Received: 25 August 2021 | Revised: 23 March 2022 | Accepted: 3 April 2022

DOI: 10.1002/jwmg.22239

RESEARCH ARTICLE

THE JOURNAL OF WILDLIFE MANAGEMENT

THE WILDLIFE SOCIETY

Accounting for residual heterogeneity in double-observer sightability models decreases bias in burro abundance estimates

Jacob D. Hennig[1] | Kathryn A. Schoenecker[2] | James W. Cain III[3] | Gary W. Roemer[4] | Jeffrey L. Laake[5]

[1]Contractor with the U.S. Geological Survey, Fort Collins Science Center, 2150 Centre Ave, Fort Collins, CO 80526, USA

[2]U.S. Geological Survey, Fort Collins Science Center, 2150 Centre Ave, Fort Collins, CO 80526, USA

[3]U.S. Geological Survey, New Mexico Cooperative Fish and Wildlife Research Unit, Department of Fish, Wildlife, and Conservation Ecology, New Mexico State University, Las Cruces, NM 88003, USA

[4]Department of Fish, Wildlife, and Conservation Ecology, New Mexico State University, Las Cruces, NM 88003, USA

[5]Independent Researcher, Escondido, CA 92025, USA

Correspondence

Jacob D. Hennig, Fort Collins Science Center, 2150 Centre Ave, Bldg C, Fort Collins, CO 80526, USA.
Email: jhennig@contractor.usgs.gov

Present address

Gary W. Roemer, 712 Stone Canyon Dr., Las Cruces, NM 88011, USA.

Funding information

U.S. Bureau of Land Management, Grant/Award Number: Interagency agreement L19PG00052; U.S. Geological Survey; U.S. Department of Defense, Fort Irwin, Grant/Award Number: MIPR 10703187

Abstract

Feral burros (*Equus asinus*) and horses (*E. ferus caballus*) inhabiting public land in the western United States are intended to be managed at population levels established to promote a thriving, natural ecological balance. Double-observer sightability (M_{DS}) models, which use detection records from multiple observers and sighting covariates, perform well for estimating feral horse abundances, but their effectiveness for use in burro populations is less understood. These M_{DS} models help minimize detection bias, yet bias can be further reduced with models that account for unmodeled variation, or residual heterogeneity, in detection probability. In populations containing radio-marked individuals, residual heterogeneity can be estimated with M_{DS} models by including a covariate that corresponds to the marked status of a group (M_H models). Another approach is to use information from detections missed by both observers to account for the characteristics that make groups more or less likely to be detected, or recaptured, by the second observer (M_R models). We used aerial survey data from 3 burro populations (Sinbad Herd Management Area, UT [2016–2018], Lake Pleasant Herd Management Area, AZ [2017], and Fort Irwin National Training Center, CA [2016–2017]) to develop M_{DS} models applicable for feral burros in the southwestern United States. Our objectives were to quantify precision and bias of standard M_{DS} surveys for feral burros and to examine which model type for incorporating residual heterogeneity (M_H or M_R) would result in the least-biased estimates of burro populations relative to the minimum number known alive (MNKA) within the Sinbad Herd Management Area. Standard M_{DS} model estimates achieved a mean coefficient of variation of 0.08, while underestimating MNKA by an average of 27.1%. Accounting for residual heterogeneity through recapture probability in M_R models resulted in estimates closer to MNKA than M_H models (9.5% vs. 16.5% less than MNKA). Our results indicate that M_{DS} models can achieve precise enough estimates to monitor feral burro populations, but they routinely produce negatively biased estimates. We encourage the use of radio-collars to reduce bias in future burro surveys by accounting for residual heterogeneity through M_R models.

KEYWORDS

abundance estimation, aerial survey, burro, detection, *Equus asinus*, feral equids, residual heterogeneity

The Wild Free-roaming Horses and Burros Act (Public Law 92-195) mandates the protection of feral equid populations on federally owned public land in the United States. The Act tasks the United States Department of Interior's Bureau of Land Management (BLM) and the United States Department of Agriculture's Forest Service to manage feral equid populations and habitat in areas where they occurred at the time of its enactment (Public Law 92-195). Management actions include maintaining site-specific population targets, allocating forage to equids, livestock, and wildlife species, applying fertility controls to select individuals, and removing individuals when the population is above an appropriate management level (National Research Council 2013). Underpinning these actions are abundance estimates derived from, in most areas, aerial survey counts. Raw counts are negatively biased owing to imperfect detection (Samuel and Pollock 1981); consequently, counts must be appropriately adjusted before they can effectively guide management.

Sightability models (Steinhorst and Samuel 1989) and double-observer designs (Graham and Bell 1989) are 2 widely employed methods for correcting aerial counts by estimating the probability of detection and its uncertainty. Sightability models use detections of known individuals to account for the influence of sighting covariates on detection probability, whereas double-observer designs imitate a 2-sample mark-recapture method (e.g., Lincoln-Petersen) to estimate detection probabilities for each observer and abundance. Each method possesses limitations restricting their usefulness. Sightability models do not account for differences among survey observers and conditions when they differ from the initial calibration survey to develop the model. Simple double-observer designs are subject to negative bias in abundance due to variation (heterogeneity) in detectability among observation units (i.e., individuals or groups; Otis et al. 1978). Huggins (1989) proposed modeling heterogeneity in detection probabilities as a function of covariates and provided an estimator of abundance and its precision. Double-observer models that use sightability covariates (i.e., M_{DS}; Walter and Hone 2003, Lubow and Ransom 2016) generate estimates precise enough to guide management practices of feral horses (*Equus ferus caballus*) in the western United States (Lubow and Ransom 2016); consequently, the BLM suggested an M_{DS} protocol as one allowable method for feral equid aerial surveys (BLM 2010). While M_{DS} models appear to provide reasonable abundance estimates for burros (*E. asinus*; Griffin 2015), the effectiveness of M_{DS} models has yet to be evaluated across multiple feral burro populations.

Burros are notoriously more difficult to detect than horses given their smaller body size, cryptic pelage, and small group sizes (D. Little, BLM, unpublished report; Griffin 2015). Further, burros predominantly inhabit more topographically rugged areas, and tend to stand still when an aircraft is overhead rather than running (flushing) like horses (Griffin 2015). The Arizona Game and Fish Department and BLM previously used a double-observer framework that did not make use of sighting covariates to generate estimates of burro abundances, but they claimed that 30–70% of burros may go undetected depending on environmental conditions (D. Little, unpublished report). The recording of sighting covariates is now standard in BLM aerial survey protocols (Griffin et al. 2020), but the sighting covariates most important to burro detection have been assessed only for a small number of populations (Griffin 2015, Gedir et al. 2021). It would be beneficial to use a larger sample size with detections from additional study areas to refine our understanding of the most important covariates.

Abundance estimates from M_{DS} surveys are typically biased low because of unmodeled (residual) heterogeneity in detection probabilities (Borchers et al. 2006, Southwell et al. 2007, Griffin et al. 2013). This means that there is still dependence in detection probabilities among observers even after the inclusion of sighting covariates. Detections among observers (i.e., capture histories) are assumed to be independent of one another, yet certain factors make some target individuals or groups more likely to be detected or missed by both observers, which induces dependence (Borchers et al. 2006). For example, a large group of burros is much more likely to be detected by both observers, and a group completely concealed by canopy cover is more likely to be missed by both. Such dependency in detection among observers can be modeled by including group size and concealment cover as covariates, but there are often unobservable or unknown dependencies that are not included in the model. That residual (unmodeled) heterogeneity induces dependence such that the recapture probability for the second observer will be greater than their initial capture probability. In other words, the second observer has increased probability of detecting a group that the first observer also detected (recapture), compared to detecting a group that the first observer missed (initial capture). When there is no residual heterogeneity, these detections are independent events, and the capture and recapture probabilities will be the same. If independence is assumed, but residual heterogeneity exists, the estimates of capture probability are positively biased leading to negative bias in abundance (Otis et al. 1978, Borchers et al. 2006).

In a standard 2-sample double-observer design, there is no information on groups missed by both observers (Buckland et al. 2010) and independence must be assumed without ancillary information. Consequently, researchers have developed several different alternative approaches to account for residual heterogeneity (Buckland et al. 2010, Barker et al. 2014, Becker and Christ 2015). Populations containing radio-marked individuals present the opportunity to include an additional capture history in the model design that corresponds to detections of marked individuals made by a telemetry observer (Griffin et al. 2013). During a survey, the telemetry receiver enables perfect detection of radio-marked individuals, and consequently, can detect groups missed by human observers. With a large enough sample size, these detections provide the information needed to account for residual heterogeneity.

One modeling approach for incorporating this information is to include a covariate corresponding to the radio-marked status (marked or unmarked) of a detected group (Griffin et al. 2013, Schoenecker and Lubow 2016, Bristow et al. 2019). This approach, mark-type heterogeneity (M_H), relies on the assumption that there is a difference in detection probability between groups with and without a radio-marked individual. Radio-marked groups are unconditionally included in a dataset because they are perfectly detected by telemetry, whereas unmarked groups are only conditionally included if they were seen by at least one observer. Thus, if residual heterogeneity exists, the sample of unmarked groups should mostly consist of those that are relatively easier to detect (Griffin et al. 2013). Further, because radio-marked individuals are assumed to be a random sample of the population, they should be in groups missed by both observers; thus, groups with a radio-marked individual should have a lower average probability of being detected by human observers, compared to unmarked groups in a set of visual-only observations (Griffin et al. 2013). Consequently, marked groups (which constitute the random sample) are thought to be harder on average to detect by human observers (Griffin et al. 2013). The estimated difference in detection between marked and unmarked groups is then used to adjust the detection probabilities of unmarked groups in M_H models, thereby decreasing the amount of negative bias in abundance estimates (Griffin et al. 2013).

Mark-type heterogeneity models have been used to correct estimates of elk (*Cervus canadensis*) abundance (Griffin et al. 2013, Schoenecker and Lubow 2016, Bristow et al. 2019), but they have never been evaluated relative to a known population size. Further, the assumption of a difference in detection between marked and unmarked groups may not always be valid, especially in populations where most groups have a non-zero chance of being detected. In these situations, the mark-type model may lead researchers to infer that heterogeneity is not present, when in fact it is. Alternatively, radio-collars can be thought of as a permanent mark, which Laake et al. (2014) reported can be used to assess the dependency of mark-loss in double-marked animals. Their approach is analogous to estimating dependency of missed observations in a two-sample mark-recapture survey. Consequently, we adapted the mark-loss model to an M_{DS} approach to account for residual heterogeneity by assessing dependency in recapture probability: recapture heterogeneity (M_R).

Both M_H and M_R approaches to accommodating residual heterogeneity depend on having radio-collars in the population. While radio-collars are regularly deployed on many large ungulate populations, researchers have faced challenges in implementing them on feral horses and burros in the United States (Hennig et al. 2020). Concerns over equid safety barred the use of radio-collars on federally managed feral equids for nearly 30 years (Collins et al. 2014); however, researchers recently reported that radio-collars can be deployed on horses and burros with minimal effects (Collins et al. 2014, Schoenecker et al. 2020). The BLM has subsequently permitted radio-collar deployment in some management areas. Consequently, through cooperation with the BLM, we were able to structure the data collection required to quantify residual heterogeneity in aerial surveys of feral equids.

Feral burro populations are increasing in the United States with estimates above appropriate management levels in most management areas (BLM 2021); therefore, the need for unbiased estimates is important. We performed a study with the overarching goals of providing context for evaluating past estimates of burro abundance and helping guide survey designs of future studies with the same goal. Our objectives were to quantify precision and bias of standard M_{DS} surveys for feral burros in the southwestern United States and to examine which model type for incorporating residual heterogeneity (M_H or M_R) would result in the least-biased estimates of burro population size.

STUDY AREA

We performed aerial surveys in the Fort Irwin National Training Center in southern California, the Lake Pleasant Herd Management Area (HMA) in central Arizona, and the Sinbad HMA in central Utah, USA. These areas embody a range of climatic and topographical variation exhibited by burro populations in the western United States. Fort Irwin houses the National Training Center, a military training base initially established for armored vehicle warfare and live-fire training missions. All branches of the military train on the base in a rotational bi-weekly on-range–off-range schedule with model cities and targets established for training. The entire training center encompasses 3,055 km^2, but our study area comprised approximately 1,560 km^2. Fort Irwin is located in the Mojave Desert and is the driest study area. The site has 2 main seasons, a hot, dry season from April through October and a cool, wet season from November through March. Mean 30-year normal precipitation was 136 mm (range = 87–233) and mean temperature was 17.3°C (range = 14.0–21.7; PRISM Climate Group 2020). Topography consists of rugged desert mountain ranges separated by bajadas and low elevation valleys. Mean elevation was 1,011 m (range = 297–1,871; United States Geological Survey [USGS] 2016). Lower elevation areas are dominated by creosote bush (*Larrea tridentata*) and white bursage (*Ambrosia dumosa*), and higher elevation areas contained mixed desert scrub with species such as blackbrush (*Coleogyne ramosissima*), California buckwheat (*Eriogonum fasciculatum*), and Nevada ephedra (*Ephedra nevadensis*). Other mammalian fauna included desert mule deer (*Odocoileus hemionus*), desert bighorn sheep (*Ovis canadensis mexicana*), coyotes (*Canis latrans*), kit foxes (*Vulpes macrotis*), desert cottontails (*Sylvilagus audubonii*), and black-tailed jackrabbits (*Lepus californicus*).

The Lake Pleasant HMA lies within the Sonoran Desert and represents the warmest and wettest study area. Similar to Fort Irwin, there is a hot, dry period between April and October and cooler, wetter period between December and March. It covers 419 km^2 with mean annual 30-year normal precipitation values and temperatures of 335 mm (range = 225–630) and 20.9°C (range = 17.0–22.3), respectively (PRISM Climate Group 2020). Mean elevation is 631 m (range = 375–1,429; USGS 2016). A flat, low elevation area comprises the southern portion of the HMA, while the northern portion is characterized by more mountainous topography. Vegetation species are typical of the Sonoran Desert consisting of Saguaro cactus (*Carnegiea gigantea*), creosote bush, and leguminous trees such as palo verde (*Parkinsonia florida*) and mesquite (*Prosopis* spp.) Mammalian fauna included desert mule deer, javelina (*Tayassu tajacu*), and cattle. Land uses are livestock grazing and human recreation.

Sinbad HMA encompasses 615 km^2 and represents the highest and coldest of the study sites with a mean elevation of 1,799 m (range = 1,310–2,144; USGS 2016) and mean 30-year normal precipitation and temperature values of 236 mm (range = 170–303) and 10.3°C (range = 7.9–12.3), respectively (PRISM Climate Group 2020). Winter typically last from November through March with precipitation falling mainly as snow, and the summer period between April and October is warmer and drier. Topography includes extremely rough areas with incised drainages and flat limestone benches (BLM 2020). Vegetation cover primarily includes pinyon (*Pinus* spp.)-juniper (*Juniperus* spp.) woodland with needle-and-thread (*Hesperostipa comata*) and Indian ricegrass (*Oryzopsis hymenoides*) in open areas (BLM 2020). Other mammalian species included pronghorn (*Antilocapra americana*), bighorn sheep, and kit fox. Main land uses are livestock grazing and energy exploration and extraction.

METHODS

Radio-collars

We deployed radio-collars in the Fort Irwin study area between August 2015 and July 2016, in Lake Pleasant between July 2016 and May 2017, and in Sinbad in April 2016. At Fort Irwin, we fit global positioning system (GPS; TGW-4500-3 store-on-board; Telonics, Mesa, AZ, USA) and very high frequency (VHF) collars (Telonics MOD-500-2) to burros captured via baited corral traps or vehicle-based ground darting (Gedir et al. 2021). We transported captured burros to a nearby BLM facility where we affixed collars before returning them back to the point of capture. At Sinbad and Lake Pleasant HMAs, we fit Iridium (Lotek Wireless, New Market, Canada) and Globalstar collars (Vectronic Aerospace GmbH, Berlin, Germany) to adult female burros (age ≥ 3 yr) with a body condition score >3 (Henneke et al. 1983). The BLM gathered burros using baited traps at Lake Pleasant and a helicopter gather in Sinbad HMA. In Lake Pleasant, we spaced traps across the HMA to distribute collars widely and herded burros into a portable manual squeeze chutes in the field to minimize their movement while we affixed radio-collars. In Sinbad, where burros were helicopter gathered and transported to a handling facility, we randomly selected study individuals by placing collars on every third or fourth adult female burro in the corral line-up until all collars were deployed. The BLM then returned radio-collared burros to the HMA via trailer transport.

Aerial surveys

We followed BLM and USGS standard operating procedures for conducting double-observer aerial surveys for feral horses and burros (Griffin et al. 2020). We used helicopters (Bell 206BIII and Bell 206L4, Bell Textron, Fort Worth, TX, USA; Hughes 500D, MD Helicopters, Mesa, AZ, USA) to fly predefined transects spaced approximately 800 m apart between 30–60 m above ground level. Helicopters flew at a speed of approximately 100 km/hour and carried a 4-person crew consisting of a pilot, 1 front seat observer, and 2 rear seat observers. During the flight, observers maintained audio and visual isolation to independently detect burro groups. Once a group passed astern of the aircraft, observers communicated with one another to record sighting covariates within a 10-m buffer of the group (Griffin et al. 2020). Covariates included the count of individuals within the group, light level (flat, high contrast, or shaded), group activity (moving or still), topographic class (rugged or smooth), visual field type (open, broken, trees), perpendicular distance from the group to the transect line (<100 m, >100–200 m, >200–300 m, >300–400 m, >400 m), proportion of concealing vegetation (to the nearest 0.10), proportion of snow-covered ground (nearest 0.10), and which side of the aircraft the group was on. If needed, observers directed the pilot to circle back to a group to ensure proper recording of covariates. For large groups (>10 individuals), an observer captured a digital photograph, which we used to confirm group sizes after the flight was completed. The front seat observer used a telemetry receiver to assess if a detected group contained a radio-marked individual. After crews flew a section of the survey area (i.e., 3–4 transects, or a topographically discrete area), the front seat observer conducted a scan of VHF signals in the immediate survey area to approximately locate all radio-marked groups. We promptly located and visually observed missed radio-marked groups that were within areas already surveyed to record sighting covariates. We recorded such groups as being detected solely by telemetry.

We conducted 2 flights in Fort Irwin with the first occurring from 4–6 March 2016, and the second taking place 2–4 February 2017 (Gedir et al. 2021). We conducted a complete survey to estimate the population within Fort Irwin in 2017, but high winds prevented us from accomplishing this in 2016. We used detections from the 2016 flight to use in our models, but we did not attempt to estimate the burro abundance for this survey. We completed a survey of the Lake Pleasant HMA between 19–23 June 2017 but also incorporated detections from 2 flights in Lake Pleasant occurring between 30–31 October 2017 to provide extra information for our models. These flights did not cover the entire survey area; thus we do not provide an abundance estimate. At Sinbad, we flew 5 complete surveys occurring between 20–21 January 2016, 15–16 June 2016, 12–13 May 2017, 11–13 October 2017, and 12–13 October 2018. Both of the 2017 surveys contained extra flights conducted over the same area that we used to increase the sample size of detections but not for calculating abundance estimates.

Statistical analyses

We pooled data from all surveys (Schoenecker et al. 2022) and fit detection models with the Huggins (1989, 1991) closed-capture estimator for mark-recapture data with individual covariates in MARK (White and Burnham 1999) using the RMark interface (Laake 2013) within the R statistical software (R Core Team 2021). The Huggins model is similar to many of the other closed capture models in MARK except that abundance (N) is not included in the likelihood with inference conditioned on the number (n) and set of observations. For a 2-occasion Huggins model, the likelihood (L) can be expressed as:

$$L\left(p_1, p_2, c_2 | x_i^{10} x_i^{01}, x_i^{11}, n\right) = \prod_{i=1}^{n} \frac{\left[(p_1(1 - c_2))^{x_i^{10}}((1 - p_1)p_2)^{x_i^{01}}(p_1 c_2)^{x_i^{11}}\right]}{(p_1 + p_2 - p_1 p_2)},$$

where p_1 is the probability observer 1 detects the group, p_2 is the probability observer 2 detects the group given that observer 1 missed it, c_2 is the probability observer 2 detects the group given that observer 1 saw it, and $x_i^{jk} = 1$ if capture history for ith observation is jk (10, 01, or 11) and 0 otherwise. Both under independence (M_{DS}) and for model M_H, $p_2 = c_2$ and in M_R typically $p_2 < c_2$. The likelihood is slightly more complicated for the inclusion of detection of marked animals, which expands the capture history for 3 occasions but is constructed in a similar fashion. The above likelihood is for the unmarked groups of animals, whereas for the m marked groups that are all found by telemetry the likelihood is:

$$L\left(p_1, p_2, c_2 | x_i^{00} x_i^{10} x_i^{01}, x_i^{11}\right) = \prod_{i=1}^{m}\left[((1 - p_1)(1 - p_2))^{x_i^{00}}(p_1(1 - c_2))^{x_i^{10}}((1 - p_1)p_2)^{x_i^{01}}(p_1c_2)^{x_i^{11}}\right].$$

The recapture probability c_2 can only be estimated with the inclusion of marked groups.

Although, not shown in the likelihood formula above, the capture and recapture probabilities can vary across observations based on covariates. We used the logit link to incorporate covariates into the models for initial (p_1, p_2) and recapture probabilities (c_2). Abundance is estimated as:

$$\widehat{N} = \sum_{i=1}^{n}\frac{1}{(p_1 + p_2 - p_1p_2)_i}$$

where the subscript i is used to recognize that the denominator can vary for each observation based on covariate values. Marked groups are simply added to the estimate without error.

We used the DoubleObs (Laake 2021) package in R to handle abundance estimation because MARK does not incorporate group size into estimating abundance within the Huggins model. Also, MARK will not generate the correct abundance estimates under model M_H, but DoubleObs does. When animals are in groups, the total abundance is estimated using:

$$\widehat{N} = \sum_{i=1}^{n}\frac{g_i}{(p_1 + p_2 - p_1p_2)_i} + \sum_{i=1}^{m}g_i$$

where g_i is the size of the ith group. The DoubleObs package also provides model-averaged estimates of abundance.

We first developed a standard M_{DS} model using 2 capture histories coinciding with the pooled detections of the pilot and front seat observer (observer 1), the pooled detections of the rear seat observers (observer 2), and records of sighting covariates. For these models, we withheld all detections made solely with the telemetry unit. This represents the data collected during standard double-observer aerial surveys where radio-collars are absent (Griffin et al. 2020). We used burro groups as the unit of observation to account for the lack of independence of multiple burros in a group; however, some groups consisted of only 1 burro.

We then used a dataset including all detections to develop models incorporating residual heterogeneity. We structured these models to create capture histories with 3 occasions: the telemetry observer, observer 1, and observer 2. Because the telemetry observer had perfect detection of radio-marked groups and no ability to detect unmarked groups, we fixed detection probabilities for the telemetry observer to $p = 1$ for radio-marked groups and $p = 0$ for unmarked groups. If we consider the observer portion of the capture history, the probability structure for the 4 capture histories (00, 10, 01, 11) are $(1 - p_1)(1 - p_2)$, $p_1(1 - c_2)$, $(1 - p_1)p_2$, and p_1c_2. The first capture history is only available for marked groups and for those groups the sum of the probabilities is 1. For unmarked groups, the sum of the probabilities is $p_1 + p_2 - p_1p_2$, which is used to condition on the number of observations in the likelihood (e.g., probability for 10 is $p_1(1 - c_2)/[p_1 + p_2 - p_1p_2]$). We used this model structure for Huggins in MARK to fit models accounting for residual heterogeneity in 3 different ways. The first (M_I) is a model that assumes independence in the initial and recapture probabilities of observer 2 (i.e., no residual heterogeneity; $p_2 = c_2$). Second, we fit the M_H mark-type heterogeneity models by including a covariate corresponding to the marked status of a group to estimate the detection difference between marked and unmarked groups (Griffin et al. 2013). The coefficient (α) for the covariate will be negative for marked groups if it incorporates residual heterogeneity. In this model, recapture probability is still assumed to be the same as initial detection probability of observer 2 ($p_2 = c_2$). Lastly, we fit the M_R recapture heterogeneity model. This is done by including a covariate for c_2 (recapture probability of observer 2) corresponding to whether observer 1 detected a group (1) or not (0). The models are constructed such that the coefficient of the covariate (β) is the difference between c_2 and p_2 on the logit scale (log $(c_2/(1 - c_2)) - \log(p_2/(1 - p_2)) = \beta$). If residual heterogeneity is not present, $\beta = 0$ and $\beta > 0$ if there is residual heterogeneity because $p_2 < c_2$. It may be possible to have a model M_{RH} combining both approaches, but this would require a substantial number of marked groups. Schoenecker et al. (2022) provides the data and the R code used to structure all model types is in Supporting Information (SI.1)

For each of the 4 model types (standard M_{DS} without telemetry detections, M_I independence model with telemetry detections, M_H mark-type heterogeneity, and M_R recapture heterogeneity), we first fit a base model containing only an observer position effect (front vs. rear seat). We then ran univariate models of all sighting covariates and interactions between each sighting covariate and observer position. We ranked models using Akaike's Information Criterion corrected for small sample sizes (AIC_c; Akaike 1998, Burnham and Anderson 2002) to assess which covariates improved model fit relative to the base model. We brought forth all covariates ranked better than the base model to compete in the full model set but kept only the highest-ranked covariate among highly correlated variables ($r > 0.5$). We then ran models containing all combinations of the sighting covariates that indicated better fit than the base model. To obtain abundance estimates per model and survey, we used a modified Horvitz-Thompson estimator, dividing the group size of each detected burro group by its estimated detection probability and summed across all groups detected per survey (Steinhorst and Samuel 1989). For residual heterogeneity models, we applied this correction only to unmarked groups because radio-marked groups were detected without error. We then model-weighted (w_i; Burnham and Anderson 2002) the resulting model-specific abundance estimates to obtain a final abundance estimate per survey.

We evaluated the accuracy of each model by comparing model-generated abundance estimates to the minimum number known alive (MNKA) within Sinbad HMA. The BLM conducted a gather in April 2016, attempting to remove all burros from the HMA and then re-released a subset of these individuals for a concurrent behavioral ecology research study. Not all burros were removed during the capture operation, but subsequent ground-based observations led to reliable identification of those that had remained. Of 236 burros removed, 103 were returned to the HMA, all of which were marked with a unique freeze-brand. Ground observers maintained a constant field presence between March and September 2016–2019 to record individuals seen, with additional observations in winter when weather permitted. Observers identified individuals by freeze-marks or distinct markings in their pelage. While the Sinbad HMA covers a large area (615 km^2), we know from GPS telemetry location data and aerial survey data that burros do not use the entire HMA and tend to concentrate in certain areas. Therefore, we are confident that our calculations of the MNKA per year were close to the true abundances within the HMA.

RESULTS

We detected 610 burro groups across 9 helicopter surveys, 549 of which followed survey protocol and were thus included in detection models. Detections not following protocol ($n = 61$) included situations where a new burro group was spotted while circling to count a previously spotted group, or when there was communication between observers. Though these detections violated survey protocol, the groups were still sighted by >1 observer so we applied model-specific correction factors and included them in abundance estimates. At Fort Irwin, we detected 79 burro groups in the March 2016 survey and 149 groups in the February 2017 survey, with observers detecting 2 of 8 and 7 of 13 radio-marked groups, respectively. At the Lake Pleasant HMA, we detected 81 groups during the June 2017 survey, but radio-collars were not present at the time of this survey. In October 2017, we detected 108 groups and 12 of 28 radio-marked groups. We detected 32 groups in the January 2016 survey of the Sinbad HMA, but no radio-marked groups were present. We recorded 26, 42, 58, and 35 detections during the June 2016, May 2017, October 2017, and October 2018 surveys, respectively. Of the radio-marked groups present, observers detected 8 of 17, 12 of 18, 23 of 31, and 7 of 10, respectively. Across all surveys, human observers detected 78 of the 125 (62.4%) radio-marked groups.

Of the residual heterogeneity models, the M_R model performed best, with the top-ranked M_R model 11.43 AIC_c units lower than the top-ranked M_H model (Table 1). The difference between the M_R model and the other models is depicted by lower initial detection probabilities, whereas the recapture probability (c_2) is on par with the initial detection probabilities of observer 2 (p_2) estimated by the other models (Figure 1; Table 2). The lower mean detection estimated in M_R models translated into higher abundance estimates (Table 3).

TABLE 1 Model-selection information for estimating feral burro abundances from helicopter surveys conducted within the National Training Center, Fort Irwin, California, Lake Pleasant Herd Management Area, Arizona, and Sinbad Herd Management Area, Utah, USA, 2016–2018. Model sets included a standard double-observer sightability model (M_{DS}), and 3 sets that included detections of radio-marked individuals: M_I models assuming no residual heterogeneity in detection probability (independence), M_H models accounting for residual heterogeneity as the difference in detection between marked and unmarked groups (mark-type heterogeneity), and M_R models accounting for residual heterogeneity through recapture probability (recapture heterogeneity). Columns represent the number of parameters (K), Akaike's Information Criterion adjusted for small sample sizes (AIC_c), difference between a given model and the top-ranked model within a set (ΔAIC_c), and model weight (w_i). Models with $\Delta AIC_c \leq 4$ along with a base model are shown.

Model[a]	K	AIC_c	ΔAIC_c	w_i
M_{DS} models				
Pos + site + size + conceal + dist:fr + dist:r + topo:fr + topo:r + flat:r + ps:fr	11	990.86	0.00	0.67
Pos + site + size + conceal + topo:fr + topo:r + dist:fr + dist:r + ps:fr	10	992.78	1.92	0.25
Pos + site	3	1,068.07	77.21	0.00
M_I models				
Pos + site + size + conceal + dist:fr + dist:r + topo:f + topo:r + flat:r + ps:fr	11	1,138.34	0.00	0.57
Pos + site + size + conceal + dist:fr + dist:r + topo:f + topo:r + ps:fr	10	1,139.60	1.26	0.30
Pos + site + size + conceal + dist:fr + dist:r + topo:f + topo:r + flat:r	10	1,142.15	3.81	0.08
Pos + site	3	1,257.56	119.22	0.00
M_H models				
Pos + type + site + size + conceal + dist:fr + dist:r + topo:r + topo:r + flat:r + ps:fr	12	1,140.19	0.00	0.57
Pos + type + site + size + conceal + dist:fr + dist:r + topo:fr + topo:r + ps:fr	11	1,141.46	1.27	0.30
Pos + type + site + size + conceal + dist:fr + dist:r + topo:fr + topo:r + flat:r	11	1,144.04	3.85	0.08
Pos + type + site	4	1,257.45	117.26	0.00
M_R models				
Pos + c:het:r + site + size + conceal + dist:fr + dist:r + topo:fr + topo:r + flat:r + ps:fr	12	1,128.76	0.00	0.62
Pos + c:het:r + site + size + conceal + dist:fr + dist:r + topo:fr + topo:r + ps:fr	11	1,130.65	1.89	0.24
Pos + c:het:r + site + size + conceal + dist:fr + dist:r + topo:fr + topo:r + flat:f	11	1,132.49	3.73	0.10
Pos + c:het:r + site	4	1,233.30	104.54	0.00

[a]pos = observer position (front vs. rear seat), site = site effect (Fort Irwin vs. Sinbad or Lake Pleasant), size = natural logarithm of group size, conceal = proportion of concealing vegetation within 10 m of detected group, dist:fr = distance effect on front observer, dist:r = distance effect on rear observer, topo:fr = rugged topographic class effect on front observer, topo:r = rugged topographic class effect on rear observer, flat:r = flat lighting class effect on rear observer, ps = pilot side effect on front observer, type = effect of mark type (marked vs. unmarked), c:het:rear = effect of the front observer capture history (het) applied to recapture probability (c) and rear seat capture history.

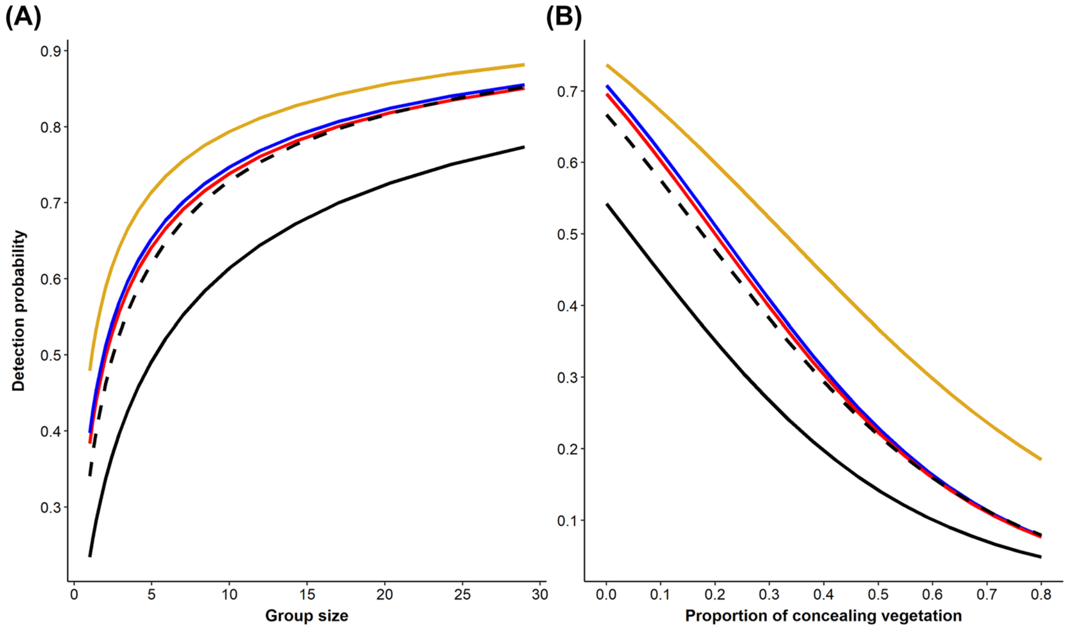

FIGURE 1 Estimates of rear seat observer detection probability (p_2) from double-observer sightability models as a function of burro group size (A) and the proportion of concealing vegetation within 10 m of a detected group (B). Black lines represent estimates from recapture heterogeneity (M_R) models incorporating residual heterogeneity in detection through recapture probability (solid = initial detection probability [p_2]; dashed = recapture probability [c_2]), blue lines represent estimates from independence (M_I) models assuming independence in detection among observers, red lines correspond to estimates from mark-type heterogeneity (M_H) models accounting for residual heterogeneity as the difference in detection between marked and unmarked groups, and gold lines represent standard double-observer sightability (M_{DS}) models. We pooled data from helicopter surveys conducted in the National Training Center, Fort Irwin, California, Lake Pleasant Herd Management Area, Arizona, and Sinbad Herd Management Area, Utah, USA, 2016–2018.

TABLE 2 Model-averaged mean detection probability estimates and standard errors for each parameter in different models estimating detection of burros across helicopter surveys in National Training Center, Fort Irwin, California, Lake Pleasant Herd Management Area, Arizona, and Sinbad Herd Management Area, Utah, USA, 2016–2018. Model types included a standard double-observer sightability model (M_{DS}), and 3 model sets that included detections of radio-marked individuals: M_I models assuming no residual heterogeneity in detection probability (independence), M_H models accounting for residual heterogeneity as the difference in detection between marked and unmarked groups (mark-type heterogeneity), and M_R models accounting for residual heterogeneity through recapture probability (recapture heterogeneity).

Parameter	M_{DS}	M_I	M_H	M_R
Front seat (p_1)	0.52 (0.03)	0.46 (0.03)	0.47 (0.03)	0.39 (0.03)
Rear seat (p_2)	0.67 (0.03)	0.60 (0.03)	0.61 (0.03)	0.43 (0.06)
Front seat (p_1)[a]			0.45 (0.04)	
Rear seat (p_2)[a]			0.59 (0.04)	
Recapture (c_2)				0.56 (0.03)

[a]Detection probability of radio-marked groups.

TABLE 3 Mean estimate of overall detection probability accounting for both observers (*p*), raw counts, abundance estimates, minimum number of known alive (MNKA) burros, standard errors (SE), coefficients of variation (CV), and lower (LCI) and upper (UCI) 95% confidence intervals from 7 helicopter surveys of feral burros in National Training Center, Fort Irwin, California, Lake Pleasant Herd Management Area, Arizona, and Sinbad Herd Management Area, Utah, USA, 2016–2018. We generated estimates using a standard double-observer sightability model (M_{DS}), and 3 model sets that included detections of radio-marked individuals: M_I models assuming no residual heterogeneity in detection probability (independence), M_H models accounting for residual heterogeneity as the difference in detection between marked and unmarked groups (mark-type heterogeneity), and M_R models accounting for residual heterogeneity through recapture probability (recapture heterogeneity).

Survey	*p*	Raw count	Estimate	MNKA	SE	CV	LCI	UCI
M_{DS} models								
Fort Irwin – Feb 2017	0.79	573	680		34	0.05	631	771
Lake Pleasant – Jun 2017[a]	0.81	379	453		29	0.06	415	533
Sinbad – Jan 2016[a]	0.86	123	138	236	10	0.07	128	173
Sinbad – Jun 2016	0.85	73	92	136	14	0.14	78	141
Sinbad – May 2017	0.86	102	117	159	8	0.06	108	141
Sinbad – Oct 2017	0.78	128	148	182	12	0.08	135	188
Sinbad – Oct 2018	0.82	145	178	213	18	0.10	157	236
M_I models								
Fort Irwin – Feb 2017	0.76	602	723		37	0.05	670	818
Lake Pleasant – Jun 2017[a]	0.76	379	495		40	0.08	439	601
Sinbad – Jan 2016[a]	0.83	123	146	236	14	0.09	131	190
Sinbad – Jun 2016	0.82	118	125	136	6	0.06	120	149
Sinbad – May 2017	0.81	116	130	159	8	0.06	121	154
Sinbad – Oct 2017	0.71	143	168	182	15	0.09	151	216
Sinbad – Oct 2018	0.78	158	186	213	16	0.08	168	236
M_H models								
Fort Irwin – Feb 2017	0.75	602	734		48	0.07	669	865
Lake Pleasant – Jun 2017[a]	0.75	379	500		43	0.09	441	616
Sinbad – Jan 2016[a]	0.82	123	148	236	15	0.10	131	195
Sinbad – Jun 2016	0.81	118	125	136	6	0.05	120	150
Sinbad – May 2017	0.80	116	131	159	8	0.06	122	158
Sinbad – Oct 2017	0.70	143	169	182	16	0.09	152	220
Sinbad – Oct 2018	0.77	158	187	213	16	0.09	168	239
M_R models								
Fort Irwin – Feb 2017	0.62	602	855		85	0.10	736	1,080
Lake Pleasant – Jun 2017[a]	0.66	379	571		64	0.11	480	742
Sinbad – Jan 2016[a]	0.73	123	165	236	21	0.13	140	230
Sinbad – Jun 2016	0.72	118	130	136	9	0.07	121	163

(Continues)

TABLE 3 (Continued)

Survey	*p*	Raw count	Estimate	MNKA	SE	CV	LCI	UCI
Sinbad – May 2017	0.71	116	143	159	13	0.09	127	182
Sinbad – Oct 2017	0.61	143	184	182	22	0.12	158	253
Sinbad – Oct 2018	0.68	158	204	213	22	0.11	176	271

[a]No radio-collars present in population at time of survey.

TABLE 4 Estimates and standard errors of beta coefficients found in the top-ranked models for estimating detection probability of burros across helicopter surveys in National Training Center, Fort Irwin, California, Lake Pleasant Herd Management Area, Arizona, and Sinbad Herd Management Area, Utah, USA, 2016–2018. Model types included a standard double-observer sightability model (M_{DS}), and 3 model sets that included detections of radio-marked individuals: M_I models assuming no residual heterogeneity in detection probability (independence), M_H models accounting for residual heterogeneity as the difference in detection between marked and unmarked groups (mark-type heterogeneity), and M_R models accounting for residual heterogeneity through recapture probability (recapture heterogeneity).

Variable[a]	M_{DS}	M_I	M_H	M_R
Intercept	1.00 (0.31)	0.97 (0.29)	0.90 (0.33)	0.48 (0.35)
Rear	−0.79 (0.26)	−0.66 (0.25)	−0.67 (0.25)	−1.11 (0.26)
Site	−0.96 (0.23)	−0.96 (0.22)	−0.97 (0.22)	−1.00 (0.24)
Size	0.62 (0.13)	0.65 (0.12)	0.65 (0.12)	0.71 (0.13)
Conceal	−3.16 (0.84)	−4.21 (0.74)	−4.16 (0.74)	−3.94 (0.78)
Dist:fr	−1.56 (0.48)	−1.80 (0.45)	−1.80 (0.45)	−1.56 (0.49)
Dist:r	1.07 (0.61)	0.44 (0.51)	0.47 (0.51)	0.69 (0.52)
Topo:fr	−1.43 (0.33)	−1.45 (0.28)	−1.43 (0.28)	−1.37 (0.29)
Topo:r	−0.73 (0.38)	−0.84 (0.28)	−0.81 (0.29)	−0.55 (0.29)
Flat:r	0.55 (0.28)	0.44 (0.25)	0.44 (0.25)	0.46 (0.23)
Ps:fr	−0.52 (0.20)	−0.46 (0.20)	−0.46 (0.19)	−0.45 (0.19)
Type			0.08 (0.20)	
C:het:r				0.93 (0.27)

[a]Rear = effect of rear seat observer position, Site = site effect (Fort Irwin vs. Sinbad or Lake Pleasant), Size = natural logarithm of group size, Conceal = proportion of concealing vegetation within 10 m of detected group, Dist:fr = distance effect on front observer, Dist:r = distance effect on rear observer, Topo:fr = rugged topographic class effect on front observer, Topo:r = rugged topographic class effect on rear observer, Flat:r = flat lighting class effect on rear observer, Ps:fr = pilot side effect on front observer, Type = effect of mark type (marked vs. unmarked), C:het:rear = effect of the front observer capture history (het) applied to recapture probability and rear seat capture history.

On average, M_R models resulted in abundance estimates 19.2%, 8.3%, and 7.8% greater than standard M_{DS}, M_H, and M_I model estimates, respectively. Recapture heterogeneity models (M_R) also produced the least biased estimates when compared to the surveys for which we knew MNKA (Table 3). The M_R model estimates ranged from 30.1% below to 1.1% above MNKA ($\bar{x}$ = 9.5% below). The M_H model estimates ranged from 7.1–37.2% ($\bar{x}$ = 16.4%) below MNKA, the M_I estimates ranged from 7.7–38.1% ($\bar{x}$= 17.0%) below MNKA, and the standard M_{DS} ranged from 16.4–41.5% ($\bar{x}$ = 27.1%) below MNKA. The M_R model produced 95% confidence intervals that encompassed the MNKA in 4 of 5 surveys (Table 3), compared to 95% confidence intervals capturing MNKA in only 3 of 5 surveys for the other models. Average precision, measured by the mean coefficient of variation, was 0.07 for the M_I model, 0.08 for the M_{DS} and M_H models, and 0.10 for the M_R models.

The sighting covariates included in the top models were identical across all model types (Tables 1 and 4). Detection probability predictably increased with larger group sizes and decreased with greater proportions of concealing vegetation surrounding a burro group (Table 4; Figure 1). Detection probabilities averaged higher for rear versus front seat observers (Table 2). The coefficient for rear seat observers was negative for all model types (Table 4), but this was offset by rear seat observers being better than front seat observers at detecting burro groups at farther distances, in rugged topography, and in flat lighting (Table 4). Moreover, detection probability by front seat observers declined when burro groups were on the pilot's side of the aircraft (Table 4). Detection was also lower in Fort Irwin compared to the Lake Pleasant and Sinbad HMA study areas (Table 4).

DISCUSSION

Unbiased abundance estimates of feral burros are imperative for guiding management plans that promote the balance of healthy herds and sustainable ecosystems. Our results indicated that standard M_{DS} models used for burros can achieve estimates on par in accuracy and precision with M_{DS} surveys of other ungulate species (Griffin et al. 2013, Lubow and Ransom 2016). Nonetheless, standard M_{DS} estimates of burros were consistently biased low with respect to MNKA. Estimates from M_{DS} models were also more biased than the estimates from models that used detections of radio-collars to account for residual heterogeneity. Radio-marking burros provided 3 main benefits. First, the telemetry observer can locate groups missed by all human observers. Second, detections made by the telemetry observer can be included without error in abundance estimates, thereby increasing survey precision. Third, including telemetry as an additional capture history allowed us to account for residual heterogeneity in detection probability and decrease bias in abundance estimates.

The M_R recapture heterogeneity model outperformed the previously used M_H model in terms of AIC_c support, less-biased estimates, and frequency of 95% confidence intervals encompassing MNKA. The M_R model confirmed that residual heterogeneity indeed existed, while the M_H model implied that it did not. The top-ranked M_H model was nearly 2 AIC_c units worse than the top-ranked M_I model, and because the M_H model only differs from the M_I model by the inclusion of the mark-type covariate, little additional variation was explained by that covariate's inclusion. Compared to estimates from M_R, the initial detection estimates were correspondingly biased high in the M_H model, leading to negatively biased estimates of abundance. A covariate of mark-type may be warranted in populations where several groups have zero chance of being detected by human observers (i.e., a lone individual completely obscured by vegetation), but our results indicated little presence of such instances in our dataset.

Our study builds upon analyses by Griffin (2015) and Gedir et al. (2021) by examining sighting covariates that most strongly influence burro detection across multiple populations. A set of 8 covariates explained a large proportion of the variation in detection probabilities among burro groups. Larger group sizes, regardless of species, are easier to detect (Udevitz et al. 2006, Ransom 2012, Griffin 2015, Gedir et al. 2021), and this was further corroborated by our study. Predictably, detection declined with greater proportions of concealing vegetation cover surrounding burro groups. Our data revealed evidence of interactions between observer position and 4 different covariates. Rear seat observers were better than front seat observers at detecting burro groups in rugged topography, in flat lighting, and at farther distances. A site effect of Fort Irwin compared to Sinbad and Lake Pleasant indicated that detection probability was lower in Fort Irwin, possibly because observers had less survey experience or because of other unexplained heterogeneity not captured by the information provided by missed radio-marked groups.

As was true in 5 of 6 data sets in Griffin (2015), regardless of model type, rear seat observers had higher detection probabilities than front seat observers. We expected this result given that helicopter surveys of bison (*Bison bison*), burros, and elk have also reported greater detection by rear seat observers (Griffin 2015, Schoenecker and Lubow 2016, Bristow et al. 2019, Gedir et al. 2021, Hennig et al. 2021). This result may seem counterintuitive given the larger field of view presented to front seat observers, but a larger field of view also translates into searching a larger area; thus, front seat observers in those studies and in this study spend more effort scanning a wide area closer to the aircraft and need to identify burros from multiple angles compared to rear seat observers who can fully direct their attention to picking out burros in a narrower but longer viewshed. The pilot's foremost role is to fly the aircraft safely; it is well established that observations on the pilot's side of the aircraft have lower front seat detection probabilities (Lubow and Ransom 2016). Finally, for surveys such as these in which the front seat observers were also tasked with using telemetry receivers, that may have contributed to increased distraction during surveys.

The inclusion of radio-marked groups helped to decrease the amount of bias in abundance estimates, yet there was consistent negative bias across surveys. The January 2016 survey in the Sinbad HMA produced particularly biased estimates. Survey practitioners noted snow cover on the ground with contrasting dark-colored conifers, which made for difficult sighting conditions. Further, there were no radio-marked burros in the population at this time, preventing us from gaining information from undetected groups in this survey. Other sources of bias could be attributed to observer-specific differences in sighting probability. Observer acuity can significantly affect detection bias (Shirley et al. 2012), but we did not include detection parameters for each observer because our pooled dataset included 17 unique observers and 5 different pilots, rendering their inclusions impractical. Undercounting is a potential source of bias, but we believe this to be a small source of bias because group sizes were small (90% of observations had <10 individuals) and we used photographs to reconcile counts of large groups. The undercounting of group size is likely on the scale of 1–2 individual burros rather than tens of individuals in species with larger group sizes (i.e., horses, elk; Lubow and Ransom 2016, Schoenecker and Lubow 2016). Some radio-collared individuals may go undetected during surveys. Our study design assumes perfect detection of groups by the telemetry observer, but some radio signals may not be picked up in rugged topography or because of technological issues; thus, information from these groups cannot be included. Further, our model assumes that all radio-collared groups are perfectly detected by the telemetry unit, and we accomplished this in our study. If radio-collared groups were detected by human observers but not the telemetry unit, then the model would have to be modified. Finally, it may be possible to use mark-recapture distance sampling (Laake and Borchers 2004, Burt et al. 2014) in this context, but it would require additional development to incorporate the marked animals and was beyond the scope of this study. This framework would allow for a non-linear response of detection to distance, which could remove the remaining negative bias; thus; it is worth future investigation.

Bias is usually unknown in population estimates. For this reason, the goal of surveys is typically to maximize precision (commonly measured by CV) because routinely collected, precise estimates facilitate detection of population fluctuations (Cochran 1977, Hodgson et al. 2016). The M_H models developed for elk in Mount Rainier National Park and in the San Luis Valley of Colorado, USA, resulted in mean coefficients of variation of 0.08 and 0.09, respectively (Griffin et al. 2013, Schoenecker and Lubow 2016). In one study, M_{DS} models developed for feral horses using detections from 4 different management areas resulted in a mean coefficient of variation of 0.15 (range = 0.02–0.25; Lubow and Ransom 2016). All of our model sets produced coefficient of variation estimates within that range reported for horses, suggesting that current BLM aerial survey protocols can be considered effective for monitoring burro populations. Nonetheless, decision-makers may put greater value on unbiased estimates because of the federal legislation that motivates these surveys, and standard M_{DS} model estimates were routinely biased low. While M_R models had the greatest mean coefficient of variation across surveys, the mean paired difference in coefficient of variation per survey was low (M_R vs. M_{DS} = 0.02; M_R vs. M_I = 0.03; M_R vs. M_H = 0.03). This minor difference in survey precision was offset by the least-biased abundance estimates; thus, we recommend M_R models for use in future burro surveys.

MANAGEMENT IMPLICATIONS

Double-observer sightability models can result in relatively precise abundance estimates of burros; therefore, we recommend maintaining standard operating procedures for conducting aerial surveys. Managers should recognize that dependency in detection among observers leads to negatively biased estimates, but this can be modeled by including detections of radio-marked groups. Deploying radio-collars ostensibly increases survey costs and requires additional effort to capture and collar individuals; therefore, managers should weigh whether decreasing bias is worth the additional resources. Future surveys will be improved with more detection data of radio-marked burros, however, so we encourage radio-collar deployment in as many management areas as possible.

ACKNOWLEDGMENTS

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the United States Government. We thank observers and air crew members (K. Birch, K. Bolton, K. D. Bristow, J. Carlisle, D. Carpenter, M. L. Crabb, J. Duncan, L. S. Ekernas, P. C. Griffin, J. Hall, L. E. Harding, C. Johnson, T. Karish, J. Kelly, T. Lyons, R. Reed, C. D. Reddell, S. Shaffer, T. Staheli, E. Thul, M. Twedell, S. Wainsgard, J. Wignall, M. Worthington, A. Yemma) for assistance with data collection, S. R. B. King for help calculating minimum number of known burros alive, and L. S. Ekernas and B. C. Lubow for help with data analyses. We thank L. M. Aker, and C. A. Everly with the National Training Center Fort Irwin Directorate of Public Works and P. C. Griffin and A. Neibergs with BLM for logistical support. This project was funded by the BLM (Interagency agreement L19PG00052), the United States Department of Defense, Fort Irwin (Military Interdepartmental Purchase Request 10703187), and the United States Geological Survey, Fort Collins Science Center.

CONFLICTS OF INTEREST

We declare no conflicts of interest.

ETHICS STATEMENT

Capture and radio-collaring procedures adhered to protocols set forth in the USGS Fort Collins Science Center (FORT IACUC 2015-10) and by the New Mexico State University Institutional Animal Care and Use Committee (permit 2015-002).

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available via the following link: https://doi.org/10.5066/P9OAEATC. The R code used to structure all model types is available in the Supporting Information (SI.1).

ORCID

Jacob D. Hennig http://orcid.org/0000-0001-6450-6906
Kathryn A. Schoenecker http://orcid.org/0000-0001-9906-911X
James W. Cain http://orcid.org/0000-0003-4743-516X

REFERENCES

Akaike, H. 1998. Information theory and an extension of the maximum likelihood principle. Selected Papers of Hirotugu Akaike. Springer, New York, New York, USA.

Barker, R. J., D. M. Forsyth, and M. Wood. 2014. Modeling sighting heterogeneity and abundance in spatially replicated multiple-observer surveys. Journal of Wildlife Management 78:701–708.

Becker, E. F., and A. M. Christ. 2015. A unimodal model for double observer distance sampling surveys. PLoS ONE 10: e0136403.

Borchers, D. L., J. L. Laake, C. Southwell, and C. G. M. Paxton. 2006. Accommodating unmodeled heterogeneity in double-observer distance sampling surveys. Biometrics 62:372–378.

Bristow, K. D., M. J. Clement, M. L. Crabb, L. E. Harding, and E. S. Rubin. 2019. Comparison of aerial survey methods for elk in Arizona. Wildlife Society Bulletin 43:77–92.
Buckland, S. T., J. L. Laake, and D. L. Borchers. 2010. Double-observer line transect methods: levels of independence. Biometrics 66:169–177.
Bureau of Land Management [BLM]. 2010. Wild horse and burro population inventory and estimation. Instruction Memorandum 2010-057. Bureau of Land Management, Washington, D.C., USA.
Bureau of Land Management [BLM]. 2020. Herd Management Areas. http://www.blm.gov/programs/wild-horse-and-burro/herd-management/herd-management-areas. Accessed 20 Nov 2020.
Bureau of Land Management [BLM]. 2021. Program data. https://www.blm.gov/programs/wild-horse-and-burro/about-the-program/program-data. Accessed 5 Aug 2021.
Burnham, K. P., and D. R. Anderson. 2002. Model selection and multi-model inference: a practical information-theoretic approach. Second edition. Springer-Verlag, New York, New York, USA.
Burt, M. L., D. L. Borchers, K. J. Jenkins, and T. A. Marques. 2014. Using mark-recapture distance sampling methods on line transect surveys. Methods in Ecology and Evolution 5:1180–1191.
Cochran, W. G. 1977. Sampling techniques. Third edition. John Wiley and Sons, New York, New York, USA.
Collins, G. H., S. L. Petersen, C. A. Carr, and L. Pielstick. 2014. Testing VHF/GPS collar design and safety in the study of free-roaming horses. PLoS ONE 9:e103189.
Gedir, J. V., J. W. Cain, III, B. C. Lubow, T. Karish, D. K. Delaney, and G. W. Roemer. 2021. Estimating abundance and simulating fertility control in a feral burro population. Journal of Wildlife Management 85:1187–1199.
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. 2015. Estimated abundance of wild burros surveyed on Bureau of Land Management Lands in 2014. U.S. Geological Survey Open-File Report 2015–1084, Reston, Virginia, USA.
Griffin, P. C., L. S. Ekernas, K. A. Schoenecker, and B. C. Lubow. 2020. Standard operating procedures for wild horse and burro double-observer aerial surveys. U.S. Geological Survey Techniques and Methods 2-A16, Reston, Virginia, USA.
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. Schaberl, and K. Beirne. 2013. A hybrid double-observer sightability model for aerial surveys. Journal of Wildlife Management 77:1532–1544.
Henneke, D. R., G. D. Potter, J. L. Kreider, and B. F. Yeates. 1983. Relationship between condition score, physical measurements and body fat percentage in mares. Equine Veterinary Journal 15:371–372.
Hennig, J. D., J. D. Scasta, J. L. Beck, K. A. Schoenecker, and S. R. B. King. 2020. Systematic review of equids and telemetry collars: implications for deployment and reporting. Wildlife Research 47:361–371.
Hennig, J. D., K. A. Schoenecker, M. L. N. Terwilliger, G. W. Holm, and J. L. Laake. 2021. Comparison of aerial thermal infrared imagery and helicopter surveys of bison (*Bison bison*) in Grand Canyon National Park. Sensors 21:5087.
Hodgson, J. C., S. M. Baylis, R. Mott, A. Herrod, and R. H. Clarke. 2016. Precision wildlife monitoring using unmanned aerial vehicles. Scientific Reports 6:22574.
Huggins, R. M. 1989. On the statistical analysis of capture-recapture experiments. Biometrika 76:133–140.
Huggins, R. M. 1991. Some practical aspects of a conditional likelihood approach to capture experiments. Biometrics 47: 725–732.
Laake, J. L. 2013. RMark: an R interface for analysis of capture-recapture data with MARK. AFSC Processed Rep 2013-01. Alaska Fisheries Science Center, National Oceanic Atmospheric Administration, National Marine Fisheries Service, Seattle, Washington, USA.
Laake, J. L. 2021. DoubleObs: analysis of double-observer survey data. R package version 0.1.3. https://github.com/jlaake/DoubleObs. Accessed 21 Jun 2021.
Laake, J. L., and D. L. Borchers. 2004. Methods for incomplete detection at distance zero. Pages 108–189 *in* S. T. Buckland, D. R. Anderson, K. P. Burnham, J. L. Laake, D. L. Borchers, and L. Thomas, editors. Advanced distance sampling. Oxford University Press, Oxford, United Kingdom.
Laake, J. L., D. S. Johnson, D. R. Diefenbach, and M. A. Ternent. 2014. Hidden Markov model for dependent mark loss and survival estimation. Journal of Agricultural, Biological, and Environmental Statistics 19:522–538.
Lubow, B. C., and J. I. Ransom. 2016. Practical bias correction in aerial surveys of large mammals: validation of hybrid double-observer with sightability methods against known abundance of feral horse (*Equus caballus*) population. PLoS ONE 11:e0154902.
National Research Council. 2013. Using science to improve the BLM Wild Horse and Burro Program: a way forward. The National Academies Press, Washington, D.C., USA.
Otis, D. L., K. P. Burnham, G. C. White, and D. R. Anderson. 1978. Statistical inference from capture data on closed animal populations. Wildlife Monographs 62:3–135.
PRISM Climate Group. 2020. PRISM climate data. http://prism.oregonstate.edu. Accessed 20 Nov 2020.

Ransom, J. I. 2012. Detection probability in aerial surveys of feral horses. Journal of Wildlife Management 76:299–307.

R Core Team. 2021. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Samuel, M. D., and K. H. Pollock. 1981. Correction of visibility bias in aerial surveys where animals occur in groups. Journal of Wildlife Management 45:993–997.

Schoenecker, K. A., J. D. Hennig, J. W. Cain, III, and G. W. Roemer. 2022. Detections of burros from helicopter aerial surveys in the southwestern US, 2016–2018. U.S. Geological Survey data release, Fort Collins Science Center, Fort Collins, Colorado, USA. https://doi.org/10.5066/P9OAEATC

Schoenecker, K. A., S. R. B. King, and G. H. Collins. 2020. Evaluation of the impacts of radio-marking devices on feral horses and burros in a captive setting. Human–Wildlife Interactions 14:73–86.

Schoenecker, K. A., and B. C. Lubow. 2016. Application of a hybrid model to reduce bias and improve precision in population estimates for elk (*Cervus elaphus*) inhabiting a cold desert ecosystem. Journal of King Saud University–Science 28:205–215.

Shirley, M. H., R. M. Dorazio, E. Abassery, A. A. Elhady, M. S. Mekki, and H. H. Asran. 2012. A sampling design and model for estimating abundance of Nile crocodiles while accounting for heterogeneity of detectability of multiple observers. Journal of Wildlife Management 76:966–975.

Southwell, C., D. Borchers, C. G. M. Paxton, L. Burt, and W. de la Mare. 2007. Estimation of detection probability in aerial surveys of Antarctic pack-ice seals. Journal of Agricultural, Biological, and Environmental Statistics 12:41–54.

Steinhorst, R. K., and M. D. Samuel. 1989. Sightability adjustment methods for aerial surveys of wildlife populations. Biometrics 45:415–425.

Udevitz, M. S., B. S. Shults, L. G. Adams, and C. Kleckner. 2006. Evaluation of aerial survey methods for Dall's sheep. Wildlife Society Bulletin 34:732–740.

U.S. Geological Survey [USGS]. 2016. 3DEP products and services: The National Map, 3D Elevation Program Web page. http://nationalmap.gov/3DEP/3dep_prodserv.html. Accessed 20 Nov 2020.

Walter, M. J., and J. Hone. 2003. A comparison of 3 aerial survey techniques to estimate wild horse abundance in the Australian Alps. Wildlife Society Bulletin 31:1138–1149.

White, G. C., and K. P. Burnham. 1999. Program MARK: survival estimation from populations of marked animals. Bird Study 46:120–138.

Associate Editor: Andrew Norton.

SUPPORTING INFORMATION

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

How to cite this article: Hennig, J. D., K. A. Schoenecker, J. W. Cain, III, G. W. Roemer, and J. L. Laake. 2022. Accounting for residual heterogeneity in double-observer sightability models decreases bias in burro abundance estimates. Journal of Wildlife Management 86:e22239. https://doi.org/10.1002/jwmg.22239