Surveys

Volunteer Observer Bias and Sharp-Tailed Grouse Lek Counts in the Upper Midwest

R. Gregory Corace III, Shelby A. Weiss, Lindsey M. Shartell*

R.G. Corace III
Seney National Wildlife Refuge, U.S. Fish and Wildlife Service, 1674 Refuge Entrance Road, Seney, Michigan 49883
Present address: Integrated Land Management, P.O. Box 17, Chassell, Michigan 49916

S.A. Weiss
471 Goethe Avenue, Kirkwood, Missouri 63122

L.M. Shartell*
Minnesota Department of Natural Resources, 1201 East Highway 2, Grand Rapids, Minnesota 55744

Abstract

For most species of prairie grouse Tympanuchus spp. standardized monitoring protocols do not exist across adjoining jurisdictions. For instance, researchers monitor relatively small and disjunct populations of prairie sharp-tailed grouse Tympanuchus phasianellus campestris in Michigan, Wisconsin, and Minnesota using different lek count protocols. Some of these state-led monitoring efforts include the widespread use of volunteers who conduct lek counts without training. To facilitate discussions regarding standardizing sharp-tailed grouse lek count protocols in the Upper Midwest and elsewhere, we quantified the magnitude of differences in data arising from lek counts conducted by paired observers and examined factors potentially associated with observer bias (e.g., training type, use of optical equipment, experience). During 2010–2015, 44 unique pairs of formally trained volunteers in the Upper Peninsula of Michigan performed 93 counts at one lek while controlling for weather, distance to the lek, and the timing of observations. All counts consisted of six 1 min observations over 1 h. At three leks in northeastern Minnesota during 2015, nine pairs (consisting of an individual new to lek counts and a more experienced biologist) performed 21 lek counts using the same protocols, but without formal training for novice observers. Across all years and all sites, our results consistently indicated more agreement in data for dancing birds than nondancing birds. Overall, the mean ± SE difference between paired observers for total birds (dancing and nondancing combined) was 1.05 ± 0.09 in Michigan and 0.37 ± 0.08 in Minnesota. Generalized linear models indicated that the use of optical equipment and training type for dancing birds and training type for nondancing birds were significant factors that improved data agreement in Michigan. In Minnesota, models indicated that experience level (number of surveys performed) of the technician or volunteer did not affect count differences for dancing birds or nondancing birds. We contend that using behavior as a proxy for sex can bias lek count data for sharp-tailed grouse. We suggest lek count protocols in the Upper Midwest be simplified and standardized. Specifically, we suggest that observers report the total number of observed birds (regardless of behavior) and that protocols establish a set duration of observation. We also suggest that the use of optical equipment be required and that training be mandatory.

Keywords: lek count; Michigan; Minnesota; population monitoring; training; volunteer

Received: November 21, 2017; Accepted: September 25, 2018; Published Online Early: October 2018; Published: December 2018

Citation: Corace RG III, Weiss SA, Shartell LM. 2018. Volunteer observer bias and sharp-tailed grouse lek counts in the Upper Midwest. Journal of Fish and Wildlife Management 9(2):666–677; e1944-687X. doi:10.3996/112017-JFWM-095

Copyright: All material appearing in the Journal of Fish and Wildlife Management is in the public domain and may be reproduced or copied without permission unless specifically noted with the copyright symbol ©. Citation of the source, as given above, is requested.

The findings and conclusions in this article are those of the author(s) and do not necessarily represent the views of the U.S. Fish and Wildlife Service.

* Corresponding author: lindsey.shartell@state.mn.us
Introduction

Each spring, male and female Gunnison sage-grouse *Centrocercus minimus*, greater sage-grouse *Centrocercus urophasianus*, greater prairie chicken *Tympanuchus cupido*, lesser prairie chicken *Tympanuchus pallidicinctus*, and sharp-tailed grouse *Tympanuchus phasianellus* spp. gather at specific sites within large grassland and shrub-steppe complexes and other nonforested, terrestrial areas. At these sites (called leks or dancing grounds) courtship and breeding occur, with males conspicuously displaying to females by vocalizing, fighting, and dancing (i.e., shaking tail feathers and stomping their feet; Hoglund and Alatalo 1995). During this time, biologists and (increasingly) volunteers conduct lek counts. When pooled across a broader geographical area, wildlife managers can use lek count data to produce population indices that help guide management (Applegate 2000; Johnson and Rowland 2007). These indices usually require the following assumptions: 1) the number of birds counted is proportional to the total bird population, 2) this proportional value is constant over time, and 3) the detection probability of birds is constant and the same for all observers (Baumgardt et al. 2017).

Standardized lek count protocols exist for most sage-grouse *Centrocercus* spp., but not for many prairie grouse *Tympanuchus* spp. (Gillette et al. 2015). This incongruity is likely the byproduct of the high conservation priority of sage-grouse in the western United States (Connelly et al. 2003; Connelly and Schroeder 2007; USFWS 2015). Although much less imperiled than sage-grouse, the prairie sharp-tailed grouse *T. phasianellus cuprestris* (hereafter, sharptail) is of conservation concern in the Upper Midwest states of Michigan, Wisconsin, and Minnesota, with populations declining significantly over the past half century. Thus, federal and state agencies and interested conservation groups have sought to promote habitat management in each state while also reinvigorating population monitoring efforts for a hunted species. A factor that may explain the lack of protocol standardization for prairie grouse may be the range of environmental conditions inhabited by different populations across the United States, from short-grass prairie to jack pine *Pinus banksiana* barrens (Connelly et al. 1998). For instance, in the northern tier of the Upper Midwest states of Michigan, Wisconsin, and Minnesota, small, disjunct populations of sharptails inhabit a diversity of habitat types, including low-intensity managed farmland, open wetlands, recent forest clear-cuts, and other large (anthropogenic or natural) forest openings (Sample and Mossman 1997; Sjogren and Robinson 1997; Gregg and Niemuth 2000; Niemuth and Boyce 2004; Losey et al. 2007; Figure 1). Consequently, lek count protocols differ by state. For instance, observers use flush counts in Michigan, but in Minnesota counts may be visual (without flushing) or flush counts. In all three states observers count the number of males, females, and birds of unknown sex based on behavioral cues, but the data reported in summary documents differs. Michigan reports the mean number of birds observed per lek and mean number of birds per occupied lek, Wisconsin reports the total number of males, and Minnesota reports the mean number of males and unknowns per occupied lek (see Table S1, Supplemental Material). These differences in protocols make the development of broader (multijurisdictional) population indices nearly impossible. Moreover, if current lek count protocols direct observers to collect data that are not useful for the development of indices relative to the time spent collecting these data or collect data that are inherently biased, these protocols may limit the ability to evaluate the distribution of active leks or may yield data ill-suited for broad-scale population monitoring.

Standardized lek count protocols in the Upper Midwest would supplement ongoing conservation efforts for sharptails, especially if managers understand factors affecting detection probability and bias. Factors influencing detection probability can be associated with the observer(s), the environment, or the organism of interest (Anderson 2001; Walsh et al. 2004). Although few studies have addressed factors affecting bias in sharptail lek counts in the Upper Midwest (Drummer et al. 2011) did quantify lek attendance rates and fidelity in the Upper Peninsula of Michigan in three different landscape contexts (e.g., wetland, forest, and agriculture). They found lek attendance to be strongly associated with date, time of day, and wind speed. They also showed that birds have a high degree of lek fidelity, suggesting that the chance that observers would double-count birds while conducting counts at multiple leks in a given morning was low. However, temporal and spatial dynamics of some leks and the inability of some observers to find leks to monitor may still add potential bias to lek count data (Gillette et al. 2013; Monroe et al. 2016). While wildlife biologists have found some leks in the Upper Midwest to exist in the same location for decades, many leks come and go due to the dynamic nature of fluctuating habitats and populations (Losey et al. 2007).

With the knowledge of some environmental factors associated with the detectability of birds as a pretext, we examined potential lek count data bias that may result from the type of data collected, how data are collected, training, and experience. We controlled for visual obstruction and distance to the lek, and only counted four leks between the study areas to limit impacts of lek variation, such as topology, which also impacts observer bias during lek counts. Previous studies have compared simultaneous counts between human observers and aerial imagery for a number of lekking grouse species (Gillette et al. 2013, 2015) and examined count differences between observers at close proximity to the lek and from a greater distance (Baumgardt et al. 2017), but we know of no study of gallinaceous species that examined differences in simultaneous lek counts conducted by paired observers counting birds from the same fixed position. Thus, our study was largely exploratory, with the goal of quantifying minimum differences between lek counts conducted by different observers and investigating the associated factors influencing these patterns. Three research questions...
guided our work:

1) Using paired observers, what is the magnitude of differences between counts of dancing (proxy for male), nondancing (proxy for female), and total birds?
2) What factors (e.g., training type, use of optical equipment, experience) are associated with differences in the number of birds counted by paired observers?
3) How do paired counts differ between a trained biologist with years of lek count experience and technicians or volunteers new to lek counts?

**Study Sites**

We conducted controlled counts of sharptails for 6 y at one lek within Seney National Wildlife Refuge (NWR) in the Upper Peninsula of Michigan, near the most eastern periphery of sharptail range in the United States, and for 1 y at three leks within the east-central portion of the sharptail range in Minnesota (Connelly et al. 1998; Figure 1). Work in Minnesota only occurred over 1 y, but took advantage of ongoing research that provided further insight into potential biases (see below).

**Michigan**

The Michigan Department of Natural Resources (MI DNR) established a hunting season for sharptails in 1935, began lek counts in 1937, and—after noting a decline in sharptail numbers from lek count data—closed the hunting season in 1998 (Ammann 1957; Luukkonen et al. 2009). Based on lek count data, the agency reopened the hunting season in 2010. Current monitoring involves occupancy surveys along section lines (Luukkonen et al. 2009) and lek counts (MI DNR unpublished protocol; see Table S1, Supplementary Material).

The Great Lakes influence the climate of Seney NWR. Lake Michigan and Lake Superior, both < 50 km away, moderate temperatures. Lake-effect snow and rain are common, and precipitation is evenly distributed over the year. For a detailed description of the climate, bedrock geology, landforms, soils, presettlement and present-day vegetation of the Seney NWR area, see Albert (1995). The overall Seney NWR matrix consists of wetlands and conifer forests. The lek used in this study is located within an old farm field on a drained wetland with large forest openings nearby. The study site (46°18’52.15” N, 86°4’56.83”W) is maintained as sharptail habitat through a combination of prescribed fire and mowing every 5–10 y. The lek itself consists of a series of four raised and interconnected mounds approximately 2 m above the level field. Each mound is approximately 10 m in length, with the total area on which birds congregate approximately 40 m in length and 10 m in depth. Sharptails have used this specific lek for decades (Losey et al. 2007). No woody vegetation that would obstruct the views of observers is within 30 m of the lek. Because lek counts throughout the Upper Midwest occur during or shortly after snowmelt, nearly all herbaceous vegetation is still dormant and lying flat against the ground. (We recognize...
that visual obstruction is a common factor influencing detection of grouse attending leks, especially in the Intermountain West, and for this reason alone our data represent minimum estimates of observer bias.)

Minnesota

The Minnesota Department of Natural Resources (MN DNR) has used the same lek count protocols annually since 1976 in two regions of the state, the northwest and the east-central (Roy 2016). Lek counts are conducted primarily by MN DNR biologists, but also include volunteers. Data have shown that the trend for the mean number of sharptails (males and unknowns) per lek over the past decade has been negative, declining 15% from 11.2 sharptails per lek in 2004 to 9.5 sharptails per lek in 2016 (Roy 2016). In addition, the number of active leks has declined in the east-central portion of the sharptail range in the state. Despite some indications of a reduced sharptail population, hunting remains open across much of the range of the species in Minnesota and sharptail monitoring is of high priority (Roy 2016).

The climate in the east-central portions of Minnesota is continental. Temperature extremes of −43°C have been noted and snowfall of 122–172 cm is common (Albert 1995). The three leks in east-central Minnesota are on private lands primarily for hay production, cattle grazing, and cropland, leaving behind little to no vegetation during the lekking season. All three leks exist on flat land, with no woody vegetation present; observer views are unobstructed and similar for each. The surrounding landscape contains a mix of agricultural lands, brushlands, open wetlands, and some mixed forests. Surrounding nonagricultural lands are maintained in an early successional stage through prescribed fire, mowing, and shearing.

Methods

Michigan

Beginning in 2010, Seney NWR recruited volunteers to conduct controlled counts of sharptails at one lek for the following reasons: 1) to aid the refuge in collecting data for a conservation priority species and 2) as an outreach opportunity for the general public during which the goals and objectives of sharptail habitat management could be discussed. Seney NWR staff also considered this a novel opportunity to illustrate complexities associated with wildlife sampling; while many volunteers have interest in the size of wildlife populations, too few biologists point out potential biases and errors associated with the numbers they report. Staff designed controlled counts of sharptails at Seney NWR, in part, to evaluate and quantify potential bias inherent in a fairly simple count of a bird species of high public interest. Although a small number of volunteers had a background in biology and related disciplines, virtually all volunteers had little to no experience with lek count protocols.

Other than the fact that observers never purposely flushed birds from the lek and counted them, our count protocols closely mirrored those of the MI DNR (see Table S1, Supplementary Material). Because factors such as date, time, and weather can reduce lek activity, we used recommendations from the eastern Upper Peninsula lek attendance study of Drummer et al. (2011) to further refine the MI DNR lek count protocols and to control for potential bias resulting from environmental variables. In other words, we conducted counts for this study with more limitations than standard MI DNR protocols dictate, another reason why our data represent minimum estimates of observer bias. Namely, counts at Seney NWR occurred at one lek and only during nearly ideal dates, time, and weather. All counts were conducted from 1 April to 14 May 2010–2015, within 1.5 h after official sunrise for a given date, with winds < 32 km/h, and under conditions lacking persistent precipitation. To further reduce bias that results from the inability of observers to find leks (Gillette et al. 2013), all observations were made from a 2 × 2 m wooden blind located 45 m from the one lek. Visual obstruction by vegetation was minimal and similar for each observer (see above).

Paired observers independently and simultaneously conducted six 1-min counts over a period of 1 h on a given morning. For the purpose of this study, we refer to each of the six counts as an “individual count” and the average of the six counts as an “averaged count.” For many lekking bird species, repeated counts are a central feature of lek count protocols (Connelly et al. 2003; Connelly and Schroeder 2007). We conducted repeated counts for 1 h to evaluate how classifying birds as dancing or nondancing, a common metric collected in lek counts in the Upper Midwest, influences bias. By limiting our observations in Michigan to one lek we controlled for many other potential sources of bias that are common to broad-scale monitoring programs devoted to lekking bird species.

Observers arrived at the blind approximately 30 min before sunrise and recorded weather data. If wind speed was > 32 km/h or rain or snow was persistent, they did not perform the count. Barring these conditions, volunteers then recorded the number of dancing birds and the number of nondancing birds during the same six 1-min count intervals each separated by a 9-min rest period (six counts per 1 h). We used dancing and nondancing as a proxy for sex in this study, with dancing birds assumed to be males. While we acknowledge that solely using dancing to infer sex of a bird is problematic for a number of reasons (Jenni and Hartzler 1978; Hoglund and Alatalo 1995), we were unable to provide a primer to volunteers on all sharptail behaviors that may occur on the lek. Moreover, because all current state-led protocols in the Upper Midwest use dancing/nondancing birds as a proxy, and no formal training of observers has been done in the past as far as we are aware in Michigan or Minnesota, we believed training volunteers on other
behavioral cues indicating sex would be problematic and confound our study. This study therefore allowed us an imperfect evaluation of the potential bias caused by counting dancing birds as male and nondancing birds as female as is typical in lek count protocols in the Upper Midwest.

During each count at Seney NWR, we instructed volunteers to use separate data sheets and to not compare observations. We provided volunteers with binoculars, but they were not required to use them. Some volunteers brought their own binoculars and/or spotting scopes. Regardless, we instructed volunteers to indicate whether or not they used optical equipment for counts. We also provided volunteers with datasheets, a sunrise chart, handheld weather meter, maps, compass, stopwatch, writing implements, and copies of instructional materials. We excluded counts from analysis if we suspected that volunteers did not follow protocols correctly or if they observed no birds.

Volunteers performed lek counts as “unique pairs.” We defined a unique pair as a unique combination of two individuals who conducted at least one count together. If any member of the unique pair conducted an averaged count with a different individual, that was treated as a new unique pairing. Refuge staff paired the volunteers or they signed up with someone they already knew. We publicized the opportunity for volunteers to participate in counts widely each year via press releases, newsletter articles, the Seney NWR website, and on social media.

Seney NWR staff led in-person training for the first 4 y (2010–2013). Volunteers were required to attend one of several training sessions offered. Each training session lasted for approximately 1 h, during which participants received detailed instructions from one refuge staff on protocols and how to use the provided equipment and datasheets. In the final 2 y (2014–2015), we delivered the same information to volunteers via a three-part training video, and did no in-person training. Under this method, volunteers took a quiz (15 questions) of the information presented in a 36-min video. Volunteers were able to view the video as many times as necessary to understand the protocol.

Minnesota

Minnesota DNR staff, temporary technicians, and volunteers conducted paired lek counts between 27 March and 15 May 2015. Pairs included the following three possible combinations with related experience: 1) a MN DNR staff with 5 y of experience and a temporary technician new to lek counts, 2) same MN DNR staff and a volunteer new to lek counts, and 3) a temporary technician new to lek counts and a volunteer new to lek counts. Lek count protocols were the same as for Seney NWR (see above). Pairs conducted lek counts from a temporary or permanent blind approximately 10 m from the center of each of three leks. Staff provided no formal training to volunteers or technicians. Instead, MN DNR staff or technicians provided brief instructions prior to the first lek count. We noted experience level (i.e., number of surveys performed in 2015) of each technician or volunteer and analyzed differences in data comparing the results of participants of varying levels of experience.

Paired observers conducted lek counts at the same time as unrelated research involving intensive trapping, sexing, and radio-collaring of sharptails occurred (Shartell 2018). The unrelated research efforts involved daily monitoring of leks from 1 h before sunrise until birds left the lek (i.e., approximately 3–4 h of continuous observation for a given lek per day). This extended observation time (relative to only 6 min of individual counts done during lek counts for this study) provided more precise knowledge of male and female attendance at a given lek. This enhanced knowledge allowed for the comparison of short (1-min) observations based on bird dancing behavior to observations based on any and all bird behaviors over a longer time frame. We hereafter refer to estimates of the number of male and female birds present at the lek derived from all observations made during trapping, sexing, and radio-collaring as a “verified count.” Again, we define a verified count in our Minnesota data as one in which the number and sex of all observed birds is known due to the intensive trapping, sexing, and radio-collaring efforts and observations unrelated to our lek count protocols. We used the comparison of individual counts using the above protocol with the verified count to further assess the error associated with using dancing as a proxy for males.

Data Analysis

We analyzed count data in two ways. Depending on the question, the metric of interest was either the absolute difference (positive value) in the number of dancing, nondancing, or total (combined) birds between members of a pair for an individual count or the mean absolute difference (positive value) in number of dancing, nondancing, or total (combined) birds for all six counts conducted in a morning (i.e., averaged count). This methodology was consistent across both study areas (the one lek at Seney NWR in Michigan and the three leks in Minnesota). Complete agreement in data for an individual count or for a given averaged count yielded a “0” difference value between paired volunteers. We used two-sample Student’s t-tests to investigate mean differences for averaged counts for dancing and nondancing birds relative to the individual variables training type and optical equipment used. We performed these analyses in SigmaPlot (2012). Count data were not normally distributed and exhibited a moderate skew of <−0.5 or >0.5, so we conducted a square-root transformation prior to t-tests. Transformed data more closely fit a normal distribution with a skew of −0.5 to 0.5 (Osborne 2010).

We used generalized linear models with fixed effects and a Poisson distribution to assess factors that influenced count difference. However, because of our
generalized linear models choice we were bound by certain assumptions, such as our response or dependent variable (count difference) had a Poisson distribution and can be modeled in an unbiased fashion by our independent variables (e.g., training type, optical equipment, etc.). Using the global model, we compared all possible models using Akaike’s Information Criteria corrected for small sample size (AICc) to identify the top model. Independent variables for Seney NWR data included training type (0\(\equiv\)in-person, 1\(\equiv\)video), use of optical equipment by both observers (0 or 1), and number of surveys performed by the unique pair. To assess the effect of training type (in-person or video), we used data in which both members of the pair had the same type of training and only that type of training. We also removed data in which observers did not indicate if they used optical equipment. Thus, the sample size for Michigan regressions was \(n = 408\). The only independent variable included for Minnesota data was number of surveys performed, as we did not conduct training and observers used optical equipment in all but one case. We used only those individual counts paired with the MN DNR biologist (\(n = 102\)), and the number of surveys conducted by the technician or volunteer was the independent variable. We performed analyses in R (R Core Team 2013).

### Results

#### Michigan

**Volunteer retention.** Over 6 y of conducting counts at Seney NWR, 44 unique pairs involving 79 observers performed 93 averaged counts and 558 individual counts (93 averaged counts \(\times\) six individual counts; Data S1, Supplemental Material). We excluded only 6% of all counts from analyses due to problems associated with following protocols (e.g., not arriving at lek at the required time, stop counting early because no birds observed, data missing, etc.). Most unique pairs (75%) only collected data for 1 y (Table 1). The mean \(\pm\) 1 SE number of years a unique pair collected data was 1.43 \(\pm\) 0.11. A unique pair conducted a mean of 2.11 \(\pm\) 0.24 averaged counts together (median and mode = 1).

#### Magnitude of count differences.** The maximum number of total birds (dancing and nondancing combined) observed during any averaged count ranged from 13 to 27 across all 6 y. The maximum number of birds observed in a given year declined over time by more than 50%. The range of differences between paired observers for individual counts (regardless of whether dancing or nondancing) was 0–10, with zero differences (i.e., complete data agreement) occurring 340 times (61% of individual counts) for dancing birds, 262 times (47%) for nondancing birds, and 271 times (49%) for total birds (Figure 2). Overall, and consistently throughout our data set, there was greater data agreement for dancing birds compared to nondancing birds (Figure 3). The mean difference in dancing birds between observers for averaged counts (\(n = 93\)) was 0.75 \(\pm\) 0.07, 1.04 \(\pm\) 0.09 for nondancing birds, and 1.05 \(\pm\) 0.09 for total birds. The plot of the mean difference in averaged counts for dancing and nondancing birds by year showed a sharp decline from 2011 to 2014 (Figure 3), which mirrored the pattern observed in maximum number of birds observed each year.

### Table 1. Prairie sharp-tailed grouse *Tympanuchus phasianellus campestris* volunteer lek count effort and retention at Seney National Wildlife Refuge (2010–2015). We defined a unique pair as a unique combination of two individuals who conducted at least one averaged lek count together. Lek counts that did not follow protocols are not included.

| Year | No. of averaged lek counts completed | No. of pairs unique to a given year | No. of pairs unique to a given year plus returned pairs | % of pairs in a given year unique to that year |
|------|-------------------------------------|-------------------------------------|--------------------------------------------------------|-----------------------------------------------|
| 2010 | 8                                   | 5                                   | 5                                                      | 100                                           |
| 2011 | 13                                  | 8                                   | 9                                                      | 89                                            |
| 2012 | 18                                  | 9                                   | 14                                                     | 64                                            |
| 2013 | 15                                  | 6                                   | 10                                                     | 60                                            |
| 2014 | 14                                  | 8                                   | 10                                                     | 80                                            |
| 2015 | 25                                  | 8                                   | 16                                                     | 50                                            |
| Mean \(\pm\) 1 SE | 15.50 \(\pm\) 2.32 | 7.33 \(\pm\) 0.61 | 10.67 \(\pm\) 1.58 | 73.83 \(\pm\) 7.76 |

#### Figure 2. Proportion of all individual lek counts (\(n = 558\)) with difference values of 0–4+ for dancing prairie sharp-tailed grouse *Tympanuchus phasianellus campestris*, nondancing birds, and total birds at Seney National Wildlife Refuge (2010–2015).
Factors affecting count differences. Volunteers that received in-person training obtained identical counts in 10 of 54 averaged counts of dancing birds (19%) and six of 54 averaged counts of nondancing birds (11%). For those who received training via video (n = 31), 10 averaged counts (32%) had zero differences for dancing birds and eight averaged counts (26%) had zero differences for nondancing birds. Data derived from volunteers who received in-person training showed greater mean differences compared to those who were trained by video (Figure 4). More specifically, data for dancing birds showed more agreement (df = 83, t = 2.21, P = 0.03) for video training compared to in-person training. Data for nondancing birds were likewise more congruent from those who received video training (df = 83, t = 1.72, P = 0.09).

When no optical equipment was used, we observed greater differences in data for nondancing birds (df = 89, t = −2.18, P = 0.03) than dancing birds (df = 89, t = −1.73, P = 0.09; Figure 5). Of the 76 averaged counts conducted using optical equipment, 18 averaged counts (24%) had zero differences for dancing birds and 13 averaged counts (17%) had zero differences for nondancing birds. Of the 15 averaged counts conducted without optical equipment, one averaged count had zero differences for dancing and nondancing birds each.

For differences in the counts of dancing birds, the full model was best supported. The top model indicated that the use of optical equipment and training type influenced agreement in individual counts of dancing birds, resulting in greater agreement overall. Experience (as indexed by number of surveys completed by a unique pair) had a moderate effect, with greater experience tending to result in more agreement in counts (Table 2). For nondancing birds, the model including use of optical equipment and training type was best supported (Tables 2–4).

Minnesota

Magnitude of count differences. In Minnesota, 21 averaged counts consisting of 126 individual counts were completed in 2015. Maximum number of birds (combined dancing and nondancing) observed over all averaged counts ranged from 7 to 15. During 17 averaged counts (81%), the pairing consisted of a MN DNR biologist and a technician or volunteer new that year to lek counts. The range of differences in individual counts for dancing and nondancing birds separately was 0–14. For 126 individual counts, zero difference occurred 59 times (47%) for dancing birds and 49 times (39%) for nondancing birds. The range of differences in individual counts of total birds (dancing and nondancing combined) was 0–4, and complete agreement occurred 91 times (72%) for total birds present. The mean difference
for averaged counts was $1.66 \pm 0.27$ for dancing birds, $1.75 \pm 0.25$ for nondancing birds, and $0.37 \pm 0.08$ for total birds.

Factors affecting count differences. Models indicated that the number of surveys performed (experience level) by the technician or volunteer did not affect count differences for dancing birds (estimate $= 0.01 \pm 0.03, P = 0.68$) or nondancing birds (estimate $= 0.04 \pm 0.02, P = 0.11$). Where differences in counts occurred between the MN DNR biologist and the other observer, the MN DNR biologist counted more dancing birds 43 times and fewer dancing birds 14 times. Conversely, the biologist was almost twice as likely to count fewer nondancing birds (41 times this occurred) than more nondancing birds (23 times this occurred).

Error associated with using dancing as a proxy for sex. For 252 individual counts, complete agreement between number of dancing birds and the number of males from verified counts occurred 85 times (34%). In all cases of disagreement, the number of dancing birds counted was lower than the number of verified males. The mean difference between individual counts of dancing birds and verified males was $3.28 \pm 0.20$, and differences ranged from 0 to 14 birds. However, when using the maximum number of dancing birds from each observer during any of the six individual counts in an averaged count ($n = 42$), complete agreement between the maximum number of dancing birds and the number of verified males occurred 24 times (57%). The mean difference was $0.95 \pm 0.23$, and ranged from 0 to 5 birds. Observations made by the MN DNR biologist ($n = 17$) had a mean difference between maximum number of dancing birds and verified males of $0.47 \pm 0.17$, while technicians and volunteers had a mean difference of $1.28 \pm 0.35$.

### Discussion

While researchers have evaluated the double observer method for songbirds monitored by point-counts (Nichols et al. 2000; Forcey et al. 2006), our study is the first we are aware of that evaluated data resulting from two human observers simultaneously conducting lek counts using the same protocol and viewing location. Biases associated with lek counts are numerous and may include observer experience, observer visual acuity, observer memory and fatigue, training, counting effort, the obstruction of views, weather, and distance to the lek (Bibby et al. 1992; Anderson 2001; Drummer et al. 2011). In this study we blocked for environmental factors that may influence detectability (Drummer et al. 2011) and potential bias resulting from inability to find leks (Gillette et al. 2013). We then quantified the magnitude of differences in lek counts between volunteers and examined factors potentially associated with observer bias (e.g., training type, use of optical equipment, experience). Trained volunteers at Seney NWR had the same individual count for the total number of birds 49% of the time over 6 y and had a mean difference of approximately one bird (dancing and nondancing combined). These findings illustrate greater data agreement than was found in 25 simultaneous counts of sharptails in Idaho using ground-based and aerial infrared methods (Gillette et al. 2015). Working with leks that had approximately twice the number of birds as

### Table 2

| Sex          | Variable                             | Estimate | Standard error | $P$ value |
|--------------|--------------------------------------|----------|----------------|-----------|
| Dancing birds| Count difference ~ optical equipment used + training + surveys performed |          |                |           |
|              | Optical equipment used               | $-0.40$  | $0.14$         | $< 0.01$  |
|              | Training                             | $-0.45$  | $0.12$         | $< 0.01$  |
|              | Surveys performed                    | $-0.11$  | $0.06$         | $0.05$    |
| Nondancing   | Count difference ~ optical equipment used + training |          |                |           |
|              | Optical equipment used               | $-0.19$  | $0.10$         | $0.10$    |
|              | Training                             | $-0.25$  | $0.11$         | $0.02$    |

### Table 3

Comparison and ranking of candidate generalized linear models predicting the differences in individual counts of dancing prairie sharp-tailed grouse *Tympanuchus phasianellus campestris* at Seney National Wildlife Refuge (2010–2015). Models used a Poisson distribution. Model predictors, number of parameters (df), Akaike’s Information Criterion (corrected for smaller sample sizes, AIC$_C$), AIC$_C$ rescaled from the lowest score ($\Delta$ AIC$_C$), and Akaiake weights (w.AIC$_C$) for each model are listed.

| Predictors                                           | df | AIC$_C$ | $\Delta$ AIC$_C$ | w.AIC$_C$ |
|------------------------------------------------------|----|---------|------------------|-----------|
| Optical equipment + training type + surveys performed | 4  | 1,035.3 | 0.00             | 0.65      |
| Optical equipment + training type                     | 3  | 1,037.1 | 1.83             | 0.26      |
| Training type                                         | 2  | 1,040.3 | 5.03             | 0.05      |
| Training type + surveys performed                     | 3  | 1,041.3 | 6.04             | 0.03      |
| Optical equipment + surveys performed                 | 3  | 1,046.8 | 11.54            | $< 0.01$  |
| Optical equipment                                     | 2  | 1,047.1 | 11.77            | $< 0.01$  |
| Intercept only                                        | 1  | 1,049.3 | 14.00            | $< 0.01$  |
| Surveys performed                                     | 2  | 1,050.8 | 15.54            | $< 0.01$  |
found at Seney NWR, observers in Idaho recorded identical counts of the total number of birds in 32% of the sample. Relative to the mean difference of one bird found at Seney, researchers found a mean difference of two birds across all simultaneous counts in Idaho. The smaller mean observed maximum number of birds on a lek and the fact that more environmental and sampling variables were controlled for at Seney NWR could partially explain the lower count difference at Seney NWR. A number of studies have concluded that while volunteers can collect quality data when properly trained, personal limitations should be assessed quantitatively so that the best fit occurs between volunteer and type of data collected (Fore et al. 2001; Crall et al. 2011; Kremen et al. 2011). We conducted detailed training (both in-person and video) only at Seney NWR and not in Minnesota. Lek count data collected by volunteers who had video training at Seney NWR had more agreement compared with data collected by those who received in-person training. While in-person training is typically thought to be a more effective training type in some environments, the value of video-driven training cannot be discounted (Pang 2009; Starr et al. 2014). Providing training that is consistent and limits ambiguity of the task is important to decrease bias. Video-based training may be a particularly useful tool for training volunteers to conduct sharptail lek counts since volunteers can revisit areas of uncertainty in the protocols whenever needed. Our data from Seney NWR provides further support for the inclusion of optical equipment in lek count protocols. In the western United States, concerns regarding the distribution and abundance of greater sage-grouse have prompted studies quantifying factors that affect detectability (Fremgen et al. 2016; Baumgardt et al. 2017). Sage-grouse lek counts are typically conducted using lek routes surveyed multiple times in a season. Observers often count male birds from 100–m using optical equipment. Baumgardt et al. (2017) reported that the use of spotting scopes to count sage-grouse when vegetation is either low or sparse may offset any effects of viewing from long distances. We suggest the same may be true in the Upper Midwest, where most lek counts are conducted during snowmelt and when vegetation is still dormant. We found limited evidence that experience, by itself, improved count agreement after controlling for other variables that research has demonstrated to introduce bias. Similarly, Baumgardt et al. (2017) found that observer experience did not affect visibility bias for sage-grouse. However, because retention of volunteers comprising unique pairs at Seney NWR was low, we were unable to follow changes in differences for a unique pair over time. We did, however, find more congruent data between volunteers that are both new to lek protocols (Seney NWR data) than between a trained biologist and those new to lek counts (Minnesota data). We suggest this may indicate differing skill levels cause different interpretations of observations and thus influence lek counts. The experienced biologist may draw upon more knowledge and experience to identify dancing and nondancing birds, or may be biased by knowing the correct sex of the bird, and thus whether it should be expected to be dancing or not. Data for the total number of birds had greater agreement than data for dancing and nondancing birds in the Minnesota data set, but not so in the Seney NWR data set (volunteer-only data). We suggest that too much of a focus on the behaviors of “dancing” or “nondancing” during each of the six counts made during an averaged count may have influenced differences in Minnesota. Paired observers likely did not count the same birds at the same time during 1-min observations, so they could accurately place one bird in two different classifications when it exhibited different behavior at different times. Ambiguity regarding classification can therefore lead to bias, particularly with those that have less experience and training. This was noted in a past study involving volunteers that compared surveys of white oak Quercus alba stands between students and professionals. Data suggested that there was less agreement when measuring characteristics that were more subjective in nature, such as crown shape (Galloway et al. 2006). For this reason, it is important to establish protocols that minimize as many sources of bias as possible, especially when recruiting volunteers. Tasks involving less subjectivity likely yield measurements that are more precise. There were a number of limitations to our study other than those described above. First, because we did not conduct flush counts, our estimates of total number of birds were possibly inaccurate as suggested by Gillette et

### Table 4
Comparison and ranking of candidate generalized linear models predicting the differences in individual counts of nondancing prairie sharp-tailed grouse Tympanuchus phasianellus campestris at Seney National Wildlife Refuge (2010–2015). Models used a Poisson distribution. Model predictors, number of parameters (df), Akaike’s Information Criterion (corrected for smaller sample sizes, AICc), AICc rescaled from the lowest score (Δ AICc), and Akaikes weights (w AICc) for each model are listed.

| Predictors | df  | AICc  | Δ AICc  | w AICc |
|------------|-----|-------|---------|--------|
| Optical equipment + training type | 3   | 1,255.6 | 0.00    | 0.36   |
| Training type | 2   | 1,256.2 | 0.62    | 0.27   |
| Optical equipment + training type + surveys performed | 4   | 1,257.7 | 2.02    | 0.13   |
| Training type + surveys performed | 3   | 1,258.0 | 2.41    | 0.11   |
| Optical equipment | 2   | 1,259.6 | 3.98    | 0.05   |
| Intercept only | 1   | 1,259.7 | 4.11    | 0.05   |
| Surveys performed | 2   | 1,261.2 | 5.60    | 0.02   |
| Optical equipment + surveys performed | 3   | 1,261.6 | 5.96    | 0.02   |
al. (2015). Comparing counts of flushed sharptails between volunteers and volunteers/technicians and a trained biologist may have provided more insight. Second, taking 1 h to conduct an averaged count possibly led to observer fatigue and may have contributed to differences in counts independent of training. However, as most of our observers were motivated volunteers new to lek counts and had never before observed birds, we suggest fatigue was unlikely. Third, while we used fixed blinds in our lek counts, most lek counts are conducted from a vehicle and often at a greater distance to the lek. This may make some comparisons with other studies of observer bias potentially difficult. Finally, because we chose to analyze data using a generalized linear model and not an N-mixture model, factors that may have driven our response variable (count difference) may have been conflated with other factors and potentially led to spurious conclusions. However, because we used few input variables and none were complex environmental variables, we suggest that the use of an N-mixture model would not necessarily impact our findings or conclusions that seem to be well supported by other literature we have cited.

We suggest lek count protocols in the Upper Midwest be simplified and standardized. Specifically, we suggest that observers make no attempt at determining sex of birds using behavior as a proxy. Instead, we suggest that they report the total number of observed birds, regardless of behaviors, and that the mean number of birds per active lek (based on ≥2 lek counts) be the primary index used. While not a focus of this study, we also suggest that states quantify the annual effort expended on sampling and searching for leks and report the number and proportion of leks sampled that are active. Further, we suggest that agencies refine lek survey protocols by adding a fixed time period (e.g., 15 min) during which multiple observations can be made by observers using optical equipment. Observers can then record the maximum value across these observations. We also advise use of standardized, recorded instructional training methods. Finally, we suggest that existing data be analyzed by each state agency to more precisely determine when lek counts are conducted. As suggested by past studies (Drummer et al. 2011; Gillette et al. 2015), more attention to lek attendance rates may yield better data overall. In Minnesota, peak hen attendance during 2014 and 2015 occurred from mid-April to May; however, late snowpack in 2013 resulted in peak hen attendance from early to mid-May; male attendance was not significantly related to date (Shartell 2018). Because the timing of peak hen attendance can vary, we recommend conducting surveys prior to hen attendance, for example from early to mid-April in Minnesota. We contend that these refinements would reduce unnecessary noise in sharptail lek count data and yield data with greater precision.

The involvement of volunteers in natural resource research, inventory, and monitoring has increased in recent years due to declining budgets and fewer personnel in many agencies and organizations (Silver-town 2009; Hochachka et al. 2012; Bonney et al. 2014). The inclusion of volunteers in data collection provides outreach opportunities for wildlife scientists and managers and their associated projects and institutions (Dickinson et al. 2010). This outreach is of particular value for many lekking gallinaceous bird species that are experiencing long-term population declines (Silvy and Hagen 2004). In appropriate settings, wildlife managers can view sharptails as a flagship species that through hunting and wildlife viewing garners attention for related regional conservation issues. Continued volunteer-based monitoring may allow for more public outreach for management actions for sharptails and conspecifics (Corace et al. 2009, 2016). The fact that we were consistently recruiting new unique pairs at Seney NWR suggested a continued interest. Likewise in Minnesota, there is interest and opportunity to increase the use of volunteers in wildlife surveys. This may be of special importance for sharptail conservation in Minnesota as the majority of known leks are on private lands and habitat management, such as shearing and prescribed fire, require the buy-in of landowners and other citizens who make up the volunteer pool.

Supplemental Material

Please note: The Journal of Fish and Wildlife Management is not responsible for the content or functionality of any supplemental material. Queries should be directed to the corresponding author for the article.

Data S1. Prairie sharp-tailed grouse Tympanuchus phasianellus campestris lek count data from one lek at Seney National Wildlife Refuge, Upper Peninsula of Michigan, 2010–2015.

Found at DOI: https://doi.org/10.3996/112017-JFWM-095.S1 (23 KB XLSX); also available at https://ecos.fws.gov/ServCat/Reference/Profile/87080.

Data S2. Prairie sharp-tailed grouse Tympanuchus phasianellus campestris lek count data from three leks in east-central Minnesota, 2015.

Found at DOI: https://doi.org/10.3996/112017-JFWM-095.S2 (113 KB XLSX); also available at https://ecos.fws.gov/ServCat/Reference/Profile/87080.

Table S1. Summary of prairie sharp-tailed grouse Tympanuchus phasianellus campestris lek count protocols in Michigan, Wisconsin, and Minnesota, USA. Lek counts are led by state natural resource agencies. Protocols in each state have counts lasting for an unspecified time, conducted from an unspecified distance to lek, and recommend that optical equipment be used. All protocols are unpublished.

Found at DOI: https://doi.org/10.3996/112017-JFWM-095.S3 (14 KB DOCX).

Reference S1. Albert DA. 1995. Regional landscape ecosystems of Michigan, Minnesota, and Wisconsin: a
working map and classification. St. Paul, Minnesota: USDA Forest Service. GTR-NC-178.
Found at DOI: https://doi.org/10.3996/112017-JFWM-095.54 (60 MB PDF); also available at https://www.nrs.fs.fed.us/pubs/gtr/gtr_nc178.pdf.

Reference S2. Roy C. 2016. 2016 Minnesota spring grouse surveys. Grand Rapids, Minnesota: Minnesota Department of Natural Resources.
Found at DOI: https://doi.org/10.3996/112017-JFWM-095.55 (386 KB PDF); also available at https://files.dnr.state.mn.us/recreation///grouse_survey_report16.pdf.

Reference S3. Shartell, LM. 2018. Survival, nest success, and habitat selection of sharp-tailed grouse in east-central Minnesota—final project report. Grand Rapids, Minnesota: Minnesota Department of Natural Resources.
Found at DOI: https://doi.org/10.3996/112017-JFWM-095-56 (534 KB PDF).

Acknowledgments
This work was funded by Seney NWR, the Seney Natural History Association, and the MN DNR. To all volunteers at Seney NWR, thank you. S. Giles, J. McDonough, and G. McClellan facilitated the surveys at Seney NWR and D. Marsh assisted with data organization. M. Witecha, C. Pollentier, and B. Dhuey provided information on lek count protocols in Wisconsin. We are also indebted to the useful edits and comments of three reviewers and the Associate Editor. Their suggestions greatly improved this paper.

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

References
Albert DA. 1995. Regional landscape ecosystems of Michigan, Minnesota, and Wisconsin: a working map and classification. St. Paul, Minnesota: USDA Forest Service. GTR-NC-178 (see Supplemental Material, Reference S1).
Ammann GA. 1957. The prairie grouse of Michigan. Lansing, Michigan: Michigan Department of Conservation Technical Bulletin.
Anderson DR. 2001. The need to get the basics right in wildlife field studies. Wildlife Society Bulletin 29:1294–1297.
Applegate RD. 2000. Use and misuse of prairie chicken lek surveys. Wildlife Society Bulletin 28:457–463.
Baumgardt JA, Reese KP, Connelly JW, Garton EO. 2017. Visibility bias for sage-grouse lek counts. Wildlife Society Bulletin 41:461–470.
Bibby CJ, Burgess ND, Hill DA. 1992. Bird census techniques. London: Academic Press.
Bonnev R, Shirkl J, Phillips TB, Wiggins A, Ballard HL, Miller-Rushing AJ, Parrish JK. 2014. Next steps for citizen science. Science 343:1436–1437.
Cartier AT, Baldy JJ, Brenneman JM. 2011. The second Michigan breeding bird atlas, 2002–2008. Kalamazoo, Michigan: Kalamazoo Nature Center.
Connelly JW, Gratson MW, Reese KP. 1998. Sharp-tailed grouse (Tympanuchus phasianellus), version 2.0. In Pools AF, Gill FB, editors. The birds of North America. Ithaca, New York: Cornell Lab of Ornithology. Available: https://birdsna.org/Species-Account/bna/species/shhtgro/introduction (September 2018).
Connelly JW, Reese KP, Schroeder MA. 2003. Monitoring of greater sage-grouse habitats and populations. Moscow, Idaho: University of Idaho. Idaho Forest, Wildlife and Range Experimental Station Bulletin 80. Available: https://sagemap.wr.usgs.gov/docs/grouse_habitat_book.pdf (September 2018).
Connelly JW, Schroeder MA. 2007. Historical and current approaches to monitoring greater sage-grouse. Pages 3–9 in Reese KP, Bowyer RT, editors. Monitoring populations of sage-grouse. Moscow, Idaho: Proceedings of a Symposium at Idaho State University.
Corace RG III, Flaspohler DJ, Shartell LM. 2009. Geographical patterns in openland cover and hayfield mowing in the Upper Great Lakes region: implications for grassland bird conservation. Landscape Ecology 24:309–323.
Corace RG III, Korte JL, Shartell LM, Kashian DM. 2016. Upland sandpiper: a flagship for jack pine barrens restoration in the Upper Midwest? Ecological Restoration 34:49–60.
Crall AW, Newman GJ, Stohlgren TJ, Holfelder KA, Graham J, Waller DM. 2011. Assessing citizen science data quality: an invasive species case study. Conservation Letters 4:433–442.
Dickinson JL, Zuckerberg B, Bonter DN. 2010. Citizen science as an ecological research tool: challenges and benefits. Annual Review of Ecology, Evolution, and Systematics 41:149–172.
Drummer TD, Corace RG III, Sjogren SJ. 2011. Sharp-tailed grouse lek attendance and fidelity in Upper Michigan. Journal of Wildlife Management 75:311–318.
Forcey GM, Anderson JT, Ammer FK, Whitmore RC. 2006. Comparison of two double-observer point-count approaches for estimating breeding bird abundance. Journal of Wildlife Management 70:1674–1681.
Fore LS, Paulsen K, O’Laughlin K. 2001. Assessing the performance of volunteers in monitoring streams. Freshwater Biology 46:109–123.
Fremgen AL, Hansen CP, Rumble MA, Gamo RS, Millspaugh JJ. 2016. Male greater sage-grouse detectability on leks. Journal of Wildlife Management 80:266–274.
Galloway AW, Tudor MT, Haegen WMV. 2006. The reliability of citizen science: a case study of Oregon white oak stand surveys. Wildlife Society Bulletin 34:1425–1429.
Gillette GL, Coates PS, Petersen S, Romero JP. 2013. Can reliable sage-grouse lek counts be obtained using...
aerial infrared technology? Journal of Fish and Wildlife Management 4:386–394.
Gillette GL, Reese KP, Connelly JW, Colt CJ, Knetter JM. 2015. Evaluating the potential of aerial infrared as a lek count method for prairie grouse. Journal of Fish and Wildlife Management 6:486–497.
Gregg L, Niemuth ND. 2000. The history, status, and future of sharp-tailed grouse in Wisconsin. The Passenger Pigeon 62:159–174.
Hochachka WM, Fink D, Hutchinson R, Sheldon D, Wong W, Kelling S. 2012. Data-intensive science applied to broad-scale citizen science. Trends in Ecology and Evolution 27:130–137.
Hoglund J, Alatalo RV. 1995. Leks. Princeton, New Jersey: Princeton University Press.
Jenni DA, Hartzler JE. 1978. Attendance at a sage grouse lek: implications for spring censuses. Journal of Wildlife Management 42:46–52.
Johnson DH, Rowland MM. 2007. The utility of lek counts for monitoring greater sage-grouse. Pages 15–23 in Reese KP, Bowyer RT, editors. Monitoring populations of sage-grouse. Moscow, Idaho: Proceedings of a Symposium at Idaho State University.
Kremen C, Ullmann KS, Thorp RW. 2011. Evaluating the quality of citizen-scientist data on pollinator communities. Conservation Biology 25:607–617.
Losey EB, Deemer BR, Corace RG III. 2007. History of sharp-tailed grouse (Tympanuchus phasianellus) at Seney National Wildlife Refuge and surrounding areas, Schoolcraft County, Michigan. The Passenger Pigeon 69:339–348.
Luukkonen DR, Minzey T, Mapsles TE, Lederle P. 2009. Evaluation of population monitoring procedures for sharp-tailed grouse in eastern Upper Peninsula of Michigan. Lansing, Michigan: Michigan Department of Natural Resources. Wildlife Division Report No. 3503. Available: https://pdfs.semanticscholar.org/44ac/898e52cd562dbbe9b0b7b871d88e3c3ff23e.pdf (September 2018).
Monroe AP, Edmunds DR, Aldridge CL. 2016. Effects of lek-count protocols on greater sage-grouse population estimates. Journal of Wildlife Management 80:667–678.
Nichols JD, Hines JE, Sauer JR, Fallon FW, Fallon JE, Heglund PJ. 2000. A double-observer approach for estimating detection probability and abundance from point counts. Auk 117:393–408.
Niemuth ND, Boyce MS. 2004. Influence on landscape composition on sharp-tailed grouse lek location and attendance in Wisconsin pine barrens. Ecoscience 11:209–217.
Osborne JW. 2010. Improving your data transformations: applying the Box-Cox transformation. Practical Assessment Research and Evaluation 15:1–9.