N*-mixture models had near-perfect ability to detect habitat management effects at this level. With larger effect sizes, i.e., 0.075 and 0.10, a smaller sample of 4 treatment and 4 control leks was capable of consistently detecting the effects, however, none of the sample sizes we considered could detect smaller effect sizes (Fig. 3). For example, with a sample of 20 leks (10 treatment and 10 control) only 72% of iterations were able to conclusively determine that a 0.025 treatment effect existed, while for 28% of iterations 85% confidence intervals of the effect overlapped 0.0. These general pattern were consistent for max count and random count approaches as well (Appendix 1).

Fig. 3. Range of lower 85% confidence limites for the year* treatment interaction term in *N*-mixture analyses of simulated Greater Sage-Grouse (*Centrocercus urophasianus*) lek count data. Each box and whisker represents 400 iterations, which varied according to the specified mean effect size (positive increase in population growth for treatment leks relative to controls) and total number of leks in the sample (equal balance between treatment and control leks). The vertical dashed line indicates the point at which lower confidence limit falls below 0.0, suggesting no support for the effect. Therefore, box and whiskers entirely above the line indicated cases where > 90% of iterations produced modeled slope coefficients with 85% confidence intervals that did not overlap 0.0. Each boxplot contains simulations where annual variation in the population growth rate varied from 0.05 to 0.20 (see Fig. 4).

Increasing annual variation in the population growth rate, i.e., background variation not associated with a treatment, did not affect bias in the treatment effect (Fig. 2D), but did result in some small but noticeable changes in the confidence intervals of the effect (Fig. 4). For example, when 7 treatment and 7 control leks were sampled under an effect size of 0.025, the lowest amount of variation we considered (SD_r = 0.05) was able to distinguish treatment effects based on 85% confidence intervals in 65% of simulations, while under the greatest amount of variation (SD_r = 0.20) only 42% of simulations could detect that level of effect. Nevertheless, in scenarios with 14 leks, a treatment effect of 0.05, and a 10-year study design, > 98% of iterations were capable of distinguishing a treatment effect based on 85% confidence intervals, irrespective of the annual variation in growth rate we considered.

Fig. 4. Range of lower 85% confidence limits for the year* treatment interaction term in *N*-mixture analyses of simulated Greater Sage-Grouse (*Centrocercus urophasianus*) lek count data. Each box and whisker represents 300 iterations, which varied according to the specified mean effect size (positive increase in population growth for treatment leks relative to controls) and annual variability in the population growth rate (r), expressed as a standard deviation (SD) of the mean. The vertical dashed line indicates the point at which lower confidence limit falls below 0.0, suggesting no support for the effect. Therefore, box and whiskers entirely above the line indicated cases where > 90% of iterations produced modeled slope coefficients with 85% confidence intervals that did not overlap 0.0. Each boxplot contains simulations where the total number of leks monitored ranges from 8 to 20 (see Fig. 3).

Single season and dynamic *N*-mixture models produced fairly equivocal results, with dynamic models yielding a slight bias in treatment effect estimates relative to the single-season models (Fig. 5A). Single season models were more superior to dynamic models with respect to support for treatment effects, however (Fig. 5B). Only 1 out of 100 iterations produced a single-season estimate of treatment effects with 85% confidence intervals that overlapped 0, compared with 10% of estimates from the dynamic models that overlapped 0 under the same simulation specifications.

When we increased the sample of control leks, we could not consistently detect effect sizes < 0.05 even given relatively large samples of control leks (Fig 6A). With a ratio of 1:4 treatment leks (n = 7) to control leks (n = 28), 19% of iterations could not distinguish a treatment effect of 0.025 from 0.0, and results were similar at a ratio of 1:8 (n = 56 control leks). Results were similar for index-based approaches for these two scenarios (Fig 6A). Larger numbers of control leks also did not allow for shorter study duration (Fig. 6C) because both 1:4 and 1:8 ratios still produced confidence intervals on treatment effects that overlapped 0.0 for the majority of iterations (61% and 47%, respectively) during a 5-year duration study. Therefore, the total number of treatment leks ultimately limited ability to detect relatively small (≤ 0.025) or shorter term (5 year) effects of habitat management. Increasing the sample of control leks did facilitate a smaller sample of treatment leks when effects were more moderate (≥ 0.05) and study duration was 10 years. At a 1:4 ratio, an effect size of 0.05 was conclusively detected for 99% of iterations with as few as 4 treatment leks (Fig. 6B), suggesting that in situations with very few focal treatment leks, using a larger number of control leks may be a useful strategy.

Fig. 5. Comparison of model error (A) and lower 85% confidence limits (B) between multiseason (Dail and Madsen 2011) and single season (Royle 2004) N-mixture models applied to simulated Greater Sage-Grouse (Centrocercus urophasianus) lek count data. Error represents the distance between model estimates of a treatment effect on population growth and the true realized effect in the simulated data, where the dashed line reflects perfect agreement (error = 0.0). Lower 85% confidence limits below 0.0 (dashed line in panel B) represent models where the treatment effect could not be distinguished from 0.0 (no effect). For each iteration (N = 100 per simulation), models were fit to an identical dataset with the following parameters: number of leks: 14 (7 treatment, 7 control); repeated visits per lek: 3; years: 10; treatment effect (change in trend on treatment leks): 0.05; trend in detection probability: 0.0; annual variation in population growth rate (SDr): 0.10.

In scenarios that included a general trend in detection probability that was shared among treatment and control leks, results from *N*-mixture models were similar to those from scenarios with no variation in detection (Fig. 7A). Mean bias increased slightly for the index-based assessments, although a relatively large proportion of iterations still produced unbiased results (Fig. 7A). When temporal trends in detection were confounded with treatment effects, however, substantial bias was introduced in estimates from all three approaches (Fig. 7B). Inclusion of a year* treatment interaction on the detection term counteracted bias in *N*-mixture models to some extent, but this also tended to inflate the range of biases observed across iterations (Fig. 7C) making estimates from *N*-mixture models less reliable in the presence of confounding effects on detection probability, at least in the specific scenarios we explored.

Fig. 6. Lower limits of 85% confidence intervals of treatment effects under an unbalanced sampling design where a larger number of control leks were used relative to treatment leks in scenarios where (A) effect size was specified as 0.025, (B) only 4 treatment leks were included, and (C) only 5 years were included in the simulation. All other simulation parameters are described in Table 1. Each box and whisker represents 100 iterations, which varied according to the ratio of treatment: control leks as identified in the figure panel legends. The vertical dashed line indicates the point at which the confidence interval overlap 0.0, suggesting no effect. Therefore, box and whiskers entirely above the line indicated cases where > 90% of iterations produced modeled slope coefficients with 85% confidence intervals did not overlap 0.0.

Fig. 7. Assessment of bias in modeled treatment effect when (A) detection probability exhibited a -0.05 annual trend, (B) when only treatment leks exhibited a -0.05 annual trend in detection probability and the model was fit with a continuous year effect on detection probability, and (C) when only treatment leks exhibited a -0.05 annual trend in detection probability and the model was fit with a treatment*year effect on detection probability. In each case, error represents the deviation in the modeled treatment effect from the true underlying population trend on treatment leks.

Model selection was generally not able to identify biases associated with confounding detection issues. Under a scenario that included 14 leks with a balanced sample of treatment and control leks, and where there was no treatment effect on abundance but a treatment effect on detection, the correctly specified detection structure was favored by model selection in 59 of 100 simulations. In other words, 59% of iterations correctly identified the confounding effect of treatment on detection probability without incorrectly attributing the effect to changing abundance. However, under a scenario with no treatment effect on p but a true treatment effect on abundance, the same model structure (treatment*trend interaction on p) was selected as the best model in 49 of 100 iterations. When there was a full confounding, i.e., treatment effects on both detection and abundance (the scenario depicted in Fig. 7C), only 22% of iterations identified the treatment*trend interaction on p, and very few analyses (7%) identified the correctly specified model structure with the interaction applied to both p and N. Collectively these results suggest a limited ability of model selection to help elucidate confounding treatment effects on detection probability, at least with the relatively small sample sizes we explored during this study.

DISCUSSION

As management efforts continue for conservation-reliant species, like sage-grouse, the need to efficiently monitor the outcome of a rising volume of local conservation efforts will increase. We are encouraged by our findings demonstrating that such local-scale monitoring using a modest number of leks can detect trends in sage-grouse abundance associated with potential management actions. With as few as 7 treatment and 7 control leks, we could detect $\geq 5\%$ annual change in abundance over a 10-year period, but we found diminishing returns with lower sample sizes, smaller effect sizes, or fewer years of monitoring. With as few as 4 leks per observational unit, we were able to detect 10% trends in annual abundance, however this effect size may be unreasonably large to assume for many management actions. Therefore, in general, we recommend a minimum sample size of 7 treatment and 7 control leks monitored for at least 10 years to effectively monitor local conservation outcomes using lek counts.

The estimates of detection probability that we derived from lek count data in Oregon exhibited annual variation (p = 0.25-0.53) that fell within the range of those reported previously in other portions of sage-grouse range (McCaffery et al. 2016). As such, we suggest that results of our simulations may be extended to sage-grouse populations outside those in Oregon. When we increased the annual variability in population growth, we found a slight decrease in the ability to distinguish treatment effects from background noise. This suggest that in systems exhibiting more stochastic dynamics, i.e., $SD_r > 0.20$, a larger sample size of treatment leks may be preferable compared with more stable systems.

We found that under an unbalanced sample size, use of a greater number of control leks did not dramatically improve the ability to detect smaller effect sizes or enable shorter study periods, relative to a balanced design. The one area where this technique did seem to improve estimation was for smaller samples of treatment leks, where a 1:4 ratio of treatment to control leks permitted consistent detection of a 5% annual change in abundance for as few as 4 treatment leks. Therefore, under a scenario where fewer than 7 leks are included in a management plan, pairing those leks with larger sample of control leks at a ratio of 1:4 should provide sufficient power to detect management effects, if they exist. This may provide additional flexibility for managers working with small, isolated populations or in situations where resources constrain the extent of conservation actions to a small subset of leks.

Our findings contrast previous work that suggested single-counts of 150 leks were necessary for detecting annual changes of 10% (Fedy and Aldridge 2011). This difference is likely due to differences in the temporal scale of trends estimated by Fedy and Aldridge (2011) compared with those during our study; we assessed multiyear, e.g., 5-year, 10-year, systematic trend in abundance, whereas Fedy and Aldridge (2011) evaluated annual change in a smoothed population growth rate derived from generalized additive models. We did not attempt to quantify shortterm dynamics during our simulations or assess model accuracy in capturing changes in abundance on a year-to-year basis. Given the results of Fedy and Aldridge (2011), we suspect capturing such short-term dynamics would be challenging at the relatively small sample sizes we considered during our study. The necessity of longer time intervals to detect effects during our study was certainly due to the greater degree of sampling variation inherent to relatively small sample sizes. Although the true treatment effect was present in our simulations for all years, a decade of observation was required to separate it from the background annual variation inherent to sage-grouse population growth. Significant annual variation, both stochastic and cyclic, has been regularly demonstrated and is likely ubiquitous among sagegrouse populations (Connelly et al. 2011, Fedy and Doherty 2011, Blomberg et al. 2012, Rowe and Fedy 2017). The appearance of population response to conservation actions or negative impacts may therefore be inherently slow, but we are encouraged that lek counts at local scales may be used to detect reasonable changes in populations. We therefore recommend that when monitoring local-scale conservation actions, practitioners focus on, and plan for, relatively long-term assessments, e.g., > 10 years, and avoid drawing too great an influence from short-term changes immediately following treatments. Our simulations also assumed that population growth increased immediately following treatments; if in reality populations lag in response, e.g., because of delayed recovery of vegetation communities, then > 10 years may be necessary to establish positive effects of treatments. Furthermore, if treatment effects are expected to illicit < 5%change in annual population growth, study durations > 10 years may be necessary.

We found, somewhat unexpectedly, that *N*-mixture models produced only marginal improvement in trend estimation relative to those from indices based on the maximum observed count of males during repeated surveys. McCaffery et al. (2016) presented *N*-mixture models as a more robust method for lek count estimation, and demonstrated measurable differences between abundance estimates derived from *N*-mixture models and maximum counts when analyzing lek count data (n = 1322) across the entire state of Montana over a 12-year period. The authors further showed that trend estimates derived from *N*-mixture analysis deviated from those based on a population index; the former was able to distinguish a 7% annual decline, while the latter was not. The discrepancy between our results and those of McCaffery et al. (2016) may be due to differences in our treatment of the index values and approaches to trend estimation from them. McCaffery et al. (2016) estimated a mean and variance of the log population growth rate (equation 6 in McCaffery et al. 2016), while we used the slope term from an annual trend fit using a Poisson regression. Barker et al. (2018) suggested that Poisson regression may produce equivocal or slightly superior estimates relative to N-mixture models, particularly for estimates of abundance from sparse data. It is also important to recognize that we focused on population trend in our assessments and did not evaluate the models' ability to predict abundance, for which indices are clearly inferior when p < 1.0. Recent work by Monroe et al. (2019) clarifies many of these relationships for larger sample sizes (n = 100 leks). These authors found that index-based methods, modeled using either Poisson or Negative Binomial regression of maximum counts, showed clear bias in estimation of abundance but produced generally unbiased trend estimates in the absence of confounding trends in detections (Monroe et al. 2019).

Abundance estimates from N-mixture models may be susceptible to bias when there is heterogeneity in detection (Dennis et al. 2015, Veech et al. 2016, Barker et al. 2018, Monroe et al. 2019) that is not captured with a correctly specified model (Monroe et al. 2019). Our simulations illustrated that when changing detection probability was confounded with the effects of habitat treatment, N-mixture models regularly produced biased estimates of the treatment effects, even under correctly specified model structure. The small sample sizes inherent to our particular questions no doubt exacerbated this issue, and while model selection provided some ability to illustrate the presence of confounding effects, this was also inconsistent. In contrast, Monroe et al. (2019) found that N-mixture models following a repeated sampling design were capable of disentangling confounding trends between p and N. Our findings do not necessarily contradict those of Monroe et al. (2019), for two important reasons. First, Monroe et al. (2019) used comparatively large samples (n = 100 leks) in their simulations, presumably with greater power to disentangle parameter covariance. Second, our design added an additional level of confounding (that of habitat treatment) that was not present in the assessment of Monroe et al. (2019); when we considered only a general confounding between p and N that was not associated with treatment, we found that N-mixture models were unbiased. False positive detections, such as resulting from double counting, may also introduce further heterogeneity in detection probability. Our findings illustrate that independence between the effects of habitat treatments and detection probability is a fundamental and critical assumption that must be met when using lek counts to evaluate conservation outcomes at small scales, irrespective of the analytical method used to quantify the trend estimate.

The probability of counting a male during a lek survey is a product of multiple factors (Monroe et al. 2019), including the probability of male lek attendance, the probability of male display during a count, given attendance, the visibility of the male to an observer while displaying, and the ability of the observer to accurately count and record a visible male. These factors are likely ubiquitous and dynamic, varying among years, leks, and observers (Walsh et al. 2004, Blomberg et al. 2013, Baumgart et al. 2017, Monroe et al. 2019, Wann et al. 2019). Our knowledge of the mechanisms that influence male detection probability during lek counts is, generally speaking, imperfect, and we did not set out to fully disentangle these issues during our study. Monroe et al. (2019) considered the relative influence of various detection components on performance of both N-mixture and maximum count indices, and found that trends in male attendance had a greater effect on model bias than trends in detection associated with visibility, i.e., Fig. 3 in Monroe et al. 2019. The authors also found that use of ancillary data on lek attendance obtained from GPS-marked males could be used to correct for latent heterogeneity in detection and improve estimation for both index- and N-mixture-based approaches to trend estimation (Monroe et al. 2019). Further work exploring the mechanistic relationship between sage-grouse behavior and detection probability during lek counts, particularly with respect to habitat management activities, is warranted. Such information could further improve the ability to use lek counts for evaluating local-scale conservation measures while addressing assumptions related to detection probability.

Lekking species of grouse will likely require continual management at landscape scales to ensure their persistence in the United States. There are unprecedented efforts underway to conserve sage-grouse and the habitats they occupy (USFWS 2015), and finding practical methods to monitor the biological outcomes of these efforts are paramount. We are encouraged by our findings as a first step in identifying sample sizes necessary to monitor populations at the scale at which most management occurs. However, further work is needed to understand how nuanced differences in behavior among lekking grouse species and system-specific variation within species affect detection probability during count-based surveys. We encourage others to pursue this work, and to use simulations such as ours to determine the level of sampling effort required for other species and systems.

Responses to this article can be read online at: http://www.ace-eco.org/issues/responses.php/1517

Acknowledgments:

We thank numerous volunteers and professional biologists who have collected lek data in Oregon over the years. Oregon Department of Fish and Wildlife provided lek count data for model parameterization. This work was supported in part by a cooperative agreement between Bureau of Land Management and Oregon State University. We received helpful comments on the manuscript from G. Sadoti, W. Thogmartin, and an anonymous reviewer. This project was supported by the USDA National Institute of Food and Agriculture, McIntire-Stennis project no. ME041602 through the Maine Agricultural and Forest Experiment Station. This is Maine Agricultural and Forest Experiment Station Publication no. 3736.

LITERATURE CITED

Arnold, T. W. 2010. Uninformative parameters and model selection using Akaike's information criterion. *Journal of Wildlife Management* 74:1175-1178. https://doi.org/10.1111/j.1937-2817.2010. tb01236.x

Barker, R. J., M. R. Schofield, W. A. Link, and J. R. Sauer. 2018. On the reliability of N-mixture models for count data. *Biometrics* 74:369-377. https://doi.org/10.1111/biom.12734

Baumgardt, J. A., K. P. Reese, J. W. Connelly, and E. O. Garton. 2017. Visibility bias for Sage-Grouse lek counts. *Wildlife Society Bulletin* 41:461-470. https://doi.org/10.1002/wsb.800

Blomberg, E. J., J. S. Sedinger, M. T. Atamian, and D. V. Nonne. 2012. Characteristics of climate and landscape disturbance influence the dynamics of Greater Sage-Grouse populations. *Ecosphere* 3:1-20. http://dx.doi.org/10.1890/ES11-00304.1

Blomberg, E. J., J. S. Sedinger, D. V. Nonne, and M. T. Atamian. 2013. Annual male lek attendance influences count-based population indices of Greater Sage-Grouse. *Journal of Wildlife Management* 77:1583-1592. https://doi.org/10.1002/jwmg.615

Coates, P. S., and D. J. Delehanty. 2004. The effects of raven removal on Sage-Grouse nest success. *Proceedings of the 21st Vertebrate Pest Conference* 21:17-20.

Connelly, J. W., C. A. Hagen, M. A. Schroeder. 2011. Characteristics and dynamics of Greater Sage-Grouse populations. Pages 53-67 in S. T. Knick and J. W. Connelly, editors. *Greater Sage-Grouse: ecology and conservation of a landscape species and its habitats.* Studies in Avian Biology No. 38. University of California Press, Berkeley, California, USA. https:// doi.org/10.1525/california/9780520267114.003.0004

Connelly, J. W., K. P. Reese, and M. A. Schroeder. 2003. *Monitoring of Greater Sage-Grouse habitats and populations*. College of Natural Resources Experiment Station, University of Idaho, Moscow, Idaho, USA. https://doi.org/10.5962/bhl. title.153828

Cook, A. A., T. A. Messmer, and M. R. Guttery. 2017. Greater Sage-Grouse use of mechanical conifer reduction treatments in northwest Utah. *Wildlife Society Bulletin* 41:27-33. https://doi. org/10.1002/wsb.742

Dail, D., and L. Madsen. 2011. Models for estimating abundance from repeated counts of an open metapopulation. *Biometrics* 67:577-587. https://doi.org/10.1111/j.1541-0420.2010.01465.x

Dennis, E. B., B. J. T. Morgan, and M. S. Ridout. 2015. Computational aspects of N-mixture models. *Biometrics* 71:237-246. https://doi.org/10.1111/biom.12246

Fedy, B. C., and C. L. Aldridge. 2011. The importance of withinyear repeated counts and the influence of scale on long-term monitoring of Sage-Grouse. *Journal of Wildlife Management* 75:1022-1033. https://doi.org/10.1002/jwmg.155

Fedy, B. C., and K. E. Doherty. 2011. Population cycles are highly correlated over long time series and large spatial scales in two unrelated species: Greater Sage-Grouse and cottontail rabbits. *Oecologia* 165:915-924. https://doi.org/10.1007/s00442-010-1768-0

Fiske, I., and R. Chandler. 2011. unmarked: An R package for fitting hierarchical models of wildlife occurrence and abundance. *Journal of Statistical Software* 43:1-23. https://doi.org/10.18637/jss.v043.i10

Foster, L. J., K. M. Dugger, C. A. Hagen, and D. A. Budeau. 2019. Greater Sage-Grouse vital rates after wildfire. *Journal of Wildlife Management* 83:121-134. https://doi.org/10.1002/jwmg.21573 Fremgen, A. L., C. P. Hansen, M. A. Rumble, R. S. Gamo, and J. J. Millspaugh. 2019. Weather conditions and date influence male Sage Grouse attendance rates at leks. *Ibis* 161:35-49. https://doi.org/10.1111/ibi.12598

Gaston, K. J., T. M. Blackburn, and K. K Goldewijk. 2003. Habitat conversion and global avian biodiversity loss. *Proceedings* of the Royal Society B: Biological Sciences 270:1293-1300. https:// doi.org/10.1098/rspb.2002.2303

Gibson, D., E. J. Blomberg, M. T. Atamian, S. P. Espinosa, and J. S. Sedinger. 2018. Effects of power lines on habitat use and demography of Greater Sage-Grouse (*Centrocercus urophasianus*). *Wildlife Monographs* 200:1-41. https://doi.org/10.1002/wmon.1034

Hagen, C. A., J. A. Pitman, T. M. Loughin, B. K. Sandercock, R. J. Robel, and R. D. Applegate. 2011. Impacts of anthropogenic features on habitat use by Lesser Prairie Chickens. Pages 63-76 *in* B. K. Sandercock, K. Martin, and G. Segelbacher, editors. *Ecology, conservation, and management of grouse.* Studies in Avian Biology Vol. 39. University of California Press, Berkeley, California, USA.

Johnson, D. H., and M. M. Rowland. 2007. The utility of lek counts for monitoring Greater Sage-Grouse. Pages 15-23 in K. P. Reese and R. T. Bowyer, editors. *Monitoring populations of Sage-Grouse. Proceedings of a symposium at Idaho State University. Station Bulletin* 88.

Kéry, M., and A. Royle. 2016. *Applied hierarchical modelling in ecology*. Academic, Amsterdam, The Netherlands.

McCaffery, R., J. J. Nowak, and P. M. Lukacs. 2016. Improved analysis of lek count data using N-mixture models. *Journal of Wildlife Management* 80:1011-1021. https://doi.org/10.1002/ jwmg.21094

McNew, L. B., V. L. Winder, J. C. Pitman, and B. K. Sandercock. 2015. Alternative rangeland management strategies and the nesting ecology of Greater Prairie Chickens. *Rangeland Ecology and Management* 68:298-304. https://doi.org/10.1016/j.rama.2015.03.009

Miller, R. F., D. E. Naugle, J. D. Maestas, C. A. Hagen, and G. Hall. 2017. Special Issue: targeted woodland removal to recover at-risk grouse and their sagebrush-steppe and prairie ecosystems. *Rangeland Ecology & Management* 70:1-8. https://doi. org/10.1016/j.rama.2016.10.004

Monroe, A. P., D. R. Edmunds, and C. L. Aldridge. 2016. Effects of lek count protocols on Greater Sage Grouse population trend estimates. *Journal of Wildlife Management* 80:667-678. https://doi.org/10.1002/jwmg.1050

Monroe, A. P., G. T. Wann, C. L. Aldridge, and P. S. Coates. 2019. The importance of simulation assumptions when evaluating detectability in population models. *Ecosphere* 10(7):e02791. https://doi.org/10.1002/ecs2.2791

Nichols, J. D., L. Thomas, and P. B. Conn. 2009. Inferences about landbird abundance from count data: recent advances and future directions. Pages 201-235 *in* D. L. Thomson, E. G. Cooch, and M. J. Conroy, editors. *Modeling demographic processes in marked populations*. Springer-Verlag, New York, New York, USA. https:// doi.org/10.1007/978-0-387-78151-8_9

Palmer, W. E., T. M. Terhune, and D. F. McKenzie, editors. 2011. The National Bobwhite Conservation Initiative: a range-wide plan *for recovering bobwhites.* National Bobwhite Technical Committee Technical Publication, ver. 2.0 Knoxville, Tennessee, USA.

Row, J. R., and B. C. Fedy. 2017. Spatial and temporal variation in the range-wide cyclic dynamics of Greater Sage-Grouse. *Oecologia* 185:687-698. https://doi.org/10.1007/s00442-017-3970-9

Royle, J. A. 2004. N-mixture models for estimating population size from spatially replicated counts. *Biometrics* 60:108-115. https://doi.org/10.1111/j.0006-341X.2004.00142.x

Scott, J. M., D. D. Goble, A. M. Haines, J. A. Wiens, and M. C. Neel. 2010. Conservation-reliant species and the future of conservation. *Conservation Letters* 3:91-97. https://doi.org/10.1111/j.1755-263X.2010.00096.x

Scott, J. M., D. D. Goble, J. A. Wiens, D. S. Wilcove, M. Bean, and T. Male. 2005. Recovery of imperiled species under the Endangered Species Act: the need for a new approach. *Frontiers in Ecology and Environment* 3:383-389. https://doi.org/10.1890/1540-9295 (2005)003[0383:ROISUT]2.0.CO;2

Severson, J. P., C. A. Hagen, J. D. Maestas, D. E. Naugle, J. T. Forbes, and K. P. Reese. 2017*a*. Short-term response of Sage-Grouse nesting to conifer removal in the northern Great Basin. *Rangeland Ecology & Management* 70:50-58. https://doi.org/10.1016/j.rama.2016.07.011

Severson, J. P., C. A. Hagen, J. D. Tack, J. D. Maestas, D. E. Naugle, J. T. Forbes, and K. P. Reese. 2017*b*. Better living through conifer removal: a demographic analysis of Sage-Grouse vital rates. *PLoS ONE* 12:e0174347. https://doi.org/10.1371/journal. pone.0174347

U.S. Fish and Wildlife Service (USFWS). 2015. Endangered and threatened wildlife and plants: 12-month finding on a petition to list Greater Sage-Grouse as endangered or threatened species, proposed rule. *Federal Register* 80:59857-59942.

Van Pelt, W. E., S. Kyle, J. Pitman, D. Klute, G. Beauprez, D. Schoeling, A. Janus, and J. Haufler. 2013. *The Lesser Prairie-Chicken range-wide conservation plan*. Western Association of Fish and Wildlife Agencies, Cheyenne, Wyoming, USA.

Veech, J. A., J. R. Ott, and J. R. Troy. 2016. Intrinsic heterogeneity in detection probability and its effect on *N*-mixture models. *Methods in Ecology and Evolution* 7:1019-1028. https://doi. org/10.1111/2041-210X.12566

Walsh, D. P., G. C. White, T. E. Remington, and D. C. Bowden. 2004. Evaluation of the lek-count index for Greater Sage-Grouse. *Wildlife Society Bulletin* 32:56-68. https://doi.org/10.2193/0091-7648 (2004)32[56:EOTLIF]2.0.CO;2

Wann, G. T., P. S. Coates, B. G. Prochazka, J. P. Severson, A. P. Monroe, and C. L. Aldridge. 2019. Assessing lek attendance of

male Greater Sage-Grouse using fine resolution GPS data: implications for population monitoring of lek mating grouse. *Population Ecology* 61:183-197. https://doi.org/10.1002/1438-390X.1019

Sponsored by the Society of Canadian Ornithologists and Birds Canada

Parrainée par la Société des ornithologistes du Canada et Oiseaux Canada

Figure A1.1 Map of study region, and delineation of the Warners Priority Area for Conservation (PAC) of greater sage-grouse in southern Oregon, USA.

Figure A1.2. Range of lower 85% confidence limits for the year*treatment interaction term in Poisson regressions fit to simulated sage-grouse lek count indices based on the maximum value of three repeated counts at each lek during each year. Each box and whisker represents 400 iterations, which varied according to the specified mean effect size (positive increase in population growth for treatment leks relative to controls) and total number of leks in the sample (equal balance between treatment and control leks). The vertical dashed line indicates the point at which lower confidence limit falls below 0.0, suggesting no support for the effect. Therefore, box and whiskers entirely above the line indicated cases where >90% of iterations produced modeled slope coefficients with 85% confidence intervals that did not overlap 0.0. Each boxplot contains simulations where annual variation in the population growth rate varied from 0.05 to 0.20 (see manuscript Fig. 4).

Figure A1.3. Range of lower 85% confidence limits for the year*treatment interaction term in Poisson regressions fit to simulated sage-grouse lek count indices based on a single randomlyselected count at each lek during each year. Each box and whisker represents 400 iterations, which varied according to the specified mean effect size (positive increase in population growth for treatment leks relative to controls) and total number of leks in the sample (equal balance between treatment and control leks). The vertical dashed line indicates the point at which lower confidence limit falls below 0.0, suggesting no support for the effect. Therefore, box and whiskers entirely above the line indicated cases where >90% of iterations produced modeled slope coefficients with 85% confidence intervals that did not overlap 0.0. Each boxplot contains simulations where annual variation in the population growth rate varied from 0.05 to 0.20 (see manuscript Fig. 4).