The Journal of Wildlife Management 82(6):1149–1160; 2018; DOI: 10.1002/jwmg.21463

Research Article

Evaluation of Desert Bighorn Sheep Abundance Surveys, Southwestern Arizona, USA

MICHAEL J. CONROY, *Warnell School of Forestry and Natural Resources, University of Georgia, Athens, GA 30602, USA*
GRANT HARRIS,[1] *U.S. Fish and Wildlife Service, P.O. Box 1306, Albuquerque, NM 87103, USA*
DAVID R. STEWART, *U.S. Fish and Wildlife Service, P.O. Box 1306, Albuquerque, NM 87103, USA*
MATTHEW J. BUTLER, *U.S. Fish and Wildlife Service, P.O. Box 1306, Albuquerque, NM 87103, USA*

ABSTRACT Designing wildlife surveys requires biologists to identify their objectives and the accuracy (i.e., bias and precision) required to inform them. In southwestern Arizona, USA, abundance estimates and trends for desert bighorn sheep (*Ovis canadensis*) rely on detection-corrected aerial surveys (i.e., group-size estimator) to inform harvest and assess management actions. The accuracy requirements and trend resolution remain undefined, rendering the surveys ability to address management needs uncertain. We used simulations based on historical surveys to optimize survey efficiency, estimate the accuracy of alternative sampling designs, and evaluate if the accuracy produced by this survey meets management needs. Simulations varied by the amount of area surveyed and temporal frequency. Given annual surveys, we examined trend detection with alternative significance levels (α spanning 0.05–0.30). Our simulations, which accepted the survey assumptions, identified many designs producing unbiased and precise estimates (CV <20%). Alternatives exist for optimizing this survey. For instance, surveying >70% of a Game Management Unit (GMU) provides similar precision to 100% coverage. Population declines $\geq$12%/year are detectable with annual surveys over a 10-year period when $\alpha = 0.05$. Setting $\alpha = 0.3$ enables detecting declines $\geq$14%/year within 3 years. This survey assumes constant detection, calculated in 1 GMU 2 decades ago, and applied to 10 other GMUs since. We tested this assumption by estimating group-size detection from field data in another GMU for 4 years. Bighorn sheep detection varied across GMUs and survey periods. Abundance estimates using the new detection rates were approximately 40% lower than current estimates. Our survey evaluation revealed that differences in abundance of approximately 50% are often not detectable, precision is insufficient to detect large trends in a timely manner (i.e., 40% decline in 7 years), and assuming a fixed detection process remains unfounded. Alternative sampling designs that estimate detectability concomitant with the survey, combined with targeted studies, would better inform management objectives for these desert bighorn sheep. Our assessment demonstrates the problems that occur when survey requirements are vague or mismatch survey design, and that monitoring designs incapable of capturing spatial and temporal variation in detectability will risk misrepresenting animals' population sizes and trends.

KEY WORDS abundance, aerial surveys, detection, double observer, *Ovis canadensis*, precision, sightability model, simulation, survey design, telemetry, trend.

Wildlife surveys often provide valuable inference about animal populations, and typically require significant commitments of time, money, and personnel. Therefore, prior to survey design, wildlife professionals should clearly identify their objectives, evaluate if a survey is the best method to fulfill their objectives, and if so, determine the amount of survey accuracy (i.e., bias and precision) required. By adhering to this process, biologists avoid potential mismatches in survey design, results, and management needs, thereby improving decision-making capacity and targeted resource allocations (Legg and Nagy 2006, Nichols and Williams 2006, Lindenmayer and Liken 2010).

In southwestern Arizona, USA, federal (U.S. Fish and Wildlife Service [USFWS]) and state (Arizona Game and Fish Department) agencies use aerial surveys to monitor population trends of desert bighorn sheep (*Ovis canadensis*). These aerial surveys have 2 main objectives: provide accurate point estimates of bighorn sheep abundance at the resolution of Game Management Units (GMUs) and detect trends in bighorn sheep population trajectories in GMUs and regionally (southwestern AZ). Agencies use these data to

Received: 29 June 2017; Accepted: 27 February 2018

[1]*E-mail: grant_harris@fws.gov*

inform harvest quotas, trigger management (i.e., predator removal), and appraise the effects of prior management (e.g., utility of water catchments, predator control, translocation). The amount of survey bias and precision (accuracy) required to achieve these aims remains undefined. Further, the amount of bias and precision these surveys produce remains unknown. Clearly, this identifies a problem. Managers operate surveys and use the results uncertain of whether the abundance estimates adequately address the management requirements.

These aerial surveys rely on a group-size detection function, and occurred from 1992–present in 11 GMUs throughout southwestern Arizona (Conroy et al. 2014; Fig. 1). Recently, Conroy et al. (2014) evaluated and improved this survey by developing a new, state-space modeling approach to quantify variation in bighorn sheep abundance estimates in each GMU, and to describe regional abundance and population dynamics. Conroy et al. (2014) identified 2 factors contributing to low precision in the survey estimates. First, data used to build the group detection probability (i.e., sightability) were restricted to group sizes <9 bighorn sheep, although subsequent surveys often contained larger group sizes. Addressing this issue requires additional sightability trials. The second factor is poor temporal continuity in survey data within and across GMUs because annually sampling of 100% of habitat for desert bighorn sheep in each GMU is cost prohibitive.

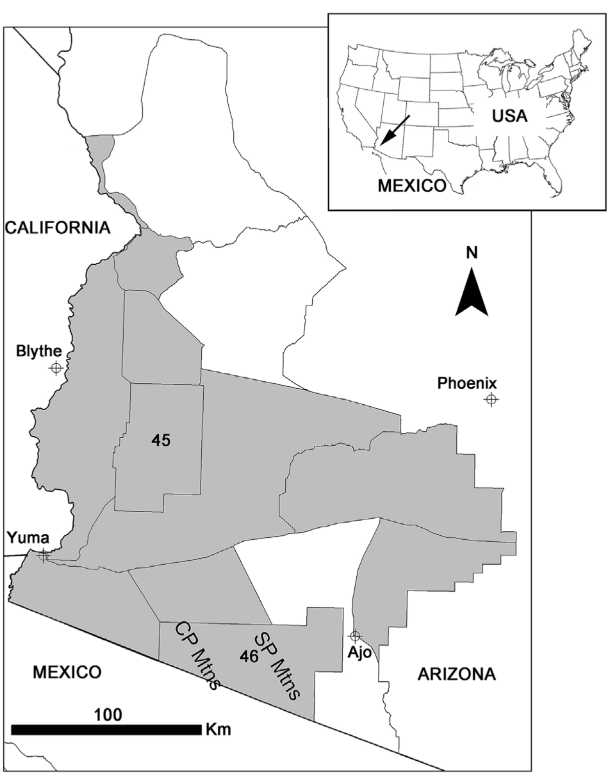

Figure 1. Location of 11 Game Management Units (GMUs, gray shading) receiving abundance surveys for desert bighorn sheep in southwestern Arizona, USA. This project occurred at GMU 45, GMU 46, and the Cabeza Prieta Mountains (CP Mtns) and Sierra Pinta Mountains (SP Mtns) within GMU 46. Previous research calculated detection of desert bighorn sheep groups in GMU 45 for 1993–1995 and this detection rate has been used to estimate abundance in the other GMUs since. In combination with previous research, we calculated detection functions for desert bighorn sheep group sizes in each mountain range of GMU 46 for 2002–2005. We also used these detection functions to estimate abundance in this unit during these years.

We evaluated this survey and the information it provides to improve management of desert bighorn sheep, and to inform survey design and implementation for other wildlife elsewhere. Specifically, we addressed 3 topics. We first focused on tradeoffs between 2 survey elements: the amount of spatial coverage of a survey within a GMU, and the time between surveys on any given GMU. We examined these tradeoffs to identify the most informative and efficient survey designs for estimating bighorn sheep abundance and trend. The most intensive survey design covers 100% of a GMU's area every year (i.e., complete spatial and temporal coverage). We simulated this scenario to exemplify the greatest precision and least bias obtainable (i.e., best-case approach). Other designs we tested reduced spatial and temporal coverage of GMUs, allowing us to assess tradeoffs between the losses in precision and gains in efficiency. Obviously, providing similar results with reduced effort saves costs and increases operational safety (i.e., reducing flight time). Second, we calculated the bias and precision these surveys produce, and examined if survey accuracy is suitable for answering the management needs. By examining these topics, we determined how much reduction in effort, spatially and temporally, could ensue before compromising the accuracy of abundance estimates and trends given the current survey design and its assumptions.

Operationally, these bighorn sheep surveys rely on a detection function built in 1 GMU during 1993–1995 and applied to the other 10 GMUs in southwestern Arizona since. Detection functions calculated in 1 location and period may not apply to other locations and periods, potentially confounding inferences about variation in abundance over space (GMU) and time (survey period). Therefore, as the third step, we calculated detection for groups of desert bighorn sheep in a different GMU (different from the original study GMU) for 4 years (each year separately), assessed variation in detection across locations and time, and compared abundance estimates to those predicted using the current survey methods. Our objectives center on optimizing this survey design for estimating bighorn sheep abundance and trend, given the current group-size estimator approach, and evaluating if this survey design produces credible estimates with acceptable levels of bias and precision to inform management objectives and needs.

STUDY AREA

The study area (i.e., GMU 45) has annual precipitation ranging from approximately 76–400 mm. Elevations span broad valleys (elevations of ~500 m) interspersed with steep and rugged mountains (elevations of ~1,500 m). The GMU 46 study area encompassed approximately 350,000 ha in Pima and Yuma counties, Arizona. The surveys occurred in the Sierra Pinta and Cabeza Prieta mountains, which had rugged topography with slopes commonly >56° and elevations peaking at ~900 m (Cain et al. 2008). Game

Management Unit 46 experiences approximately 50–160 mm of rainfall/year, with more occurring eastward. Wildlife refuges in both GMUs (Kofa National Wildlife Refuge [NWR] and Cabeza Prieta NWR) were established to bolster desert bighorn sheep populations. According to the present survey method, mean abundances of desert bighorn sheep are approximately 500 in each area (Conroy et al. 2014, 2015). The areas also contain populations of mule deer (*Odocoileus hemionus*), Sonoran pronghorn (*Antilocapra americana sonoriensis*), coyotes (*Canis latrans*), and numerous reptile species. Vegetation and other characteristics of the study areas are addressed in Hervert et al. (1998) and Cain et al. (2008).

METHODS

Desert bighorn sheep have been counted at GMU 45 (centered on Kofa National Wildlife Refuge [NWR]) and 10 other GMUs throughout southwestern Arizona since the early 1980s via aerial surveys (Fig. 1). Because the group-size estimator was developed at GMU 45, and GMU 45 has the longest regular series of surveys in comparison to other portions of the region, data from GMU 45 have been used to evaluate the group-size estimator (Conroy et al. 2014, 2015). Individual GMUs have been completely surveyed (100% of the area of a GMU flown) since 1992, but typically not every year.

The survey approach produces abundance estimates by using a modified group-size estimator (Samuel et al. 1987; Fieberg 2012; Conroy et al. 2014, 2015). The method combines sightings of radio-marked groups (i.e., telemetry; Hervert et al. 1998) to estimate group size-specific detection rates for adjustment of counts from operational surveys. This estimation approach has been described in detail, evaluated, and used to obtain estimates of abundance and trends for GMU 45 and the other 10 GMUs from 1992 to 2012 (Conroy et al. 2014, 2015 [1992 estimate calculated retrospectively, after construction of the detection function]). We used these estimates for examining alternative spatial and temporal sampling frequencies. The GMU 45 study area and survey procedures are described in Hervert et al. (1998), with additional operational details provided by Conroy et al. (2014).

Simulation to Examine Bias and Precision of Abundance Estimates

Our method combined historical survey data and simulation modeling to evaluate the performance of alternative survey designs, focusing on the estimation of abundance and detection of trends while accounting for variations in incomplete spatial and temporal coverage. We first addressed the influence of spatial coverage on the accuracy of abundance estimates for a single survey (specified GMU and year). The density of desert bighorn sheep varies throughout the landscape covered by the survey. Our simulation accounts for this spatial heterogeneity in density. To begin, for specified GMU abundance (N), we specified density of individuals (λ) per each of the M sample units in the GMU as:

$$\lambda = N/M.$$

We considered the area sampled as a fraction of $M = 100$ equal-area spatial sampling units comprising the entire area, with $m = fM$ of these taken as sample plots, where f indicates the proportion of area sampled (M and m correspond to N and n in Fieberg (2012) and are used here to avoid confusion with N signifying abundance). We assumed a constant proportional allocation of individuals to group sizes across the GMU $i = 1,2, \dots M$ for sampling unit i, so that numbers per group size were simulated (via a Poisson distribution):

$$N_{g_i} \sim \text{Poisson}(\lambda \pi_g), i = 1, \dots, M,$$

where π_g was the proportion of the population in each group size g, estimated from the previous group-size study (Conroy et al. 2014). Finally, we obtained numbers of groups of each size as:

$$C_{g_i} = N_{g_i}/n_g,$$

where n_g is the number of individuals in group-size category g (i.e., group size 3 has 3 sheep). In this study we used 6 group-size categories: 1–5 and 7, the latter representing the pooled original categories of 6, 7, and ≥ 8 sheep/group (Conroy et al. 2014).

We simulated group-size data for the detection models by passing through each of the m units sampled and generating observed counts as binomial outcomes conditional on the numbers of groups in each unit and predicted detection probability:

$$c_{g_i} \sim \text{Binomial}\left(C_{g_i}, \hat{p}_g\right),$$

where $\hat{p}_g$ is the probability of detection for each group-size category, predicted from the calibration study in Conroy et al. (2014). We then used the group frequencies over all m sampled units as input for the detection model and repeated the process 500 times to evaluate estimator precision and bias. The approach assumes a constant, proportional group-size distribution. In earlier simulations, we generated group-size frequencies C_{gi} by a multinomial process but found that the resulting simulations exhibited higher variability than observed empirically, and thus we employed the above, simpler approach.

We then used the calibration (experimental) data from GMU 45 (Hervert et al. 1998) and simulated data for each year to estimate abundance by detection models (Fieberg 2012; Conroy et al. 2014, 2015). We considered hypothetical, known abundance (N) in the range of 100 to 1,000 per GMU and f of 0.1 to 1.0. For each combination of N and f we replicated the simulations 500 times and computed relative root mean squared error (RMSE) as:

$$\text{RMSE}(\hat{N}) = \frac{\sqrt{\text{MSE}(\hat{N})}}{N},$$

where

$$\mathrm{MSE}(\hat{N}) = 1/500\sum_{j=1}^{500}(\hat{N}_j - N)^2$$

and $\hat{N}_j$was estimated abundance in replication j. We also computed relative bias (RBIAS) and coefficient of variation (CV) as:

$$\mathrm{RBIAS}(\hat{N}) = \frac{\mathrm{BIAS}(\hat{N})}{N}$$

and

$$\mathrm{CV}(\hat{N}) = \frac{\mathrm{Var}(\hat{N})}{N},$$

where

$$\mathrm{BIAS}(\hat{N}) = 1/500\sum_{j=1}^{J}(\hat{N}_j - N)$$

and

$$\mathrm{Var}(\hat{N}) = \mathrm{MSE}(\hat{N}) - \mathrm{BIAS}(\hat{N})^2.$$

Finally, we obtained bootstrapped confidence intervals for $\hat{N}$ from the lower and upper 2.5% quantiles of the simulated estimates.

Simulation to Examine Tradeoffs in Survey Design to Measure Trend

To explore tradeoffs between survey period and spatial coverage, we conducted a simulation study motivated by the state-space model for estimating abundance (Conroy et al. 2014). First, we simulated population growth according to a state-space model developed by Conroy et al. (2014):

$$N_{t+1} = N_t \exp[r_t],$$

where N_t is abundance (whether directly estimated, or not) for a single GMU and r_t is log-scale growth rate in year t. We modeled log-scale growth r_t as a normally distributed random effect

$$r_t \sim N(\bar{r}, \sigma_t),$$

where $\bar{r}$ is average log-scale growth and σ_t is a scale parameter controlling for temporal variability. We took initial abundance as $N_1 = 500$ (in the range of values estimated for GMU 45 during 2007) and $\sigma_t = 0.179$ from the 1992 to 2012 state-space analysis (Conroy et al. 2014). The results in Conroy et al. (2014) indicated strongest support for a model specifying the above random, temporal effect and GMU-specific mean growth, with some GMUs exhibiting positive and others negative growth. However, we were interested in the ability of alternative designs to detect hypothetical trends. Therefore, we specified trend from -1% to -30%, corresponding to values for $\bar{r} = \log(1 + \text{trend}/100)$ of approximately -0.01 to -0.36. We also simulated trajectories for N_t with this range of values for $\bar{r}$, from the initial value of $N_1 = 500$ and $\sigma_t = 0.179$. The simulated trajectory thus contained a fixed effect given a loglinear trend in N from the state-space model and a random effect of normally distributed r_t, conditioned on $\bar{r}$ and σ_t.

For each simulated trajectory of N_t, $t = 1, \ldots, 20$, we first selected a subset of years that would be used for estimation of trends, using intervals between surveys of 1 (annual surveys) to 10 years. For each subset of years, we simulated observed group-size frequencies as above, now conditional on N_t. Once we obtained the estimates of N_t for each year in the state-space model, we fit the estimates to a log-linear model using ordinary least-squares regression (package lm in R; R Foundation for Statistical Computing, Vienna, Austria):

$$\log(\hat{N}_t) = b_0 + b_1 t$$

with $\hat{b}_1$ taken as an estimate of mean log-scale growth. This approach, though not conforming to the state-space process model, produces estimates that are similar to the state-space estimates of $\bar{r}$ but are computationally faster. The process allows us to replicate multiple simulation trials and explore the implications of alternative sampling designs on detecting specified trend levels. For each simulated trajectory of N we likewise estimated $\hat{b}_1$ and took this as realized true growth for computation of bias and mean squared error. We replicated the simulations 1,000 times for each combination of trend, spacing, and the fraction of area sampled. We computed RMSE and bootstrapped confidence intervals for $\widehat{\bar{r}}$ from the lower and upper 2.5% quantiles of the simulated estimates. We provide the programs and data to simulate abundance and trends for these surveys of desert bighorn sheep at https://sites.google.com/site/mjconroybiometrics/resources/dbs/optimal-design.

Lastly, because a 20-year period is too long for biologists to act toward addressing population changes, we also examined the amount of population decline per year that this survey detects over a 3- to 10-year period given annual surveys. We relaxed α from 0.05 to 0.10, 0.20, and 0.30 to examine tradeoffs between trend resolution, time, and Type 1 error (GMU spatial coverage held at 100%). By tolerating a greater chance of falsely declaring a trend (Type 1 error), one decreases the chance of missing a real trend (Type 2 error). In particular, under a cautionary principle, managers would be more concerned about failing to detect a population decline, and thus willing to risk the occasional false alarm. Essentially, the risks of failing to address population declines seem greater than incorrectly identifying them.

Detection Assessment

We tested the assumption that detection was constant across all GMUs by using a simultaneous double-count method to estimate detection of desert bighorn sheep in GMU 46 (Graham and Bell 1989, Cain et al. 2008). Cain et al. (2008) surveyed each mountain range in mid-October to early November from 2002 to 2005 using a helicopter (Bell 206B JetRanger and Bell 206L LongRanger; Bell Helicopter Textron, Fort Worth, TX, USA) and conducted 2 independent surveys in 2003. Cain et al. (2008) attempted to standardize factors that may influence survey results.

Researchers flew all surveys at the same time and at constant survey intensity of 2.9 min/km^2 (Hervert et al. 1998). Three observers participated in all surveys: 1 observer on each side of the helicopter in the rear seats and 1 observer in the front left seat. The pilot was in the front right seat and was not an observer; we used only the 2 left side observers for detection calculations. Although the individual observers varied somewhat throughout each survey of each mountain range, the same 5 personnel were observers on all surveys. Each observer independently searched for desert bighorn and observers did not alert others to the presence of desert bighorn sheep until the rear of the helicopter had passed the animals. Then, the pilot circled back to collect data on group size and composition. We recorded all groups as being observed by either the front observer only, the rear observer only, or both observers.

We calculated detection probabilities for single animals and groups of animals for each survey in each mountain range. We followed Graham and Bell (1989) to calculate the detection probability and variance for the front observer only ($\hat{S}_1$), the rear observer only ($\hat{S}_2$), and the probability that an animal will be observed by $\geq$1 left side observer ($\hat{S}$) using Chapman's (1951) correction as:

$$\widehat{S}_1 = \frac{B}{S_2 + B}$$

$$\widehat{S}_2 = \frac{B}{S_1 + B}$$

$$\widehat{S} = \frac{(B+1)(S_1 + S_2 + B)}{(S_1 + B + 1)(S_2 + B + 1) - (B+1)}$$

$$\mathrm{Var}\left(\widehat{S}\right) = \frac{S_1 S_2 (S_1 + S_2)\widehat{S}}{(S_1 + B)^2 (S_2 + B)^2},$$

where S_1 and S_2 are the number of groups seen by the front and rear observers only, and B is the number of groups seen by both observers. The original data collection followed institutional animal care and use protocols in effect at the time of data collection.

RESULTS

Estimate Precision

The simulated surveys provided abundance estimates that were nearly unbiased (RBIAS <3%) regardless of N or f (proportion of GMU surveyed), so we focused on precision, measured by CV (Fig. 2). In practice, the CVs realized for the 1992–2012 surveys in southwestern Arizona ranged from 0.06 to 0.56 and averaged 0.15. The intensity of spatial sampling and the size of N influenced survey precision. Specifically, surveys with 100% spatial sampling and N of 100 sheep resulted in a CV approximately 20% (Fig. 2; Table S1, available online in Supporting Information; 4 GMUs have approximately 100 sheep [Conroy et al. 2014, 2015]). When sheep abundance is 200, surveys with >50% coverage result in a CV <20% (Fig. 2; Table S1; 5 GMUs have approximately 200 sheep [Conroy et al. 2014, 2015]). With abundances of 500 sheep, a CV <20% occurs when surveying >20% of a GMU (Fig. 2; Table S1; 2 GMUs have approximately 500 sheep [Conroy et al. 2014, 2015]). With 1,000 sheep, the CV is always <20%, falling to <10% when surveying >40% of the GMU (Fig. 2; Table S1; no GMUs have approximately 1,000 sheep [Conroy et al. 2014, 2015]). Irrespective of abundance, there is little gain in precision by surveying >70% of the GMU. The average reduction in CV from 100% GMU coverage to 70% coverage across population sizes of 100–1,000 (increments of 100) is $1.7 \pm 0.81\%$ (SD; Table S1).

We evaluated the ability of this survey to discriminate between 2 abundance estimates within a GMU (i.e., identify change between 2 population estimates). Results depended on estimated abundance, the proportion of GMU surveyed, the magnitude of change in the population, and the precision biologists require. For example, given N of 200, a 20% CV

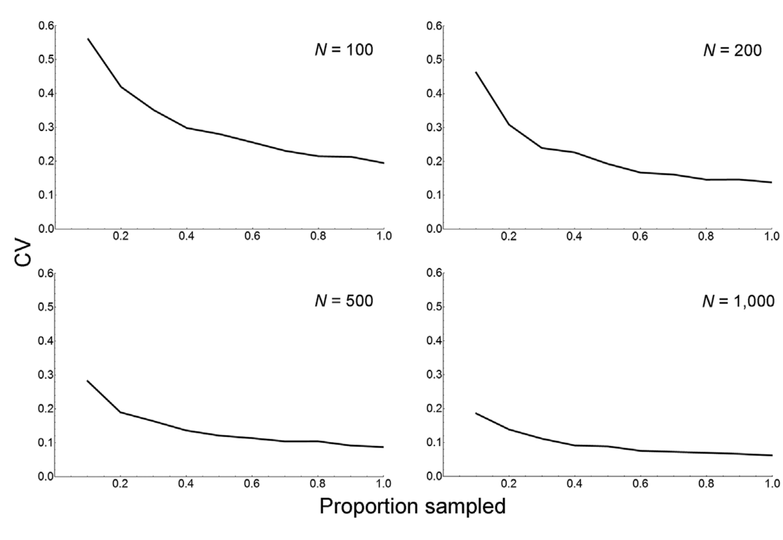

Figure 2. Coefficient of variation (CV) of estimated abundance from detection data for desert bighorn sheep in southwestern Arizona, USA, 1992–2012 (Conroy et al. 2014) versus proportion of area surveyed within a Game Management Unit.

(50% GMU sampling) provides estimates approximately ± 74 sheep (Table S1). A subsequent survey resulting in a 50% population increase (300 sheep) is unable to identify a significant change in abundance at any level of GMU sampling (i.e., the CIs overlap [$\alpha = 0.05$]; Fig. 3). A population doubling (400 sheep) would be detectable with sampling >40% of a GMU (CV 14%; Table S1).

Trend Resolution

Surveys using the group-size estimator began in 1992 and our data series ends at 2012. Over a 20-year period, trends ≤8%/year were virtually undetectable ($\alpha = 0.05$). However, trends ≥10%/year were detectable under many designs (e.g., 30% coverage and 6-year survey period; 50% coverage and 8-year survey period) suggesting opportunities for strategic tradeoffs (Fig. 4).

Understanding population trends in shorter periods will allow biologists to pursue management actions sooner. Therefore, we examined the ability of this survey to detect declining population trends over 3–10 years, given annual survey intervals, 100% GMU coverage, and varying precision (Fig. 5). Our results identify the smallest amount of trend in population decline that a scenario can reliably detect. When $\alpha = 0.05$, the survey can detect a 12% population decline over a 10-year period (Fig. 5). In other words, a sheep population must be declining by ≥12%/year for a 10-year period (95% CI = −26.00%, −0.04%) for this survey to identify that a decline is happening. Within a 3-year and 5-year period, annual surveys can reveal a population decline occurring when the population is declining by ≥26% (95% CI = −58.6%, −3.4%) and 18% (95% CI = −39.0%, −1.1%) per year, respectively (Fig. 5; i.e., decline from 500 to 226 individuals across 5 years with an 18%/year decline).

Given $\alpha = 0.10$, the survey can detect a ≥11% decline/year in 10 years (90% CI = −22.2%, −1.0%). With $\alpha = 0.20$, the survey can detect a ≥10% annual decline over 10 years (80% CI = −19.4%, −2.4%; Fig. 5). Were there 500 sheep in year 1, a 10% annual decline over 10 years would result in 193 sheep by year 10 (68 sheep in year 20). Trend detection for $\alpha = 0.20$ for 3 years is the same as $\alpha = 0.05$ for 5 years (18% annual decline; Fig. 5). With $\alpha = 0.30$, the survey can detect an 8% decline/year over 10 years (70% CI = −15.8%, −1.4%) and 14% decline in 3 years (70% CI = −30.3%, −0.2%; Fig. 5). In keeping with our example using 500 sheep, 370 would remain after 3 years given a 14% annual decline. With $\alpha = 0.30$, there exists a 30% chance of identifying a population decline when none exists.

Detection Assessment

In Unit 46, each mountain range was surveyed once per year in 2002, 2004, and 2005 and twice in 2003 (Table 1). The Sierra Pinta range was broken down into 4 survey blocks, which took an average of 5.3 ± 0.45 hour each year to cover all 4 blocks. The Cabeza Prieta range was divided into 6 survey blocks, which took an average of 7.7 ± 0.55 hour each year to cover all 6 blocks.

During aerial surveys in GMU 46 from 2002 to 2005, left-side observers recorded 91 and 126 groups of desert bighorn sheep in the Sierra Pinta and Cabeza Prieta Mountains, respectively. The number of groups per survey recorded by ≥1 left-side observer ranged from 16 to 21 ($\bar{x} = 18.2 \pm 1.93$ groups/survey) in the Sierra Pinta and 23–29 ($\bar{x} = 25.2 \pm 2.34$ groups/survey) in the Cabeza Prieta Mountains. Group size did not differ by survey ($F_{4,\ 217} = 1.43$; $P = 0.227$) or range ($F_{1,\ 217} = 2.47$; $P = 0.118$). Group sizes were small in both mountain ranges, ranging from 1 to 9 animals/group ($\bar{x}$ group size of 2.24 ± 1.96 sheep/group in the Sierra Pinta and 1.88 ± 1.47 sheep/group in the Cabeza Prieta Mountains). Single animals represented 51.6% (47 of 91 observations) and 61.1% (77 of 126 observations) of all groups observed in the Sierra Pinta and Cabeza Prieta Mountains, respectively.

Detection varied within and between mountain ranges and years (Table 1). Within the same location and group size,

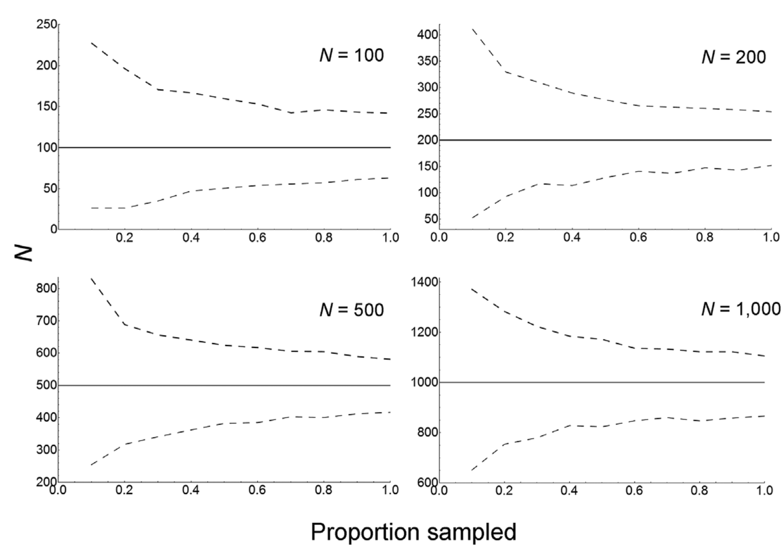

Figure 3. Simulated confidence intervals (dashed lines) for estimating abundances of desert bighorn sheep in southwestern Arizona, USA, in relation to specified proportion of area sampled within a Game Management Unit (1992–2012).

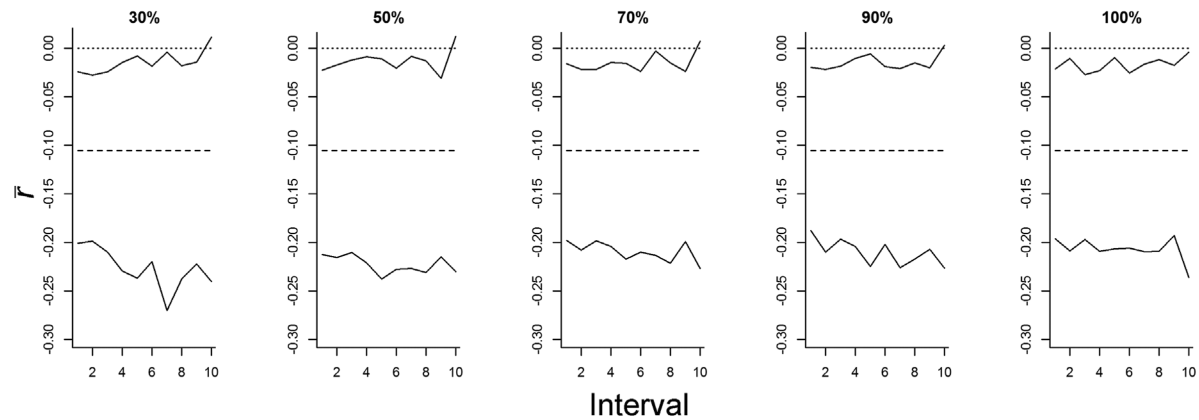

Figure 4. Simulated confidence intervals (solid lines) on estimates of mean growth rate ($\bar{r}$; dashed lines) of desert bighorn sheep in southwestern Arizona, USA, for combinations of specified $\bar{r} = \log(1 + \text{trend}/100)$ for various proportions of area sampled (30–100%) and survey intervals ($n = 1$–10 years between surveys over 20 years [1992–2012]). An upper confidence limit exceeding $\bar{r} = 0$ (dotted lines) indicates failure of design to reject the null hypothesis of population stability ($\alpha = 0.05$). Trend equals -10%.

detection could vary by 62% across time. Across locations, within the same group size and year, detection could vary by 68% (once detection was identical [Table 1]). Any similarities and differences in detection were unpredictable between locations and time (Table 1). The detection probabilities in the Sierra Pinta and Cabeza Prieta Mountains differed from the detection probability calculated in GMU 45 by 1.0–52% for group sizes of 1 to $\geq$3 sheep (Conroy et al. 2014).

We illustrated the effects that different detection estimates have on abundance estimates, by predicting abundance in GMU 46 with the detections in current use (i.e., generated in GMU 45 between 1993 and 1995) and with detections calculated in GMU 46 (2002 and 2005). The current survey (using detection estimates from GMU 45) estimates approximately 350 desert bighorn sheep in GMU 46 for 2002 and 2005 ($\bar{x}$ estimate; Fig. 6; Conroy et al. 2015). For GMU 46, we kept the group-size data and estimation model identical to the current survey but replaced the GMU 45 detections with the detections calculated for GMU 46 (in the Sierra Pinta and Cabeza Prieta mountains, respectively). The abundance estimate for desert bighorn sheep in GMU 46, using the GMU 46 detection, was between 200 and 240 sheep, or approximately 40% less ($\bar{x}$ estimates; Fig. 6). The abundance estimates resulting from the different detection estimates are so different that their confidence intervals do not overlap (Fig. 6).

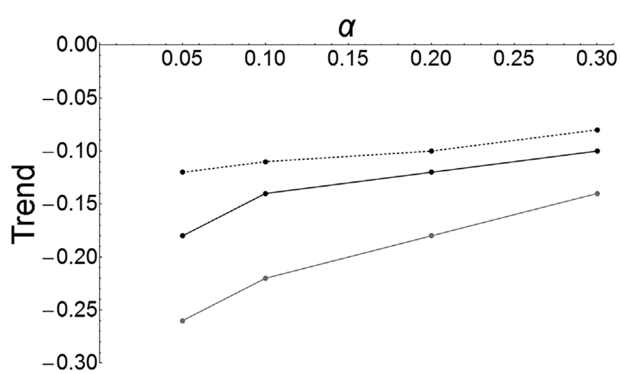

Figure 5. Simulated estimates of mean growth rate ($\bar{r}$, labeled trend) for desert bighorn sheep in southwestern Arizona, USA, given combinations of specified $\bar{r} = \log(1 + \text{trend}/100)$ with varying α (0.05–0.30) and time (3 years [gray], 5 years [solid black], and 10 years [dashed black]), given annual surveys and 100% spatial coverage of a Game Management Unit.

DISCUSSION

Estimate Precision

Assuming a constant detection process across space and time, multiple survey designs for desert bighorn sheep provide unbiased and precise estimates of abundance at the GMU level. Based on the empirical data and simulations, a CV $\leq$20%, a value commonly regarded as standard (Williams et al. 2002), can be consistently obtained using this survey. Most abundance estimates of sheep over 1992–2012 achieved this standard of precision, with some estimates approaching a CV of 10% (Conroy et al. 2015; Table S1). Our results suggest that wildlife managers could reduce the amount of GMU surveyed and acquire adequate precision ($\leq$20%), given the abundance of bighorn sheep in the GMU. For instance, populations of 500 sheep can be sampled with >40% spatial coverage and acquire 14% CV (Fig. 2; Table S1). In all cases, there is little gain in precision by sampling >70% of a GMU (Fig. 2). Managers can use this information to evaluate if the gains in precision warrant the additional costs.

The amount of effort and precision required must match the needs for the survey. For example, biologists are often concerned with changes in population size. Given a survey with 10% CV and surveying 70% of the GMU, a population of 500 sheep must increase to 800 sheep before the estimates are considered statistically different (Table S1). A 20% CV would require a population exceeding 1,000 sheep before differences were detectable (Table S1). Our results indicate that the current survey methodology generally lacks sufficient power to detect considerable changes in abundance (i.e., 50%).

Table 1. Detection and variance estimates by group size for desert bighorn sheep from aerial surveys conducted using a simultaneous double count method in the Sierra Pinta and Cabeza Prieta mountains, Cabeza Prieta National Wildlife Refuge (Game Management Unit 46), USA, 2002–2005. Year 2003 contained 2 independent surveys (a and b).

	Group size					
	1		2		≥3	
Year	Detection	Variance	Detection	Variance	Detection	Variance
Sierra Pinta						
2002	0.909	0.011	0.857	0.094	0.882	0.035
2003a	0.818	0.055	0.882	0.035	0.923	0.022
2003b	0.727	0.055	0.857	0.094	0.667	0.104
2004	0.727	0.055	0.857	0.094	0.571	0.084
2005	0.579	0.089	0.923	0.022	0.857	0.043
Cabeza Prieta						
2002	0.692	0.029	0.769	0.117	0.857	0.094
2003a	0.854	0.012	0.800	0.011	0.882	0.035
2003b	0.949	0.004	0.923	0.022	0.968	0.003
2004	0.957	0.003	0.923	0.014	0.960	0.003
2005	0.712	0.038	0.800	0.111	0.857	0.427

Trend Resolution

Relationships between the amount of GMU to sample and the frequency of surveys depend on the magnitude of the trend required. For this survey, designs are incapable of detecting declines <10%/year over a 20-year period ($\alpha = 0.05$; Fig. 4). Coarser trends (i.e., ≥10% decline/year over 20 years) are detectable under a number of design combinations, allowing flexibility in the allocation of effort, be it GMU spatial coverage or temporal frequency (Fig. 4). Allowing larger Type 1 error (e.g., $\alpha = 0.3$) enables detecting smaller declines (e.g., 8% decline/year over 10 years or 14% decline/year over 3 years; Fig. 5). This outcome requires annual surveys with 100% coverage and accepting a 30% chance of identifying population declines when none may exist.

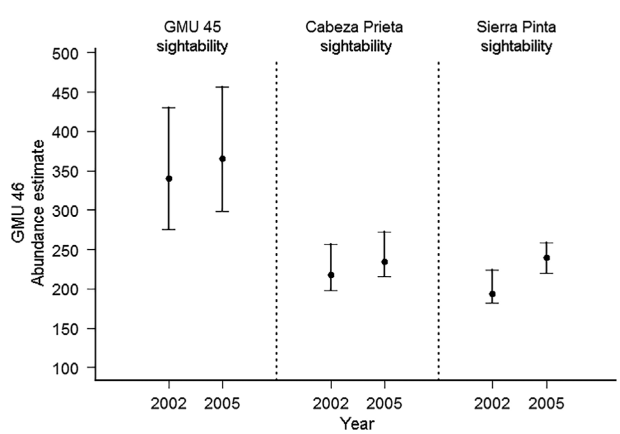

Figure 6. Three sets of abundance estimates for desert bighorn sheep at Game Management Unit (GMU) 46 (encompassing Cabeza Prieta National Wildlife Refuge) varying by the detection parameter used. Previous research in GMU 45 generated detection functions during 1993–1995 and they have been employed in GMU 46 and 10 other GMUs since. Previous research also generated detection functions (sightability) for the Cabeza Prieta Mountains and Sierra Pinta Mountains in GMU 46, independently during 2002 and 2005. These abundance estimates rely on the same calculations and group-size data for GMU 46, for a given year (only the detectability changes). Mean abundance estimates for GMU 46 using detections from GMU 46 are approximately 40% lower than estimates using the detections calculated from GMU 45, and the confidence intervals (black bars) do not overlap.

These results examining trend could be considered optimistic, best-case scenarios because they rely on a simple log-linear (exponential) model of growth, and not more complex (e.g., density-dependent) growth models. Any deviations from these conditions would further degrade trend resolution.

In practice, 9 of the 11 GMUs typically had a survey interval of 3 years (i.e., survey conducted in year 1 then again in year 4). Hence, a 3-year survey interval appears the scale that biologists attempt to judge population trends with this survey. Given this interval, the shortest, detectable trends occur on a 7-year timeframe (i.e., 3 data points). We suggest 7 years is too long for biologists to meaningfully address population changes because relatively small annual changes (i.e., −8%) can generate relatively large cumulative changes over a 7-year period (i.e., ~40% decline in abundance). Annual surveys within 3 years are capable of detecting a 14% annual decline (36% total decline in a 4-year period, $\alpha = 0.3$; Fig. 5). The measured population decline between these examples is essentially identical (i.e., registering a ~40% decline before considering management response), rendering the resolution of a 14% annual decline over a 4-year period perhaps equally unacceptable.

Whether trend detection or identifying population change assists bighorn sheep management depends on reasons for collecting these data and the risk of being wrong (i.e., identifying declining trends when such trends do not exist). If this survey continues unmodified, we consider the risk of being wrong lower than the risk of failing to detect a strong trend or change, justifying use of a larger α (i.e., 0.30) and shorter interval (i.e., 3 years). Regardless, our evaluation reveals that this survey does not provide sufficient precision in abundance estimates to identify if meaningful trends are occurring over useful periods (per above).

Survey Assumptions and Assessment

Inference from these bighorn sheep surveys in southwestern Arizona requires adhering to 3 assumptions: observers identify group sizes of desert bighorn sheep accurately (i.e., when seen, a group of 5 sheep is not classified as 4 or 8 sheep), the detection function for group size derived from 1 GMU applies to the other 10 GMUs, and the detection function remains constant across time, observers, and varying survey conditions (i.e., the ability to observe and tally sheep groups never changes). We have shown the last 2 assumptions are untenable.

Counting bighorn sheep is an imperfect process. Changes in abundance estimates can reflect an observer's ability to detect sheep and sheep behavior or movements, and not indicate true changes in the sheep population. Therefore, the group-size estimator, which produces a detection-corrected estimate of abundance, should account for the detection process of sheep groups and membership (i.e., membership meaning a group classified as 5 is truly 5 bighorn sheep; Clement et al. 2017). Instead, the parameters used in the group-size estimator that describe the detection process for groups were generated during 1993–1995 and remained in use for multiple observers, under various survey conditions (e.g., light levels, weather) in 10 different locations, for over 20 years (Conroy et al. 2014, 2015). The technique does not allow the detection process to vary among observers, sites, and conditions. Instead, the detection process only varies by bighorn sheep group sizes. Therefore, the estimates of sheep abundance produced by the group-size estimator are biased, because a non-constant detection process exists as evidenced by the varying detection rates obtained over time for GMU 46 using the double observer approach. Such bias, in addition to the uncertainty we have already identified, overshadow the true population dynamics of sheep in southwestern Arizona.

When estimation of the detection process is not inherently part of the survey technique, a survey will remain plagued by assumptions of a constant detection process across observers, sites, and time (Lancia et al. 2005). Changing detection rates across location and time often occurs for multiple taxa elsewhere (Lancia et al. 2005, Griffin et al. 2013, Gibson-Reinemer et al. 2016, Stewart et al. 2017).

Using telemetry to determine the detection process forms a good approach because it attempts to account for animals that are in the survey area but unavailable to be seen and recorded by observers (i.e., accounts for probability of detection and availability probability). For desert bighorn sheep, this could be sheep behind rock outcrops or in caves. The double observer method for estimating the detection process does not account for such hidden animals (i.e., accounts for detection probability but not availability probability; Hervert et al. 1998). Therefore, the double observer approach is more likely to estimate a higher detectability than a technique based on telemetry. If true, then the population estimates generated from the double observer technique would be lower than estimates using telemetry (Caughley et al. 1976). Data from Hervert et al. (1998) and our results provide evidence for this outcome. Unfortunately, the actual (i.e., true) abundance of desert bighorn sheep remains unknown in these GMUs, negating formal comparisons between these techniques. Given this situation, the better method is the one that meets sampling assumptions, is logistically feasible, and remains sustainable. In our assessment, because the double observer approach meets sampling assumptions better (i.e., does not assume a constant detection process across locations and years), it provides the more credible abundance estimate for the survey years in GMU 46.

Evaluating the performance of double observer or telemetry-based approaches to a known abundance was beyond the scope of this project. Regardless, irrespective of the absolute veracity of the double observer approach we used, it reveals that the detection process varies considerably through time and space.

We are not promoting one survey technique over the other. Instead, we emphasize that detection must be estimated during each survey process, irrespective of the technique used, to ensure that the resulting abundance estimates are unbiased by observers and survey conditions. A survey using telemetry that generates detection concomitant with the survey would also provide a credible abundance estimate.

Support to continue using a constant detection process across time and space is based on the costs necessary to obtain detection rates (especially with telemetry), the idea that constant detection represents the best information available, and the assumption that if the surveys' detection rates remain consistent over time, then population trends and estimates are close enough. We contend that approximate data are obtainable via other, more affordable methods (i.e., minimum count), best information may misrepresent truth, and the tool (method) should match the job requirements (objectives). The accuracy of data required depends on how they will be used, the risk of being wrong, and the importance of being right.

The sustainability of using telemetry to estimate detection for each survey location and period remains challenged by the expense and logistical effort to maintain telemetry on animals. This may cause other techniques, like double observer to become more tenable. Formal comparisons of abundance estimates obtained from double observer, telemetry, and similar methods form the only way to examine their accuracy and tradeoffs.

Survey Consequences

The unbiased abundance estimates and precision we calculated herein occur because data used to simulate and test bias and precision adhered to this survey's assumptions. Because these assumptions are untenable, the abundance estimates produced by this survey do not reflect the true abundance of sheep in a GMU. The resulting estimates are biased by unknown and varying magnitudes because true changes in the detection process through time are unknown. Tangible outcomes of not accounting for variation in the detection process include wide fluctuations in abundance estimates between years within GMUs and across GMUs (Conroy et al. 2015). The wildly stochastic ($\sigma_t = 0.179$) nature of these sheep populations provides further evidence for such a problem (Conroy et al. 2015).

Presently, federal and state wildlife agencies use these surveys for informing harvest levels (i.e., hunting), and appraising management actions to promote abundance increases (i.e., translocation, predator control [USFWS 2009]). Harvest of desert bighorn sheep in Arizona is male-only, conservative, and based primarily on 15–25% of the estimated class III and IV males observed during surveys. Little precision in abundance estimates is required for such assessments, making lower quality data appear sufficient for informing harvest quotas.

Pitfalls occur, however, when such low-quality data (i.e., minimum counts or this survey in its current design) are used to assess deeper ecological questions such as the causes for population trends or testing the efficacy of management actions. Biologists use data from this survey to recommend when to conduct and cease predator control (USFWS 2009). When the bighorn sheep population falls below a numerical threshold, predator control can be implemented and when it reaches a higher numerical threshold, predator control can cease (USFWS 2009). Given the ethical and political sensitivities surrounding predator control, greater certainties in survey results seem desirable.

Instead of using an abundance survey to monitor the effects of predation or translocation on population growth, a concerted study of cause-specific mortality and other factors affecting population growth (e.g., disease, habitat condition, immigration and emigration; Sæther 1997) would likely be more informative. Granted, given the multitude of factors influencing population growth, these types of studies are ecologically complex and expensive.

If the group-size estimator had accounted for variability in detection across GMU and time, would it have changed the management of desert bighorn sheep in this area? The answer depends on the decisions made based on these data. We lack information about these decisions. However, within the field of wildlife biology, too frequently wildlife surveys will dictate the management, when the opposite should be true. This appears to be the case for the desert bighorn sheep surveys in southwestern Arizona. Therefore, any management shaped by abundance and trend data (e.g., harvest, translocations, predator control, target abundance thresholds) would be affected (to unknown degrees) because they were influenced by biased estimates.

Potential Improvements

The technique of using detection of group sizes for estimating abundances seems sound if the relationship between group size and detectability is accounted for during each survey. Further, future surveys could reduce variance more. For example, the survey could incorporate additional explanatory variables into its calibration and operation (e.g., aircraft speed, altitude, lighting conditions, topography, vegetation, and observer experience) in addition to a wider range of group sizes (Samuel et al. 1987; Strobel and Butler 2014; Conroy et al. 2014, 2015). A parameter accounting for individuals within groups that were missed by observers would also reduce bias and uncertainty in the abundance estimates (Conroy et al. 2014, Clement et al. 2017).

The major flaw centers on implementation that uses the same detection across locations and time. Hervert et al. (1998) recognized that detection probabilities they reported in GMU 45 may not apply to other areas having significantly different habitat characteristics. Our analyses confirm that detection probabilities gained in 1 location may not apply to another location, and we extend this premise to include changes in time. The only way to address varying detection between surveys is by having the detection process estimated during each survey. For instance, the USFWS Breeding Ground Waterfowl Surveys combine aerial counts with a subsample of ground counts (taken as a close approximation of truth) to build predictive relationships for deriving visibility correction factors (VCF) by vegetation type, species, and survey (Zimmerman et al. 2012). A VCF is analogous to the detection model adjustments for group sizes of desert bighorn sheep except it can vary across space and time. A similar approach could use telemetry or a double observer approach on a subsample of sheep surveys to build the detection models (as we performed herein). For large mammals, abundance surveys incorporating detection probability are becoming increasingly common (Jacques et al. 2014, Lubow and Ransom 2016, Smyser et al. 2016). Elsewhere, many other examples of detection applications (e.g., distance sampling, mixture models based on repeated counts, mark-resight, subsampling, hybrid models) include a calibration (i.e., experimental) sample within the survey design (Fieberg 2012, Schmidt et al. 2012, Schmidt and Rattenbury 2013). For example, The New Mexico Department of Game and Fish is evaluating a hybrid double observer-sightability approach for desert bighorn sheep surveys (Griffin et al. 2013). After telemetered groups are used to build the detection function, double observer data refine and calibrate it over time (Griffin et al. 2013).

Currently, surveys for bighorn sheep in GMU 45 and GMU 46 cost approximately $50,000 and $44,000 per survey (US$). Given GMU 45 was surveyed 10 times during 1994–2012, and GMU 46 surveyed 7 times, wildlife agencies spent nearly a million dollars in pursuit of estimating bighorn sheep abundance and trends in these GMUs alone. If this survey continues to follow the same basic protocols and assumptions, then alternatives exist for optimizing it, especially regarding spatial coverage. Pursuing these alternative designs sacrifice little in estimate precision and trend while reducing costs. The survey results, however, will remain suspect until the survey technique incorporates an estimation of the detection process within it (Lancia et al. 2005).

Moving forward, estimating detection, during subsets of the survey flights, seems an appealing approach. Additional pilot work would be required, however, to evaluate how improvements to the detection process affect abundance and trend resolution.

Prior to survey modification or redesign, we suggest that the biologists managing these desert bighorn sheep identify the type of information they need, and specify why they need it, for informing their management actions and decisions (e.g., population management, prescribing sustainable hunts; Lindenmayer and Likens 2009, 2010). This information will steer construction of credible surveys and studies to address their objectives and identify the amount of accuracy (precision and bias) required. Such an approach will advance the management of desert bighorn sheep in southwestern Arizona.

MANAGEMENT IMPLICATIONS

Biologists should identify specific management or conservation objectives that wildlife surveys can inform, prior to survey design and implementation. Then, the survey design and effort required match the amount of bias and precision needed. Concomitantly, a survey must rely on tenable assumptions, to ensure a credible methodology.

We commend designers of this abundance survey for desert bighorn sheep in southwestern Arizona, for recognizing the importance of detectability and working to incorporate it. Unfortunately, our assessment of the current survey identifies survey deficiencies: differences in abundance of approximately 50% are often not detectable using standard levels of precision, precision remains insufficient to detect substantial trends within short periods (i.e., 4 years), and the assumption of a fixed detection process is untenable. The detection of bighorn sheep varied by time and location, demonstrating the inapplicability of using 1 detection process through time and across locations. Therefore, the current survey does not produce credible data. Further, the questions these surveys inform appear uncertain, abundance data may inappropriately address those questions, and the required amount of survey accuracy remains undefined. Identifying and describing these problems helps justify and explain why survey re-evaluation and a likely revision are necessary.

Evaluating survey effectiveness and efficiency forms an important part of adaptive management that agencies strive for. Our research identified weaknesses in survey methodology and suggested improvements that could be implemented depending on the management needs. These improvements center on estimating the detection process for each survey event. The evaluation we provide is instructive for initiation and design of wildlife surveys elsewhere.

ACKNOWLEDGMENTS

The findings and conclusions in this article are those of the authors and do not necessarily represent the views of the USFWS. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government. We thank L. A. Smythe and R. S. Henry for project initiation and support. We appreciate the survey design and considerable amount of surveying effort conducted by personnel from the Arizona Game and Fish Department and the USFWS. J. R. Heffelfinger, J. D. Wehausen, K. M. Proffitt, and an anonymous reviewer provided constructive comments on earlier drafts.

LITERATURE CITED

Cain, J. W. III, P. R. Krausman, J. R. Morgart, B. D. Jansen and M. P. Pepper. 2008. Responses of desert bighorn sheep to removal of water sources. Wildlife Monographs 17:1–32.

Caughley, G., R. Sinclair, and D. Scott-Kemmis. 1976. Experiments in aerial survey. Journal of Wildlife Management 40:290–300.

Chapman, D. G. 1951. Some properties of the hypergeometric distribution with applications to zoological sample censuses. University of California Publications in Statistics 1:131–159.

Clement, M. J., S. J. Converse, and J. A. Royle. 2017. Accounting for imperfect detection of groups and individuals when estimating abundance. Ecology and Evolution 18:7304–7310.

Conroy, M. J., R. S. Henry, and G. Harris. 2014. Estimation of regional sheep abundance based on group sizes. Journal of Wildlife Management 78:904–913.

Conroy, M. J., R. S. Henry, and G. Harris. 2015. Estimation of regional sheep abundance based on group sizes. Erratum. Journal of Wildlife Management 79:865–867.

Fieberg, J. R. 2012. Estimating population abundance using sightability models: R SightabilityModel Package. Journal of Statistical Software 51:1–20.

Gibson-Reinemer, D. K., D. R. Stewart, M. W. Fritts, J. A. DeBoer, and A. F. Casper. 2016. Estimating the effects of environmental variables and gear type on the detection and occupancy of large-river fishes in a standardized sampling program using multiseason Bayesian mixture models. Transactions of the American Fisheries Society 36:1445–1456.

Graham, A., and R. Bell. 1989. Investigating observer bias in aerial survey by simultaneous double-counts. Journal of Wildlife Management 53: 1009–1016.

Griffin, P. C., B. C. Lubow, K. J. Jenkins, D. J. Vales, B. J. Moeller, M. Reid, P. J. Happe, S. M. McCorquodale, M. J. Tirhi, J. P. Schaberi, and K. Beirne. 2013. A hybrid double-observer sightability model for aerial surveys. Journal of Wildlife Management 77:1532–1544.

Hervert, J. J., R. S. Henry, M. T. Brown, and R. L. Kearns. 1998. Sighting rates of bighorn sheep during helicopter surveys on the Kofa National Wildlife Refuge, Arizona. Desert Bighorn Council Transactions 20:11–26.

Jacques, C. N., J. A. Jenks, T. W. Grovenburg, R. W. Klaver, and C. S. Deperno. 2014. Incorporating detection probability into northern Great Plains pronghorn population estimates. Journal of Wildlife Management 78:164–174

Lancia, R. A., W. L. Kendall, K. H. Pollock, and J. D. Nichols. 2005. Estimating the number of animals in wildlife populations. Pages 106–153 *in* C. E. Braun, editor. Techniques for wildlife investigations and management. Sixth edition. The Wildlife Society, Bethesda, Maryland, USA.

Legg, C. J., and L. Nagy. 2006. Why most conservation monitoring is, but need not be, a waste of time. Journal of Environmental Management 78:194–199.

Lindenmayer, D. B., and G. E. Likens. 2009. Adaptive monitoring: a new paradigm for long-term research and monitoring. Trends in Ecology and Evolution 29:482–486.

Lindenmayer, D. B., and G. E. Likens. 2010. The science and application of ecological monitoring. Biological Conservation 143:1317–1328.

Lubow, B. C., and J. I. Ransom. 2016. Practical bias correction in aerial surveys of large mammals: validation of hybrid double-observer with sightability method against known abundance of feral horse (*Equus caballus*) populations. PLoS ONE 11(5):e0154902. https://doi.org/10.1371/journal.pone.0154902

Nichols, J. D., and B. K. Williams. 2006. Monitoring for conservation. Trends in Ecology and Evolution 21:668–673.

Sæther, B. E. 1997. Environmental stochasticity and population dynamics of large herbivores: a search for mechanisms. Trends in Ecology and Evolution 12:143–149.

Samuel, M. D., E. O. Garton, M. W. Schlegel, and R. G. Carson. 1987. Visibility bias during aerial surveys of elk in northcentral Idaho. Journal of Wildlife Management 51:622–630.

Schmidt, J. H., and K. L. Rattenbury. 2013. Reducing effort while improving inference: estimating Dall's sheep abundance and composition in small areas. Journal of Wildlife Management 77:1048–1058.

Schmidt, J. H., K. L. Rattenbury, J. P. Lawler, and M. C. MacCluskie. 2012. Using distance sampling and hierarchical models to improve estimates of Dall's sheep abundance. Journal of Wildlife Management 76:317–327.

Smyser, T. J., R. J. Guenzel, C. N. Jacques, and E. O. Garton. 2016. Double-observer evaluation of pronghorn aerial line-transect surveys. Wildlife Research 43:474–481.

Stewart, D. R., M. J. Butler, G. Harris, L. A. Johnson, and W. R. Radke. 2017. Estimating abundance of endangered fish by eliminating bias from non-constant detectability. Endangered Species Research 32:187–201.
Strobel, B. N., and M. J. Butler. 2014. Monitoring whooping crane abundance using aerial surveys: influences on detectability. Wildlife Society Bulletin 38:188–195.
United States Fish and Wildlife Service [USFWS]. 2009. Limiting mountain lion predation on desert bighorn sheep on Kofa National Wildlife Refuge, Yuma and La Paz counties, AZ. Federal Register 74:38667–38668.
Williams, B. K., J. D. Nichols, and M. J. Conroy. 2002. Analysis and management of animal populations. Academic Press, San Diego, California, USA.
Zimmerman, G. S., J. R. Sauer, W. A. Link, and M. Otto. 2012. Composite analysis of black duck breeding population surveys in eastern North America. Journal of Wildlife Management 76: 1165–1176.

Associate Editor: Kelly Proffitt.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's website.