Tools and Technology Article

Validating Aerial Photographic Mark–Recapture for Naturally Marked Feral Horses

BRUCE C. LUBOW, *Natural Resource Ecology Laboratory, Colorado State University, Fort Collins, CO 80523, USA*
JASON I. RANSOM,[1] *United States Geological Survey, Fort Collins Science Center, 2150 Centre Avenue, Building C, Fort Collins, CO 80526, USA*

ABSTRACT Accurately estimating large mammal populations is a difficult challenge because species of interest often occupy vast areas and exhibit low and heterogeneous visibility. Population estimation techniques using aerial surveys and statistical design and analysis methods provide a means for meeting this challenge, yet they have only rarely been validated because wild populations of known size suitable for field tests are rare. Our study presents field validations of a photographic aerial mark–recapture technique that takes advantage of the recognizable natural markings on free-roaming feral horses (*Equus caballus*) to accurately identify individual animals and groups of animals sighted on multiple occasions. The 3 small populations of feral horses (<400 animals each) in the western United States used in the study were all closely monitored on a weekly basis by local researchers, thus providing test populations of known size. We were able to accurately estimate these population sizes with aerial surveys, despite rugged terrain and dense vegetation that created substantial heterogeneity of sighting probability among horse groups. Our best estimates at the 3 sites were within −6.7%, 2.6%, and −8.6% of known truth (−4.2% mean error, 6.0% mean absolute error). In contrast, we found undercount bias as large as 32% before any statistical corrections. The necessary corrections varied both temporally and spatially, in response to previous sighting history (behavioral response), and by the number of horses in a group. Despite modeling some of the differences in horse-group visibility with sighting covariates, we found substantial residual unmodeled heterogeneity that contributed to underestimation of the true population by as much as 22.7% when we used models that did not fully account for these unmeasured sources. We also found that the cost of the accurate and validated methods presented here is comparable to that of raw count (so called, census) methods commonly employed across feral horse ranges in 10 western states. We believe this technique can assist managers in accurately estimating many feral horse populations and could be applied to other species with sufficiently diverse and distinguishable visible markings. (JOURNAL OF WILDLIFE MANAGEMENT 73(8):1420–1429; 2009)

DOI: 10.2193/2008-538

KEY WORDS aerial survey, *Equus caballus*, feral horse, heterogeneity, mark–recapture, population estimation, sighting probability, validation.

Accurately estimating population sizes of large free-roaming animals is a challenging and critical task for successful wildlife management (Williams et al. 2002), yet up to one third of ungulates in the western United States are missed by standard visual aerial surveys (Samuel et al. 1987, Ackerman 1988, Singer and Garton 1994, Bodie et al. 1995, Bowden and Kufeld 1995). Visibility of ungulates can vary tremendously among survey sites and occasions, depending upon transect spacing and sighting factors such as snow cover, group size, activity of the animals, tree cover, and experience of the observers (Pollock and Kendall 1987, Samuel et al. 1987, Unsworth et al. 1994, Bodie et al. 1995, Lubow and Ransom 2007). Despite these well-known biases that result in variable and unknown degrees of undercounting, the use of so-called census methods that make the unjustified assumption of 100% sighting probability remains commonplace (Rabe et al. 2002), while published evaluations of population estimation methods for feral horses (*Equus caballus*) are rare. Modern survey methods based on statistical models have been applied in Australia (Bayliss and Yeomans 1989, Graham and Bell 1989, Walter and Hone 2003, Laake et al. 2008), and recently Lubow and Ransom (2007) applied a technique to a North American feral horse population. All of these studies employed the simultaneous double-count method of mark–recapture and were limited by incorporating only 2 occasions (mark and recapture sightings), making testing for and correcting of biases due to unmodeled heterogeneity impossible. Furthermore, none of these prior studies was able to validate the methods in a population of accurately known size.

Our study focuses on a form of mark–recapture sampling technique that adjusts for sightability bias similar to a sightability bias correction model. Mark–recapture methods do not necessarily require physically capturing animals, only that individual animals or coherent groups can be reliably identified by natural or artificial marks or other unique characteristics on ≥ 2 occasions; individual capture histories can thus be recorded and used to estimate the number of unobserved animals (Seber 1973; Huggins 1989, 1991; Neal et al. 1993; Pledger 2000). Repeated observations can be made simultaneously by multiple observers or at different times. Sighting heterogeneity among individual animals or groups is common; some groups, due to their size, distance, coloration, location within cover, or other factors, are easier or more difficult for observers to see (Pollock and Kendall 1987). Unless heterogeneity is measured and modeled to correct for these differences, the unmodeled heterogeneity will result in underestimating population size (Borchers et al. 2006, Laake et al. 2008). There are 2 fundamental approaches to modeling heterogeneity, which we refer to as explicit and implicit. The more intuitive explicit method is to record a set of covariates that can be used to explicitly model the differing sighting probabilities of different animals or groups (Samuel et al. 1987). The alternative method requires more sighting occasions (typically >4) and estimates heterogeneity implicitly from the distribution of

[1] *E-mail: ransomj@usgs.gov*

A B

Figure 1. Initial mark photo (A) of a feral horse (*Equus caballus*) group at Little Book Cliffs Wild Horse Range (Mesa County, CO, USA, 2006) and a resight photo (B) of the same gronp on a subsequent flight, showing the unique individual markings that enable identification of groups without artificial marking. Individuals with uniform pelage coloration, such as the lead horse in photo A, were identified by unique combinations of pelage color and leg and face markings, whereas individuals exhibiting pinto coloration have unique pelage patterns. In this group, the uniformly brown horse exhibits a wide white blaze on its face and 2 white socks on its rear legs, and the pinto horses all have distinct patterns on their left side.

sighting frequencies rather than attempting to explicitly explain most of the sighting variation through covariates. This method is based on fitting mixture models in which multiple subpopulations are assumed to exist, each with its own sighting probability (Pledger 2000).

Feral horses present an opportunity to employ a photographic mark–recapture methodology because, unlike most North American terrestrial mammals, they often have sufficient distinctive natural markings to be uniquely identifiable; therefore, artificial marking is unnecessary for use with a mark–recapture estimation technique (Fig. 1). Natural markings for feral horses include unique pelage colors, and face and leg patterns (Gower 2000, Green 2001). A similar identification strategy has been used for whale (*Megaptera novaeangliae*) body pigmentation patterns (Hammond 1986, da Silva et al. 2000, Schweder 2003), manatee (*Trichechus manatus*) scar patterns (Langtimm et al. 2004), and zebra (*Equus grevyi*) stripe patterns (Rubenstein 1986, Williams 1998). Additionally, the composition of identifiable individuals in family gtoups can be used to help identify the group. Another advantage of using natural markings is that the cost of marking (sighting once) is no higher than for recapture (sighting again). Animals become marked the first time they are observed during a survey, which can be on any survey occasion. Therefore, precision continues to improve as larger portions of the population become marked when they are first observed during the course of the survey.

Our objective was to field-test a mark–recapture method for feral horses using natural markings for identification. We also sought to explore the importance of sighting heterogeneity and the ability to correct for it with appropriate statistical models. We were able to fully evaluate the bias, precision, and cost of this method under some of the most difficult sighting conditions (dense vegetation and complex terrain) often encountered in North American feral horse habitats.

STUDY AREA

We conducted 4 tests of the photographic mark–recapture technique with sightability bias correction covariates on 3 known populations of feral horses: Little Book Cliffs Wild Horse Range (WHR), McCullough Peaks Herd Management Area (HMA), and Pryor Mountain WHR. Survey 1 was conducted at the McCullough Peaks HMA, located in Park County, Wyoming, USA, 32 km east of the town of Cody. The area consisted of 44,440 ha of predominantly flat, open sagebrush park, with rugged badlands along the western edge of the range. Vegetation was characterized by small shrubs, grasses, and forbs with sparse stands of cottonwood (*Populus* sp.) and willow (*Salix* spp.) along the ephemeral streambeds. Elevations ranged from 1,200 m in the lowlands to 1,964 m at the summit of McCullough Peaks. Horses were distributed across the entire area, although the more rugged western portion of the area was known to have lower densities of horses than the eastern portion.

Surveys 2 and 3 were conducted at the Little Book Cliffs WHR, located in Mesa County, Colorado, USA, 13 km northeast of Grand Junction. This area consisted of 14,614 ha of rugged terrain, with sloping plateaus, sagebrush parks, and 4 major canyon systems. Dense stands of Colorado piñon (*Pinus edulis*) and Utah juniper (*Juniperus osteosperma*) occurred across the area and were intermixed with small sagebrush parks. Elevations ranged from 1,500 m to 2,250 m, with some impassable steep-walled shale and sandstone cliffs. Horses were distributed across the entire area.

Survey 4 was conducted at the Pryor Mountain WHR, located 21 km north of Lovell, Wyoming, within northern Bighorn County, Wyoming, and southeastern Carbon County, Montana (USA). The study area ranged in elevation from 1,190 m to 2,625 m and consisted of 16,046 ha of lowland desert, foothill slopes, forested

montane slopes, steep canyons, and isolated grassy plateaus. Vegetation at lower elevations included small desert shrubs and grasses, with mid-elevations dominated by curl-leaf mountain mahogany (*Cercocarpus ledifolius*), Utah juniper, and various grasses, and high elevations dominated by limber pine (*Pinus flexilus*), subalpine fir (*Abies lasiocarpa*), and alpine bluegrass (*Poa alpinum*). Horses were located mostly between mid- and high elevations in moderate to sparse tree cover.

METHODS

Aerial Surveys

We conducted the first 2 surveys with 2 flights each and the final 2 surveys with 6 flights each; we conducted each survey over a 2-day period. We changed the number of flights and amount of time allocated to the latter 2 surveys to better address heterogeneity in group sighting variables and provide greater accuracy and precision to the estimate, which proved necessary for the populations living in relatively dense tree cover and complex topography. We spaced transects closer together and flew more slowly in 2-flight surveys to obtain greater sighting probability per occasion. We designed the time allocated for each survey to be comparable to the cost of one intensive raw-count census flight based on the flight times used by local managers prior to this study. In other words, we held the survey cost constant and examined the difference in precision and accuracy among various methods.

We conducted all surveys with helicopters to allow for mobility in navigating terrain and stability for photographing groups. For all analyses, we defined a group as ≥ 1 horses in association with each other. We conducted both Little Book Cliffs WHR surveys in a B4763B1 (Soloy Aviation Solutions, Olympia, WA) helicopter, and the McCullough Peaks HMA and Pryor Mountain WHR surveys in a 206B-III Jet Ranger (Bell Helicopter, Hurst, TX) helicopter. Survey altitudes ranged from 60 m to 150 m above ground level. We photographed every group using high-resolution digital cameras (Minolta Dimage GT™ [Ramsey, NJ] with 50–200-mm lens or Canon EOS 350D™ [Lake Success, NY] with 70–300-mm image-stabilized lens), and recorded additional covariate data for each group at the time of sighting. Covariates collected on all surveys included locomotion of group (stationary or moving), time of day, sun direction, topography type (rugged or level), vegetation type (trees, shrubs, or grasses), percent vegetation cover (to the nearest 10%), and group size, as well as Global Positioning System (GPS) coordinates. The pilot served as a secondary observer and was accompanied by one primary observer on all surveys. An additional crewmember for data recording was present on 2 flights, but we later deemed this unnecessary. During survey 1, an additional observer took part in the survey to compare film with digital photographs. We found film photography to be unnecessary for adequate resolution and more costly and cumbersome than digital photography.

Immediately following each entire survey, we matched photographs of groups to determine mark–recapture data with a colleetive effort between multiple members of our team and, in some cases, with participation of the herd manager. Horses at Little Book Cliffs WHR and McCullough Peaks HMA exhibited predominately black, bay, and pinto pelage, whereas Pryor Mountain WHR did not contain any pinto horses and was dominated by dun- and grullo-colored animals. Nevertheless, the diversity of phenotypic expression among individual horses (face and leg markings, in addition to pelage color) made this a straightforward comparison exercise and consensus was reached relatively quickly among participants analyzing each photo.

These 3 herds are part of a larger research project for which each individual horse in each population had previously been identified, catalogued, and monitored weekly on the ground (Ransom 2009); thus, the true population size, as well as individual band composition, was known prior to each aerial survey (however, this information was not shared with the aerial observers or the statistical analyst prior to producing estimates of the aerial survey). We tested our ability to identify groups in the photographs by comparing our decisions with these independently generated population composition data. We confirmed pelage color and markings for each horse, and in all cases our aerial photo decisions matched the individuals and groups of known animals on the lists, making it highly unlikely that we misidentified a photo. A small number of horses were not observed on the ground within a short time prior to each aerial survey and were, thus, considered to be of unknown status (possibly deceased or emigrated). Therefore, the known population sizes for each herd used for validation in this study are presented as a range from the minimum number known present to the maximum number representing all animals that could have been present.

Transects for all surveys were linear, predetermined, and followed by GPS route navigation while in the air. We chose spacing of transects by relying on the prior judgment of field personnel for the distance that would provide at least a moderate (30–50%) probability of sighting each horse group anywhere within the study area during the course of the entire survey. The density and extent of tree cover, together with the terrain type, which determined the ease or difficulty of detecting groups, were the primary factors influencing these judgments.

For survey 1, we stratified transect spacing such that a portion of the range known to have a greater density of horses was more intensely surveyed than areas with a lower density. We spaced the high-density transects 1.6 km apart and spaced the low-density transects 3.2 km apart. Sighting conditions across both areas were homogeneous. We performed the second flight 2 days later, with transects oriented the same direction but offset by 0.8 km in the high-density area and by 1.6 km in the low-density area. Mean transect spacing per flight based on actual GPS tracks was 2.6 km and mean flight speed on transects was 43 km/hour.

We spaced survey 2 transects 1.6 km apart uniformly across the range, and then repeated them the following day

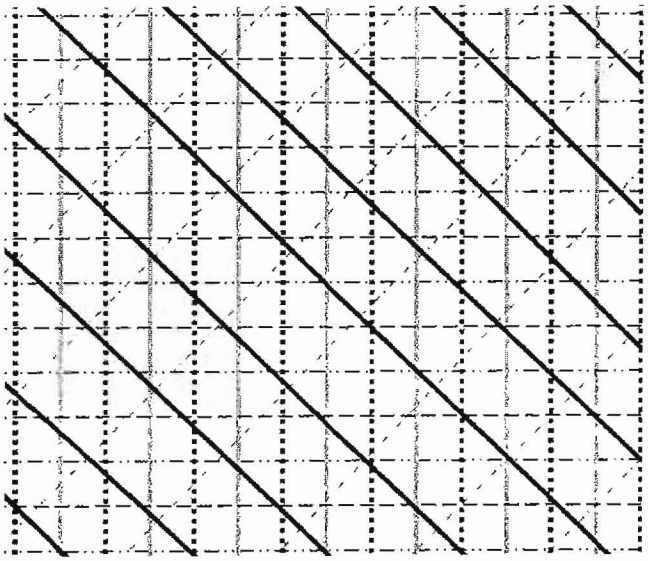


Figure 2. Schematic layout of 6 independent sets of transects used for aerial surveys 3 and 4 at Little Book Cliffs Wild Horse Range, 2006 (Mesa County, CO, USA) and Pryor Mountain WHR, 2007 (SE Carbon County, MT, USA), depicting the flight direction and spacing relative to other sets of transects for each of the 6 flights per survey.

with the same spacing, but perpendicular to the transects of the previous day. On both days, we flew additional transects to follow elevation contours within canyons and other areas where visibility was poor from the original parallel transects. Based on GPS tracks, the average transect spacing we achieved was 0.5 km. Average survey speed was 42 km/hour.

We conducted surveys 3 and 4 over 2 days each, with 6 total flights for each survey (3/day). We restructured these final 2 surveys to better address problems of the sighting probability heterogeneity encountered in survey 2. We collected data in these latter surveys on 6 independent sighting occasions, which provided the required data to apply the implicit method of heterogeneity correction. We spaced transects farther apart and flew at somewhat higher speeds to complete the 6 separate passes in roughly the same time (and therefore at the same cost) as a 2-flight survey required. Flight 1 of each survey was a parallel set of transects, with flight 2 transects rotated by 90°, and flight 3 transects rotated by 45°. The following day, we conducted the 3 flights again, with transects offset between those from the first day (Fig. 2). We spaced survey 3 transects 1.6 km apart and spaced survey 4 transects 3.2 km apart. Average transect spacing was 1.1 km in survey 3 and 1.3 km in survey 4. Average flight speed was 49 km/hour in survey 3 and 75 km/hour in survey 4.

Data Analysis

We used Program MARK (version 5.1, http://welcome.warnercnr.colostate.edu/-gwhite/mark/mark.htm, accessed 20 Mar 2008) to perform analyses with likelihood-based closed-capture population models, as described by Pledger (2000) and Huggins (1989, 1991). We fit a set of models for each analysis that included effects of time (t), behavioral response (b), heterogeneity (h), and a covariate for group size (N). We fit interactive and additive versions of models where they were supported by the data for that analysis. The models we considered (in the notation of Pollock et al. [1990] and Pledger [2000]) were M(0), M(t), M(b), M(h), M(th), M(t+h), M(t+b), M(bh), and M(b+h), each of which also considered the group-size covariate, N, as an additive effect. We excluded additional models with covariates for animal movement, sun position, tree cover, and terrain type after preliminary analyses proved them to have minimal predictive value as assessed using corrected Akaike's Information Criterion (AIC_c) weights (Burnham and Anderson 2002). We created 2-sighting occasion data from surveys with 6 actual sighting occasions by combining data from the first 3 occasions to create a single synthetic first occasion and from the last 3 occasions to create a single synthetic second occasion.

The independent observable unit was a horse group, not an individual horse. Therefore, mark–recapture analysis estimated the number of groups. We also estimated a mean group size from the data and the population estimate was the product of these 2 estimates. We computed the estimated number of groups for each analysis by averaging the individual estimates from each model actoss all of the models considered, weighting each by its AIC_c model weight (Burnham and Anderson 2002). We computed average group size by weighting the observed group size by the inverse of the probability that groups of that size would be observed on ≥ 1 flight during the survey. This procedure adjusted for the bias that would otherwise result from the average of observed groups being larger than the true average of group sizes in the population due to lower sighting probability for smaller groups.

We based standard errors for the number of groups on AIC_c model-weighted estimates, which account for model-selection uncertainty (Burnham and Anderson 2002). We corrected average group size for sighting bias. We computed the error for population using the Delta method (Seber 1973) and computed confidence intervals based on a log-normal distribution for the estimated number of horses not observed as follows:

$$c = \exp\left\{1.96 \times \sqrt{\ln\left[1 + \left(\frac{\hat{s}}{\hat{N} - N_{obs}}\right)^2\right]}\right\}$$

$$95\%\text{CI} = \left[N_{obs} + \frac{\hat{N} - N_{obs}}{c}, N_{obs} + c\left(\hat{N} - N_{obs}\right)\right]$$

Where $\hat{N}$ was the population estimate, $\hat{N}_{obs}$ was the number observed, $\hat{s}$ was the standard error of the population estimate, and the constant 1.96 was the value at which the cumulative standard normal distribution had 97.5% [1 − (0.05/2)] probability. This method was simply assuming a 95% confidence interval on the log-transformed estimate of missed horses and back-transforming that to the linear scale. This was a logical and plausible distribution of errors on a quantity with a range of [0, ∞] and excluded lower

Table 1. Strength of evidence for effects on sighting probability of time, group size, behavioral response to prior sighting, and unmeasured heterogeneity among groups based on the sum of corrected Akaike's Information Criterion model weights across all models that include indicated effect (models may have multiple effects included, so evidence for the several effects will not sum to 100%, but evidence for any single effect will always be 0–100%) for aerial surveys of feral horses at McCullough Peaks Herd Management Area (HMA; Park County, WY, USA), Little Book Cliffs Wild Horse Range (WHR; Mesa County, CO, USA), and Pryor Mountain WHR (SE Carbon County, MT, USA), 2003–2007.

	Survey 1 (McCullough Peaks HMA)	Survey 2 (Little Book Cliffs WHR)	Survey 3 (Little Book Cliffs WHR)		Survey 4 (Pryor Mountain WHR)	
Effects	2 Occ[a] (%)	2 Occ[a] (%)	2 Occ[a] (%)	6 Occ[a] (%)	2 Occ[a] (%)	6 Occ[a] (%)
Time	21.0	36.4	99.0	43.3	23.7	91.8
Group size	44.8	49.7	34.7	51.0	25.5	63.9
Behavior				53.8		14.2
Heterogeneity				46.2		69.4

[a] Occasions (Occ) are separate flights within each survey.

confidence intervals that were less than the minimum known alive (i.e., actual no. observed).

We computed true error rates using the known population sizes for each population at the time of the survey. Because of some ambiguity about the possible death or emigration of a few horses, we knew the true numbers only within a narrow range. Our comparisons of estimates to truth were based on the mean of the minimum and maximum number of horses that could have been present during the survey.

RESULTS

The strength of evidence (AIC_c model wt) supporting the various effects of covariates on sighting probability differed considerably among surveys (Table 1). Support for variation over occasion (time) ranged from as low as 21% in survey 1 to a high of 99% for survey 2. Group-size effect on sighting probability had moderate support on all occasions, ranging from AIC_c weight of 25.5% to 63.9%. Evidence for behavioral response of horses to a prior close approach by the helicopter prior to sighting was moderate in survey 3 (33.5% of groups were running), but nearly absent (only 5.4% of groups running) in survey 4 (the 2 surveys with a 6-flight design for which this effect could be examined). Finally, heterogeneity of sighting probability among different groups had moderate support in both surveys in which it was estimated by the implicit method (surveys 3 and 4).

The effect of group size on sighting probability differed substantially among surveys (Fig. 3). Small groups were most easily seen at McCullough Peaks HMA (survey 1), where trees were absent and visibility was excellent. In stark contrast, sites with tree cover (surveys 2–4) had sighting probabilities for the smallest groups (i.e., 1 horse) well under 50% (Fig. 3). Sighting probability increased with group size at different rates for different flights, but was >90% for groups ≥27 horses in all flights.

Behavioral response to prior observation differed between the 2 surveys for which this effect could be measured (the 6-flight surveys), not only in magnitude, but also in direction. In survey 3, groups had a 20.8% (±7.9% SE) probability of being sighted the first time and 59.2% (±4.5% SE) probability of being sighted again if they had already been sighted. The effect was smaller and in the opposite direction in the Pryor Mountain WHR survey. Initial sighting probability was 40.4% (±7.2% SE) and recapture probability was 29.2% (±3.9% SE). All of these probabilities were based on the model that holds all other factors constant and includes only a behavioral effect. Notice that the trap-shy response (reduced sighting probability following first sighting) is not strongly supported for the Pryor Mountains (Table 1) and that the confidence intervals overlap, but the trap-happy (increased sighting probability following first sighting) effect is moderately supported for Little Book Cliffs WHR with nonoverlapping confidence intervals.

Model-weighted averaging across all models produced estimates of the total number of groups (Table 2). We calculated mean group sizes with bias-corrected weights (Table 2) and multiplied these by the estimated number of groups to obtain the estimated size of the herd (Fig. 4). We compared these estimates to the minimum and maximum possible true number of horses that we knew to be present at the time of the survey (Fig. 4).

Raw (uncorrected) counts of horses actually seen during aerial surveys were consistently less than the known population by between −7.5% to −32.0% ($\bar{x}$ = −18.2%, ±5.1% SE), whereas estimated population size without implicitly accounting for heterogeneity was closer to the true

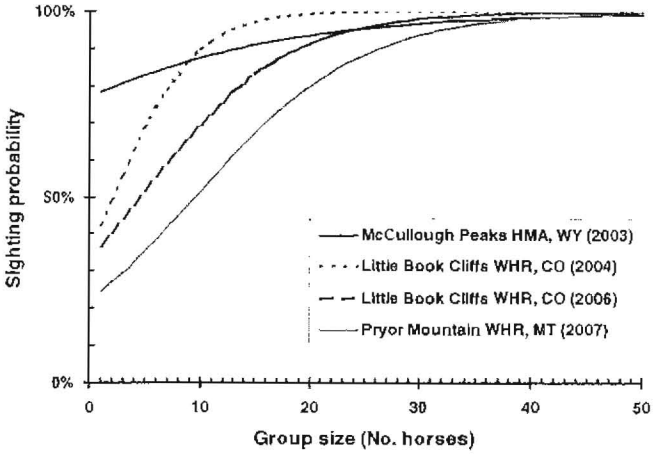


Figure 3. Estimated effect of group size on sighting probability of feral horses from aerial surveys at McCullough Peaks Herd Management Area (HMA; Park County, WY, USA), Little Book Cliffs Wild Horse Range (WHR; Mesa County, CO, USA) and Pryor Mountain WHR (SE Carbon County, MT, USA), 2003–2007. Sighting probabilities are for a single helicopter flight. Surveys are not comparable because transect spacing differed among surveys. Group-size effect is based on a model with all other sighting factors held constant [model M(0 + N)].

Table 2. Data and estimates from aerial surveys of feral horse populations at McCullough Peaks Herd Management Area (HMA; Park County, WY, USA), Little Book Cliffs Wild Horse Range (WHR; Mesa County, CO, USA), and Pryor Mountain WHR (SE Carbon County, MT, USA), 2003–2007. Estimates are based on corrected Akaike's Information Criterion–weighted model average of maximum-likelihood closed-capture models in Program MARK. Models with time, group size, and behavioral response effects were included for all surveys, and models of heterogeneity in capture probabilities were included for surveys done with 6 sighting occasions (as indicated).

	Survey 1 (McCullough Peaks HMA)	Survey 2 (Little Book Cliffs WHR)	Survey 3 (Little Book Cliffs WHR)		Survey 4 (Pryor Mountain WHR)	
Population parameter	2 Occ[a]	2 Occ[a]	2 Occ[a,b]	6 Occ[a]	2 Occ[a,b]	6 Occ
Groups (no.)						
Seen	48	42	38	38	36	36
Estimate	49.3	52.0	39.5	46.6	39.4	41.4
SE[c]	0.95	2.53	0.34	3.64	0.84	1.41
Group size (horses/group)						
Bias corrected $\bar{x}$[d]	7.1	3.0	3.6	3.6	3.4	3.4
SE[c,d]	0.16	0.26	0.27	0.27	0.34	0.34

[a] Occasions (Occ) are separate flights within each survey.
[b] Calculated using data collected on 6 occasions by combining the first 3 occasions and the last 3.
[c] Finite population correction factor applied.
[d] Group size corrected for sized-biased sighting probability.

size, but remained negatively biased ($\bar{x}$ = −13.8%, ±3.3% SE; Table 3). Error was smallest at McCullough Peaks HMA (survey 1), where sightability was excellent due to very open terrain and low vegetation. The standard deviation of errors in our surveys provided a means of estimating the range of potential errors in future surveys. For raw counts, the 90% prediction interval was (−34.9%, −1.4%), indicating that about 5% of raw count (so called census) surveys would result in undercounts of >35% and most (>90%) would result in undercounts of some magnitude.

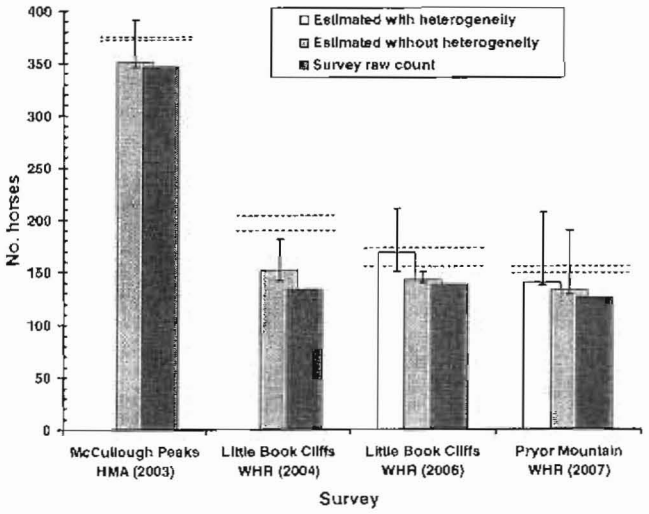


Figure 4. Population estimates (bars), 95% lognormal confidence intervals (black error bars), and minimum and maximum possible known true population size at time of survey (pairs of dashed lines) for aerial surveys of feral horse populations at McCullough Peaks Herd Management Area (HMA; Park County, WY, USA), Little Book Cliffs Wild Horse Range (WHR; Mesa County, CO, USA), and Pryor Mountain WHR (SE Carbon County, MT, USA), 2003–2007. Estimates are based on corrected Akaike's Information Criterion–weighted model average of maximum-likelihood closed-capture models in Program MARK. Models with time and group-size effects were included for all surveys, and models of behavioral response and heterogeneity in capture probabilities were either included (white bars) or not (light gray bars) when the number of occasions supported estimation of these models. Uncorrected raw counts (dark gray bars) from each survey are shown for comparison.

Estimates dramatically improved when sets of models included some that modeled heterogeneity implicitly. Errors on the 2 surveys with implicit heterogeneity information were positive once and negative once and averaged only 5.6% (±3.0% SE) in absolute magnitude (Table 3). The 90% prediction interval for the 6-occasion surveys using the implicit heterogeneity correction was (−16.0%, 10.0%). Fitting a simple exponential model to the errors as a function of the number of sampling occasions per survey provided a projection of the effort (no. of sampling occasions/flights) required for any desired level of precision (Fig. 5).

Statistically estimated precision was uniformly high (<10.5%); however, some of these estimated rates were overly optimistic due to negative bias in methods without implicit heterogeneity correction, which failed to produce confidence intervals that covered the true population size at Little Book Cliffs WHR during either survey 2 or 3. In contrast, the 95% confidence intervals for the estimates easily covered the true population sizes in both cases when implicit heterogeneity was modeled (i.e., in both 6-occasion surveys) and also in surveys 1 and 4, even without implicit heterogeneity correction (Fig. 4).

Survey cost was dominated by the cost of helicopter time. For comparison, we computed the cost per hectare based on actual survey flight time (excluding ferrying to and from the survey site), average helicopter cost of US$650/hour, and actual area covered by the transects. Cost was lowest (US$0.07/ha) at McCullough Peaks HMA (survey 1), where visibility was excellent and enabled wide transect spacing. At the 2 sites dominated by tree cover, cost was US$0.36/ha, which is 5 times the cost at McCullough Peaks HMA. The mean cost (US$0.36/ha) of the 2 surveys with 6 sampling occasions (survey 3 and 4) was nearly identical to the cost (US$0.37/ha) of the comparable 2-occasion survey in rugged, treed terrain (survey 2). This was the result of our

Table 3. Errors in estimates compared to known population size (using mean of min. and max. known animals present during each survey) from aerial surveys of feral horses at McCullough Peaks Herd Management Area (HMA; Park County, WY, USA), Little Book Cliffs Wild Horse Range (WHR; Mesa County, CO, USA), and Pryor Mountain WHR (SE Carbon County, MT, USA), 2003–2007. Errors are the differences between the known population size and the 3 types of estimates: raw counts, statistical estimates without heterogeneity correct, statistical estimates with heterogeneity correct (negative value indicates estimate was lower than actual). Error as a percent of the actual population size is shown. Mean errors and the mean of the absolute values (magnitude) of the errors are presented in the last columns on the right. Statistically estimated (not actual) errors (SE and CV) are also shown for comparison.

Error	Survey 1 (McCullough Peaks HMA)	Survey 2 (Little Book Cliffs WHR)	Survey 3 (Little Book Cliffs WHR)		Survey 4 (Pryor Mountain WHR)		$\bar{x}$	$\bar{x}$ absolute error
	2 Occ[a]	2 Occ[a]	2 Oec[a,b]	6 Oec[a]	2 Occ[a,b]	6 Occ[a]		
Raw-count errors								
No. horses	−28	−63	−26		−27			
Error (%)	−7.5	−32.0	−15.6		−17.6		−18.2	18.2
Error without heterogeneity correction								
Actual errors								
No. horses	−23	−45	−21		−20			
Error (%)	−6.7	−22.7	−13.0		−13.0		−13.8	13.8
Statistically estimated errors								
SE	7.9	9.4	2.4		13.6			
CV	2.2	6.1	1.7		10.2			5.1
Error with heterogeneity correction								
Actual errors								
No. horses				4.2		−13.2		
Error (%)				2.6		−8.6	−3.0	5.6
Statistically estimated errors								
SE				14.2		14.7		
CV (%)				8.4		10.5		9.5

[a] Occasions are separate flights within each survey.
[b] Calculated using data collected on 6 occasions by combining the first 3 occasions and the last 3.

attempt to design surveys of equal approximate cost, leaving differences only in performance to be measured and compared.

During surveys 3 and 4, the patterns of detecting horse groups on each of the 6 successive flights during each of these 2 separate surveys were similar (Fig. 6). Fewer than 50% of horse groups were seen on all but one flight at Little Book Cliffs (Fig 6A). We sighted new, unique groups on every flight except the last one at Pryor Mountain WHR (Fig. 6B). The cumulative total number of unique groups seen over the course of the 6 flights did not exceed 85% and was not approaching a clear asymptote.

DISCUSSION

The photographic mark–recapture with implicit heterogeneity correction worked well in the feral horse populations we tested, despite the difficult sighting conditions created by topography and vegetation cover. Estimated errors were <10.5% coefficient of variation, confidence intervals covered true population size well, and actual errors were substantially lower (5.6% mean absolute error). However, we observed numerous factors that affect sighting probability, which must be accounted for to obtain accurate estimates with realistic confidence intervals. Specifically, raw counts missed as many as one-third of the horses present, despite our observer's high level of aerial survey expertise. Using prediction interval estimates to project error rates on similar surveys led us to predict that undercounts of >35% of the true population would occur in 5% of surveys. This was consistent with the previously observed undercounts of as

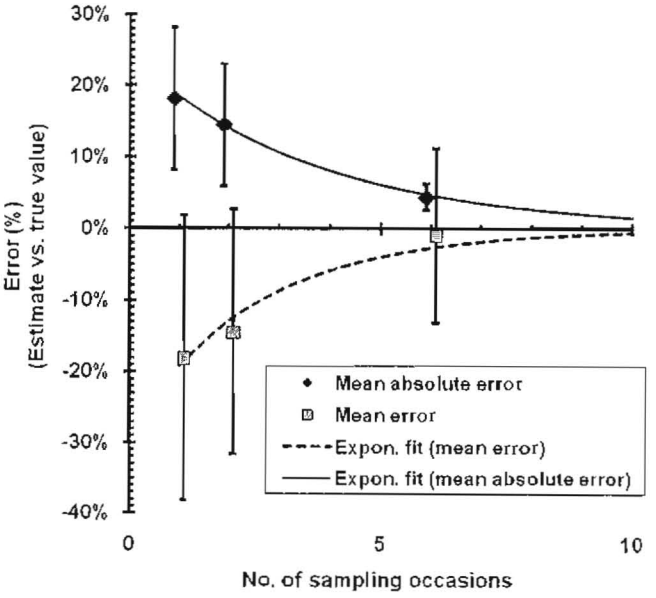
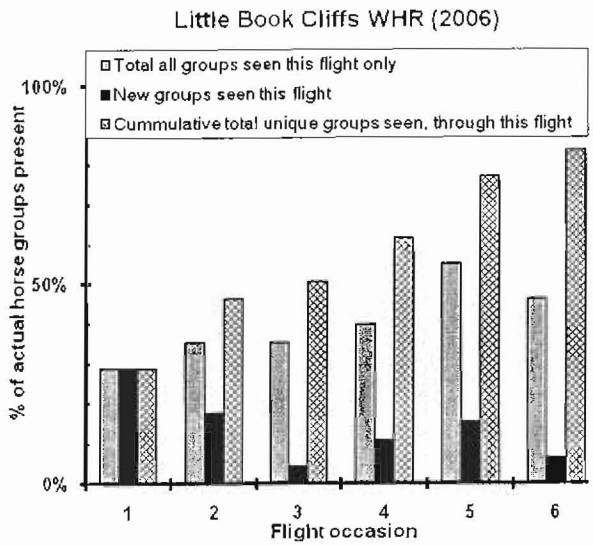


Figure 5. Errors in estimates compared to known true population size (using mean of min. and max. known animals present during survey) from aerial surveys of feral horse populations at McCullough Peaks Herd Management Area (HMA; Park County, WY, USA), Little Book Cliffs Wild Horse Range (WHR; Mesa County, CO, USA), and Pryor Mountain WHR (SE Carbon County, MT, USA), 2003–2007. Mean absolute error is the mean of the absolute values of the errors for each survey. Values for one occasion are the raw counts without correction. Values for 2 occasions are estimates corrected for time, group size, and behavioral response effects on sighting probability, but without modeling unmeasured heterogeneity. Values for 6 occasions are based on a set of models that include some for unmeasured heterogeneity. Error bars for mean absolute error are 95% log-normal confidence interval for estimation of the mean absolute error. Error bars for the mean error (bias) represent the 90% range of predicted outcomes based on a normal distribution (not CI on the mean error).

A



B

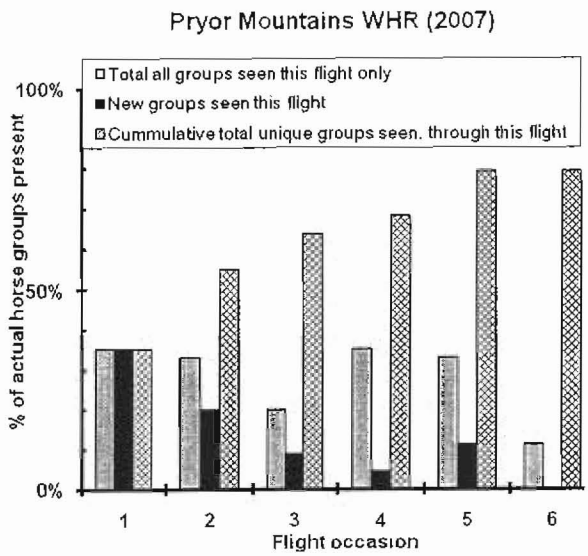


Figure 6. Percent of known feral horse-groups observed in each of 6 independent sighting occasions and cumulative across all sighting occasions, during aerial surveys of Little Book Cliffs Wild Horse Range (WHR), Mesa County, Colorado, USA, 2006 (A), and Pryor Mountain WHR, southeastern Carbon County, Montana, USA, 2007 (B).

much as 39% for feral horses on parts of Assateague Island (east coast of United States) by aerial observation compared to more accurate ground counts (Bashore et al. 1990). Without statistical models and corrections, we do not believe an observer could reliably count all or even most horses within a population unless sighting conditions were extremely good and homogeneous. More importantly, without statistical models and proper sampling methods, there would be no confidence intervals or other evidence to support claims of accuracy.

The corrections required to obtain accurate estimates varied considerably among herds in differing terrain and vegetation types, among flights for the same herd and location, in relation to horse-group size, and due to behavioral response of the horses themselves. No constant correction factor or single sightability-model calibration could account for this extensive spatial and temporal variation in sighting probability. Similar variation in sighting probability with local conditions has previously been observed in aerial surveys (Bayliss and Yeomans 1989, Graham and Bell 1989, Bashore et al. 1990, Walter and Hone 2003, Lubow and Ransom 2007).

Linklater and Cameron (2002) also reported that avoidance behavior of horses to helicopters during aerial surveys led to considerable double-counting of groups, or missing of groups entirely. Using natural markings to identify individuals allowed us to avoid the problems of double-counting and groups merging due to helicopter avoidance response, and we were able to validate this with our extensive record of known individuals in each population. We cannot assess whether helicopter avoidance response was responsible for our not observing specific individuals that were known to be present, but our overall results suggest that whatever heterogeneity in sighting probability this may have caused, our methods and models were able to adequately correct for it.

We found that heterogeneity among groups of horses (some groups are easier to see than others) was among the most important factors affecting sighting probability, particularly under difficult sighting conditions (areas with low sighting probabilities). Despite attempting to correct for differences in sighting probability through the use of covariates for tree cover, terrain, and occasion (time), unmeasured heterogeneity remained a dominant effect in some surveys. This finding casts doubt on the feasibility of applying simple sightability bias correction models (Samuel et al. 1987) that remain constant over time, location, and observers, and that require the measured covariates to explain most or all of the variation in sighting probability, to produce unbiased estimates. The greater the residual, unmodeled, component of heterogeneity, the more severe the negative bias becomes (Laake 1999). Surveys in very open habitat, such as the McCullough Peaks in survey 1, might be sufficiently accurate with only explicit corrections based on a few measured covariates affecting heterogeneity; but in difficult sighting conditions, methods that can implicitly account for heterogeneity are essential. We believe that 6 sighting occasions were needed for sufficient accuracy (precision and bias correction; Fig. 5). Future applications of this method might benefit from experimenting with >6 occasions, although we doubt that >8 would be of any additional value.

The photographic mark–recapture method does not necessarily increase survey time, cost, risk to observers, or disturbance of horses. It is important to note that the poor results we reported here for raw-count (census) surveys were based on the same data from the flights used to produce much more reliable results by applying appropriate statistical models; in other words, the time, cost, effort, risk, and disturbance were identical, and only the methods of data collection and analysis were different.

Obtaining reliable results with the photographic mark–recapture method requires that surveys be properly designed with 6–8 independent sighting occasions, minimal error in identifying groups in photographs, and adequate sighting probabilities for all groups. It is not important that most horse groups are sighted on each occasion, only that each group has a moderate probability (>30%) of being sighted at least once during the survey. Consequently, transects can be designed to merely sample a fraction of the population on each flight. They need not attempt to see all, or nearly all, of the horses on each pass or even once during the course of the survey.

It is common among experienced field biologists to assume that nearly all animals are observed during the course of an aerial survey. Often this belief is held even about the results of a single-pass survey. As our data clearly demonstrate, horses were missed even after 6 independent sighting occasions over the study area, resulting in 16–20% of the known groups never being observed on any occasion. We observed new groups on all sighting occasions except for one. Even by extrapolating beyond the data, there was no clear asymptote to the cumulative raw count that could reliably predict the total population size.

Despite the excellent results we obtained in these trials, it is unlikely that extending this method to all feral horse populations would prove successful. Application to larger herds is constrained by the ability to accurately match (recognize) groups from photographs of the same groups on multiple occasions. Greater uniformity of phenotype among individuals would also complicate the identification of groups. Very dense herds or herds with an atypical social structure that resulted in frequent merging or splitting of bands would also interfere with identification of unique groups. Managers designing surveys using this method must carefully evaluate whether the ability to consistently identify groups correctly can be assured.

We caution users of this method that the survey can only measure horses present in the study area at the time of the survey. In our experience, this often does not correspond to the political boundaries established for managing the herd. Horses often cross fences and travel great distances in search of water, forage, or mates. Whenever possible, it would be important to design surveys so that they cover the entire range of all herds that might intermingle or exchange individuals during a single survey, even if this requires multiple aircraft or multiple days of flights and extended survey boundaries to encompass potential habitat areas adjacent to the public land intended for feral horse use. One distinct advantage of this method is that horse movements do not complicate the survey and there is no requirement that every horse group be observed or that multiple sightings of the same group be avoided (as long as they can be identified accurately in the photographs).

Estimated precision of our estimates was always better (smaller estimated error) when heterogeneity was not modeled implicitly than when it was; this is probably a systematic negative bias in the error estimate that compounds the already negatively biased population estimate in the presence of heterogeneity. The actual error in survey 3 was almost 9 times larger than the statistically estimated standard error due to the combined effect of a negatively biased population estimate and a low error estimate. Thus, heterogeneity leads to estimates that appear to have high precision but are, in fact, neither as precise as they seem nor unbiased. Such estimates can be particularly problematic for managers because they could be misled to believe that they have precise and accurate information when, in fact, they may have neither. Larger confidence intervals at least would alert managers to the weakness of the estimate; seemingly precise but wrong estimates may be worse than none at all, leading to unwarranted overconfidence. Consequently, managers seeking accurate estimates would need to measure and model implicitly the heterogeneity in every population where this method is applied, unless and until multiple repeated surveys demonstrate that a particular population, due to uniform and excellent sighting conditions, such as the one at McCullough Peaks in our study, is not sufficiently heterogeneous to require such corrections. However, given the nearly equal cost of 6-occasion surveys to 2-occasion surveys with similar estimated precision, there should be little incentive to abandon the higher occasion alternative.

Estimation using mark–recapture over multiple sighting occasions provides several important advantages over rival survey methods, including 1) correction for multiple and (temporally or spatially) varying sighting bias; 2) correction for sighting heterogeneity among horse groups using the implicit method; 3) increase in effective sample size in small populations by simply adding additional sighting occasions to the survey; and 4) elimination of potential violation of the assumption that animals not be counted multiple times as they move during the survey. Thus, mark–recapture surveys conducted and analyzed properly can overcome many of the shortcomings that threaten accuracy in other methods such as simple uncorrected censuses and sightability bias correction models. Furthermore, the ability to uniquely and reliably identify individuals or groups over multiple sighting occasions using natural markings in feral horses avoids the prohibitive cost and risk (to both field workers and animals) of capturing and handling large free-ranging animals that is traditionally required to implement mark–recapture with unique artificial markings such as radiocollars or numbered neck-bands. Thus, we believe that the method evaluated here offers one of the best options available for estimating feral horse population size, both for accuracy and cost-effectiveness.

MANAGEMENT IMPLICATIONS

Surveys not employing appropriate sampling designs and statistical models to correct for the pervasive sources of bias and variation present during many aerial surveys may lead to inaccurate population estimates and are, thus, vulnerable to criticism from public stakeholders and unsuitable as a basis for sound management decisions. We have demonstrated that, with the proper field and analysis methods, accurate and precise estimates of feral horse populations are possible and no more costly than many traditional methods employing simple raw counts. The photographic mark–

recapture method is particularly useful for smaller herds and has the distinct advantage of working well under difficult sighting conditions or when horses are moving rapidly and double-counting cannot be reliably avoided; however, it is not applicable to all herds. Our ability to verify this method in multiple populations of known size gives us confidence that results produced by this method will withstand the scrutiny of various public stakeholders, politicians, and advocacy groups that frequently take an interest in the scientific basis for feral horse management decisions.

ACKNOWLEDGMENTS

We thank the Grand Junction, Cody, and Billings Field Offices of the United States Bureau of Land Management for cooperation and assistance with special-use aviation needs over the federal lands they manage. T. R. Stanley, J. E. Roelle, A. B. Shepherd, and 2 anonymous referees provided much appreciated expertise and comments on drafts of this manuscript. This research was funded by the United States Geological Survey Wildlife Program in cooperation with the United States Bureau of Land Management and Colorado State University. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the United States Government.

LITERATURE CITED

Ackerman, B. B. 1988. Visibility bias of mule deer aerial census procedures in southwest Idaho. Dissertation, University of Idaho, Moscow, USA.

Bashore, T. L., R. Keiper, J. W. Turner, Jr., and J. F. Kirkpatrick. 1990. The accuracy of fixed-wing aerial surveys of feral horses on a coastal barrier island. Journal of Coastal Research 6:53–56.

Bayliss, P., and K. M. Yeomans. 1989. Correcting bias in aerial survey population estimates of feral livestock in northern Australia using the double-count technique. Journal of Applied Ecology 26:925–933.

Bodie, W. L., E. O. Garton, E. R. Taylor, and M. McCoy. 1995. A sightability model for bighorn sheep in canyon habitats. Journal of Wildlife Management 59:832–840.

Borchers, D. L., J. L. Laake, C. Southwell, and C. G. M. Paxton. 2006. Accommodating unmodeled heterogeneity in double-observer distance sampling surveys. Biometrics 54:1221–1237.

Bowden, D. L., and R. C. Kufeld. 1995. Generalized mark–resight population size estimation applied to Colorado moose. Journal of Wildlife Management 59:840–851.

Burnham, K. P., and D. R. Anderson. 2002. Model selection and multimodel inference: a practical information-theoretic approach. Second edition. Springer-Verlag, New York, New York, USA.

da Silva, C. Q., J. Zeh, D. Madigan, J. Laake, D. Rugh, L. Baraff, W. Koski, and G. Miller. 2000. Capture–recapture estimation of bowhead whale population size using photo-identification data. Journal of Cetacean Research and Management 2:45–61.

Gower, J. 2000. Horse color explained. Trafalgar Square, North Pomfret, Vermont, USA.

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

Green, B. K. 2001. The color of horses: the scientific and authoritative identification of the color of the horse. Montana Press, Missoula, USA.

Hammond, P. S. 1986. Estimating the size of naturally marked whale populations using capture–recapture techniques. International Whaling Commission Special Issue 8:253–282.

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. 1999. Distance sampling with independent observers: reducing bias from heterogeneity by weakening the conditional independence assumption. Pages 137–148 *in* G. W. Garner, S. C. Amstrup, J. L. Laake, B. F. J. Manly, L. L. McDonald, and D. G. Robertson, editors. Marine mammal survey and assessment methods. A.A. Balkema, Rotterdam, Netherlands.

Laake, J., M. J. Dawson, and J. Hone. 2008. Visibility bias in aerial survey: mark–recapture, line-transect or both? Wildlife Research 35:299–309.

Langtimm, C. A., C. A. Beck, H. H. Edwards, K. J. Fick-Child, B. B. Ackerman, S. I. Barton, and W. C. Hartley. 2004. Survival estimates for Florida manatees from the photo-identification of individuals. Marine Mammal Science 20:438–463.

Linklater, W. L., and E. Z. Cameron. 2002. Escape behaviour of feral horses during a helicopter count. Wildlife Research 29:221–224.

Lubow, B. C., and J. I. Ransom. 2007. Aerial population estimates of feral horses (*Equus caballus*) in Adobe Town and Salt Wells Creek Herd Management Areas using an integrated simultaneous double-count and sightability bias correction technique. U.S. Geological Survey Open File Report 2007-1274, Fort Collins, Colorado, USA.

Neal, A. K., G. C. White, R. B. Gill, D. F. Reed, and J. H. Olterman. 1993. Evaluation of mark–resight model assumptions for estimating mountain sheep numbers. Journal of Wildlife Management 57: 436–450.

Pledger, S. 2000. Unified maximum likelihood estimates for closed capture–recapture models using mixtures. Biometrics 56:434–442.

Pollock, K. H., and W. L. Kendall. 1987. Visibility bias in aerial surveys: a review of estimation procedures. Journal of Wildlife Management 51:502–510.

Pollock, K. H., J. D. Nichols, C. Brownie, and J. E. Hines. 1990. Statistical inference for capture–recapture experiments. Wildlife Monographs 107.

Rabe, M. J., S. S. Rosenstock, and J. C. deVos, Jr. 2002. Review of big-game survey methods used by wildlife agencies of the western United States. Wildlife Society Bulletin 30:46–52.

Ransom, J. I. 2009. Understanding the influences of immunocontraception on equids through competing models of behavior. Thesis, Colorado State University, Fort Collins, USA.

Rubenstein, D. I. 1986. Ecology and sociality in horses and zebras. Pages 282–303 *in* D. I. Rubenstein and R. W. Wrangham, editors. Ecological aspects of social evolution. Princeton University Press, Princeton, New Jersey, USA.

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

Schweder, T. 2003. Abundance estimation from multiple photo surveys: confidence distributions and reduced likelihoods for bowhead whales off Alaska. Biometrics Journal 59:974–983.

Seber, G. A. F. 1973. The estimation of animal abundance and related parameters. Griffin, London, United Kingdom.

Singer, F. J., and E. O. Garton. 1994. Elk sightability model for the Super Cub. Pages 47–49 *in* J. W. Unsworth, F. A. Leban, D. J. Leptich, E. O. Garton, and P. Zager, editors. Aerial survey manual. Second edition. Idaho Department of Fish and Game, Boise, USA.

Unsworth, J. W., F. A. Leban, D. J. Leptich, E. O. Garton, and P. Zager. 1994. Aerial survey: user's manual. Second edition. Idaho Department of Fish and Game, Boise, USA.

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

Williams, B. K., J. D. Nichols, and M. J. Conroy. 2002. Analysis and management of animal populations—modeling, estimation, and decision making. Academic Press, San Diego, California, USA.

Williams, S. D. 1998. Grevy's zebra: ecology in a heterogeneous environment. Dissertation, University College, London, United Kingdom.

Associate Editor: Forsyth.