Population Ecology

Detection Probability in Aerial Surveys of Feral Horses

JASON I. RANSOM,[1] *U.S. Geological Survey, Fort Collins Science Center, 2150 Centre Avenue, Building C, Fort Collins, CO 80526, USA*

ABSTRACT Observation bias pervades data collected during aerial surveys of large animals, and although some sources can be mitigated with informed planning, others must be addressed using valid sampling techniques that carefully model detection probability. Nonetheless, aerial surveys are frequently employed to count large mammals without applying such methods to account for heterogeneity in visibility of animal groups on the landscape. This often leaves managers and interest groups at odds over decisions that are not adequately informed. I analyzed detection of feral horse (*Equus caballus*) groups by dual independent observers from 24 fixed-wing and 16 helicopter flights using mixed-effect logistic regression models to investigate potential sources of observation bias. I accounted for observer skill, population location, and aircraft type in the model structure and analyzed the effects of group size, sun effect (position related to observer), vegetation type, topography, cloud cover, percent snow cover, and observer fatigue on detection of horse groups. The most important model-averaged effects for both fixed-wing and helicopter surveys included group size (fixed-wing: odds ratio = 0.891, 95% CI = 0.850–0.935; helicopter: odds ratio = 0.640, 95% CI = 0.587–0.698) and sun effect (fixed-wing: odds ratio = 0.632, 95% CI = 0.350–1.141; helicopter: odds ratio = 0.194, 95% CI = 0.080–0.470). Observer fatigue was also an important effect in the best model for helicopter surveys, with detection probability declining after 3 hr of survey time (odds ratio = 0.278, 95% CI = 0.144–0.537). Biases arising from sun effect and observer fatigue can be mitigated by pre-flight survey design. Other sources of bias, such as those arising from group size, topography, and vegetation can only be addressed by employing valid sampling techniques such as double sampling, mark–resight (batch-marked animals), mark–recapture (uniquely marked and identifiable animals), sightability bias correction models, and line transect distance sampling; however, some of these techniques may still only partially correct for negative observation biases.

KEY WORDS abundance, bias, *Equus caballus*, observation error, population estimation, simultaneous double-count, visibility, wild horse.

Aerial survey techniques are often employed to overcome the obstacles of large spatial expanses, limited access to all areas of interest, and dense vegetation that may prevent observers from counting large animals from the ground. These surveys use fixed-wing and rotary-wing aircraft to view animals from above and can be effective for estimating population size across large areas (Caughley et al. 1976, Tracey et al. 2008, Lubow and Ransom 2009), but heterogeneous sighting conditions can still lead to observation bias that may significantly underestimate the true population (Caughley 1974, Borchers et al. 2006, Laake et al. 2008). Observation bias in aerial surveys may arise from internal factors (aircraft type, observer fatigue, observer skill, and observer seat position), external factors (animal behavior, distance from the aircraft, and group size), and environmental factors (angle of the sun, landscape shading from cloud cover, percent snow cover, percent vegetation cover, topography, and vegetation type), but many of these potential influences are often ignored when estimating animal abundance (Fleming and Tracey 2008).

Some of the numerous and variable biases that may arise from such factors can often be addressed, with varying effectiveness, by using aerial survey methods that employ statistical sampling theory to make corrections to the actual number of animals observed. The 5 main categories of sampling techniques are double sampling, mark–resight (batch-marked animals), mark–recapture (uniquely marked and identifiable animals), sightability bias correction models, and distance sampling (Barker 2008). Such estimators have been historically used for small mammals and animals inhabiting small spatial areas, though use for large wildlife species has increased rapidly over recent decades (see Schwarz and Seber 1999, Williams et al. 2002 for review). However, the use of so-called census methods, which assume that all animals are seen, remains a common practice. This can lead to up to one-third of large mammals on the landscape being unaccounted for in aerial surveys (Pollock and Kendall 1987, Samuel et al. 1987, Ackerman 1988, Bodie et al. 1995).

Received: 4 October 2010; Accepted: 15 May 2011;
Published: 5 August 2011

Additional Supporting Information may be found in the online version of this article.

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

Statistical sampling has rarely been employed to estimate feral horse (*Equus caballus*) populations in the United States (Lubow and Ransom 2009), yet this species is often the subject of political and public interest, intense management, and is at the center of numerous ecological debates.

Estimating large animal abundance is an important task for scientists and practitioners, and is especially critical when species of concern are non-native and potentially overabundant. Feral horses, for example, inhabit grasslands, desert, montane, and forest environments on every continent except Antarctica. Populations may increase over 20% annually (Eberhardt et al. 1982, Garrott et al. 1991) and have a wide array of influences on native flora, fauna, and ecosystem processes (Smith 1986, Fahnestock and Detling 1999, Levin et al. 2002, Beever and Herrick 2006, Beever et al. 2008). Regardless of the sampling technique employed, understanding sources of bias and the magnitude of effects influencing aerial detection of wildlife can lead to better project planning and more informed statistical analyses, and thus may help produce better estimates of abundance.

I used data collected from aerial surveys of feral horses that employed dual observers independently collecting data in a mark–resight framework similar to that described by Graham and Bell (1989) for feral horses, which provided trials with the opportunity for each observer to simultaneously observe or miss horse groups on the landscape. I did not evaluate the simultaneous double-count method here because most of these surveys were performed on populations of unknown size and abundance estimates could not be validated. Instead, I used the repeated paired sighting trials to test the influence of several internal, external, and environmental factors on group detection in feral horse surveys. I sampled multiple populations in the western USA and assessed the strength of 45 candidate models incorporating these effects for both fixed-wing and helicopter surveys.

STUDY AREAS

The populations used in this study occupied public lands in the western United States consisting of flat, rolling, and mountainous terrain populated predominately by sagebrush (*Artemisia* spp.). Some tree species occurred in the study areas, primarily piñon (*Pinus edulis*) and juniper (*Juniperus* sp.), along with sparse stands of cottonwood (*Populus deltoides*). Ungulates sympatric with feral horses in the study areas included livestock (*Bos* spp. and *Ovis aries*), elk (*Cervus elaphus*), mule deer (*Odocoileus hemionus*), bighorn sheep (*Ovis canadensis*), and pronghorn (*Antilocapra americana*). Aerial surveys were conducted throughout all seasons from 2004 to 2009.

The 15 areas surveyed were the Adobe Town-Salt Wells Creek Herd Management Area (HMA) Complex, Wyoming (850,115 ha at latitude 41°21′N, longitude 108°30′W), Buck and Bald HMA Complex (Buck-Bald, Butte, Cherry Creek, and Maverick-Medicine HMAs), Nevada (1,386,871 ha at latitude 40°14′N, longitude 115°13′W), Cedar Mountain HMA, Utah (86,625 ha at latitude 40°29′N, longitude 112°57′W), Lander HMA Complex (Conant Creek, Rock Creek Mountain, Muskrat Basin, and Dishpan Butte HMAs), Wyoming (151,880 ha at latitude 42°47′N, longitude 107°51′W), Little Owyhee and Snowstorm Mountains HMAs, Nevada (233,490 ha at latitude 41°41′N, longitude 116°59′W), McCullough Peaks HMA, Wyoming (44,440 ha at latitude 44°35′N, longitude 108°40′W), and Sand Wash HMA, Colorado (63,390 ha at latitude 40°47′N, longitude 108°21′W).

METHODS

Aerial Surveys

Observers performed aerial surveys using a simultaneous double-count method (described by Caughley and Grice [1982] and adapted to feral horses by Graham and Bell [1989]) in which 2 observers independently collected data from the same side of either a fixed-wing aircraft (24 flights) or helicopter (16 flights). Observers conducted all surveys between 0700 hours and 1800 hours with daily flight times ranging 1–8 hr and survey duration ranging 1–6 days. Each survey included 3 experienced observers accompanied by 1 pilot, and observers systematically rotated seat position at each fuel stop. Flight time between fuel stops varied per flight, but each period incorporated multiple transects. Sixteen different observers collected data during the course of this study. Individual observers were therefore not common across all surveys, but each observer participated in multiple flights.

Pilots performed fixed-wing flights using a Cessna 210 aircraft (Cessna, Wichita, Kansas) and totaled 124.72 hr of survey time. Fixed-wing aircraft maintained an above-ground level (AGL) altitude of approximately 152–183 m and airspeed of approximately 259–296 km/hr. Pilots performed helicopter flights using either a 206BIII Jet Ranger (Bell Helicopter, Hurst, Texas) or 206LIII Long Ranger (Bell Helicopter) and totaled 91.96 hr of survey time. Observers did not remove aircraft doors for any surveys and there was no fundamental difference in observer view from each type of helicopter; Long Rangers were only necessary for conditions where additional power was required to safely operate with 4 people on board. Helicopters maintained an AGL of approximately 60–80 m and airspeed of 148–166 km/hr. All transects were pre-determined and spaced 1.6–2.1 km apart in parallel arrangement, but transect direction varied by survey. Observers only recorded horse groups from the transect where the horses were first detected. Transect spacing and survey altitude were based on manager recommendations from previous experience and maintained here for continuity with management practice. All transects, flight paths, and horse group locations were followed and recorded using a handheld Global Positioning System (GPS) unit with an external antenna mounted in the front window of the aircraft.

Observers maintained audio (radio silence) and visual (seat partition) isolation during the surveys, with the provision that once a group of horses had passed the rear observer, all observers were free to discuss the group size and circle back if confirmation was needed. This procedure did not affect the independent observation record, but ensured that observers recorded the agreed number of animals. Covariate data collected during all surveys consisted of: approximate distance to the group when perpendicular to the aircraft (d; <0.8 km, >0.8 km), behavior of horses when first observed (b; still, moving), cloud cover (l; clear, partly cloudy, overcast), hour of day (h), number of horses per group (a), observer name, observation cardinal direction, percent snow cover ($\hat{w}$), seat position of the observer (p), topography (e; uniform, complex), and vegetation type (g; none or grass, shrub, tree). Observers recorded distance by visually locating topographic or anthropogenic landmarks midway between transects and then estimating group position relative to these points. Although imprecise, the 2 broad categories allowed observers to easily determine if groups were closer to the current transect or the adjacent transect and assign a category. I defined uniform topography as terrain with topographic features less than the size of a horse, and complex terrain as terrain with topographic features greater than or equal to the size of a horse. Likewise, I defined shrubs as woody vegetation less than the size of a horse, and trees as woody vegetation greater than or equal to the size of a horse. Observers assigned snow, topography, and vegetation values for the area immediately surrounding each horse group, including the area containing all members of large groups and within an approximately 10-m radius from the center of small groups. Observers visually estimated amount of snow cover to the nearest 10%.

I calculated covariate data for fatigue (f), cumulative fatigue (f'), and sun effect (s) a posteriori. I defined observer fatigue as the number of consecutive hours an observer had performed an aerial survey when an observation was made on a given day and cumulative fatigue as the number of consecutive days an observer had performed an aerial survey. I calculated the effect of sun position on observations for each survey location and date (t) of observation using the National Oceanic and Atmospheric Administration solar calculator (Department of Commerce 2010). This provided precise values for apparent sunrise (r_t), apparent sunset (u_t), solar declination (c_t), and solar noon (o_t), such that for observations directed south, I calculated sun effect as

$$s_{\mathrm{s}} = \frac{23.44 - c_t}{46.88}$$

thus transforming c_t to a scale of 0–1, with 1 representing the winter solstice when the sun is lowest on the horizon (directly facing an observer) as viewed from the northern hemisphere (Fig. 1). For observations directed east,

$$s_{\mathrm{e}} = \frac{((o_t - h_t)/(o_t - r_t)) + (1 - s_{\mathrm{s}})}{2}$$

Likewise, for observations directed west,

$$s_{\mathrm{w}} = \frac{((h_t - o_t)/(u_t - o_t)) + (1 - s_{\mathrm{s}})}{2}$$

Equations for east- and west-directed observations place equal value on solar declination and distance of the sun from the observation meridian, thus allowing the minimum effect of sun on east–west observations to occur at solar noon on the winter solstice, increasing to the maximum effect at sunrise or sunset on the summer solstice (Fig. 1).

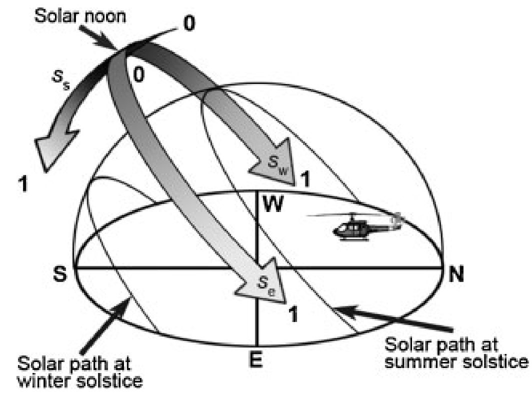

Figure 1. Diagram of sun movement at 40° North latitude depicting the relationship of sun effect variables s_{s}, s_{e}, and s_{w} to a south-facing aircraft (diagram adapted from National Oceanic and Atmospheric Administration [Department of Commerce 2010]).

Statistical Analyses

I performed analyses on all observations when at least 1 observer (e.g., front seat) recorded the presence of a group and the paired observer (e.g., seat directly behind the front observer) had an equal opportunity to do so. Groups detected in front of or directly beneath aircraft by the front seat observer were not available to the rear seat observer and I therefore excluded them from analyses. I omitted observations made from the seat behind the pilot because the pilot was not an observer and therefore a double-count was not possible on that side of the aircraft. Analyses did not statistically account for groups that may have been missed by both observers since covariate values were unknown and I did not attempt to estimate them.

I formulated a candidate set of 45 a priori mixed-effect logistic regression models (see Tables S1 and S2 available online at www.onlinelibrary.wiley.com), using additive and interactive formulations of the variables a, d, e, f, f', g, l, p, s, and $\hat{w}$ as fixed effects, and observer|population as a random effect. Following the notation of Gelman and Hill (2007), candidate models took the form of:

$$Pr(y_i = 1) = \mathrm{logit}^{-1}(\alpha_{0jk[i]} + \beta_0 + \beta_1 X_1[i])$$

for $i = 1, \ldots, n$ observations, where the random effect intercept $\alpha_0 \sim N\ (\mu = 0,\ \sigma^2_{jk})$, for $j = 1, \ldots, n$ observers in k = 1,. . ., n horse populations. β_0 represents the intercept for fixed effect X_1 with coefficient β_1. I analyzed all models using the lme4 package of R version 2.11.1 (http://www.r-project.org, accessed 14 Jul 2010). I obtained estimates by maximum likelihood (Harville 1977), and strength of evidence for models using Akaike's Information Criteria (AIC; Burnham and Anderson 2002). I evaluated model fit by calculating area under the receiver operating curve (AUC) with function somers2 in R package Hmisc. I computed AUC using the fixed effect predictions while holding random effects at zero.

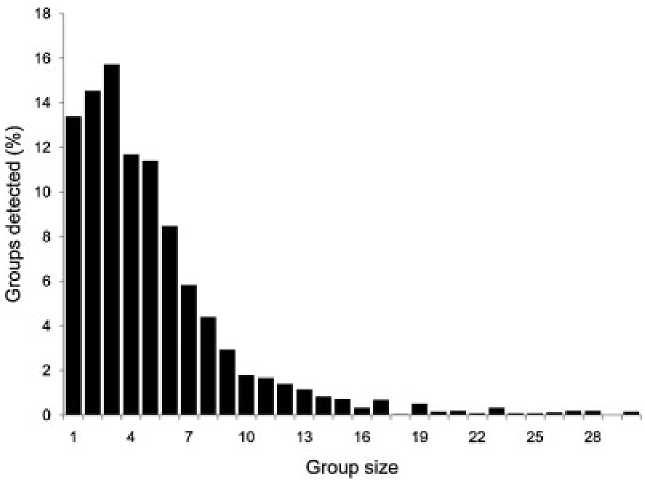

Figure 2. Size and frequency of feral horse (*Equus caballus*) groups detected during 24 fixed-wing flights and 16 helicopter flights over rangelands in the western United States, 2004–2009 ($n = 2{,}517$). Groups >30 animals were rare and are excluded from the figure.

RESULTS

Observers detected 1,792 feral horse groups during 24 fixed-wing flights and 725 groups during 16 helicopter flights. The size of groups encountered during these surveys ranged from 1–86 horses, with mean group size = 5.50 horses (95% CI = 5.26–5.75). Over 90% of the groups encountered numbered <14 animals and groups of 1–3 horses occurred most frequently (Fig. 2). Modeling was stratified between fixed-wing and helicopter surveys due to considerable variation in sighting conditions related to the horses' awareness of aircraft. In helicopter surveys, 85.89% (95% CI = 83.35–88.44) of horse groups were running before being sighted, whereas only 7.89% (95% CI = 6.64–9.14%) of horse groups were running from fixed-wing aircraft before being sighted. Consequently, 22.04% (95% CI = 19.91–28.16%) of detected horses were >0.8 km away from helicopter transects, but only 3.22% (95% CI = 2.41–4.04%) of detected horses were >0.8 km away from fixed-wing transects.

AUC for the fixed-wing global model was 0.613 (predicted probabilities ranging 0.650–0.998), and for the helicopter global model was 0.803 (predicted probabilities ranging 0.346–1.000). Four models were plausible among the fixed-wing candidate model set and each of them included the factors group size and sun effect, with 2 models also inclusive of vegetation and 2 of distance (Table 1). Likewise, the 1 well-supported model in the helicopter candidate set also included group size and sun effect, as well as topography, vegetation type, and observer fatigue (Table 1). Variation in detection probability among observers by geographic location was an important influence in both fixed-wing and helicopter models (Table 2; fixed-wing σ_{OBS} ranged 0.274–0.294, helicopter $\sigma_{\mathrm{OBS}} = 0.156$). For example, in the model-averaged estimate for fixed-wing surveys, the poorest observer missed 32.12% (95% CI = 21.46–45.04%) of 3-horse groups whereas the best observer only missed 9.22% (95% CI = 5.54–14.95%) of 3-horse groups in good conditions (looking away from the sun). In the model-averaged estimate for helicopter surveys, detection probability was very high in good conditions (looking away from the sun, no vegetation, open topography, fatigue <4 hr) with the poorest observer only missing 1.83% (95% CI = 1.36–2.48%) of 3-horse groups and the best observer only missing 0.79% (95% CI = 0.24–2.56%) of 3-horse groups. However, in poor conditions (looking into the sun, shrub cover, complex topography, fatigue >3 hr), the poorest observer missed 70.26% (95% CI = 63.48–76.24) of 3-horse groups and the best observer missed 50.25% (95% CI = 42.64–57.85) of 3-horse groups (Fig. 3).

Group size was an important effect in all supported models (Table 1; 95% CI for odds ratios do not overlap 1 in any model), and models containing this effect accounted for 99.99% of AIC weight in both fixed-wing and helicopter candidate model sets. Observers in fixed-wing aircraft missed fewer groups as group size increased, with the model-averaged fixed-wing estimate of 17.51% (95% CI = 16.83–18.21%) single-horse groups missed in good conditions and 34.60% (95% CI = 33.53–35.70%) single-horse groups missed in poor conditions. However, observers detected

Table 1. Most-supported ($\Delta_i < 4$), and intercept-only, a priori mixed-effect logistic regression models of detection probability from 24 fixed-wing and 16 helicopter flights to survey feral horses (*Equus caballus*) in the western United States, 2004–2009. Each model incorporates random effects of observer|population on the intercept term. Candidate model sets included 45 models for each aircraft type with fixed-wing observations $n = 1{,}792$, helicopter observations $n = 725$, and K number of parameters. Models are ranked by change in Akaike's Information Criteria (AIC; Δ_i) and AIC weight among all candidate models (w_i).

Model no.	Model	*K*	Δ_i	w_i
Fixed-wing				
20	Group size + sun	5	0.00	0.190
22	Group size + sun + vegetation	6	0.05	0.185
16	Group size + sun + distance	6	0.78	0.129
44	Group size + sun + vegetation + distance	7	0.81	0.127
12	Group size + sun + group size × sun	6	1.99	0.070
1	Group size	4	2.30	0.060
45	Group size + sun + vegetation + distance + fatigue	8	2.49	0.055
27	Group size + distance	5	3.08	0.041
40	Group size + percent snow	5	3.65	0.031
14	Intercept only	3	32.51	0.000
Helicopter				
45	Group size + sun + topography + fatigue + vegetation	8	0.00	0.989
14	Intercept only	3	53.82	0.000

Table 2. Odds ratios for fixed effects in the most-supported a priori mixed-effect logistic regression models of detection probability from 24 fixed-wing ($n = 1{,}792$ observations) and 16 helicopter ($n = 725$) flights to survey feral horses (*Equus caballus*) in the western US, 2004–2009. Variance is given as σ_{OBS} for the random effects of observer|population on the intercept term of each model.

					Vegetation	
Model no.	**Model**	σ_{OBS}	**Group size**	**Sun effect**	**None to shrub**	**None to tree**
Fixed-wing						
20	Group size + sun effect	0.283	0.71	0.599		
	95% confidence interval		0.678–0.744	0.371–0.968		
22	Group size + sun effect + vegetation	0.274	0.707	0.585	0.745	0.498
	95% confidence interval		0.521–0.959	0.361–0.948	0.549–1.011	0.082–3.030
16	Group size + sun effect + distance	0.294	0.712	0.599		
	95% confidence interval		0.680–0.746	0.371–0.968		
44	Group size + sun effect + vegetation + distance	0.285	0.708	0.585	0.744	0.492
	95% confidence interval		0.676–0.742	0.361–0.947	0.548–1.010	0.081–2.997
Helicopter						
45	Group size + sun effect + topography + fatigue + vegetation	0.156	0.640	0.195	0.371	0.422
	95% confidence interval		0.587–0.698	0.080–0.472	0.128–1.077	0.102–1.741

nearly all 30-horse groups from fixed-wing aircraft regardless of conditions (Fig. 3A). The same trend occurred in helicopter surveys; however, the contrast in detection probabilities of good versus poor conditions was dramatic. In good conditions, 98.16% (95% CI = 97.99–98.31%) of single-horse groups were detected, but in poor conditions only 29.47% (95% CI = 27.69–31.30%) of single-horse groups were detected and detection still did not reach 100% when group size approached 30 horses (Fig. 3B).

This disparity in group size effect between good and poor conditions for helicopter surveys is explained in part by sun effect (odds ratio = 0.194, 95% CI = 0.080–0.470). In good conditions, detection probability only dropped 5.31% from no sun effect (looking directly away from the sun) to full sun effect (looking directly toward the sun) at a group of 3 horses (the most frequently observed group size; Fig. 4B), but in poor conditions detection probability of the same group size dropped 38.30% from no sun effect to full sun effect (Fig. 4B). Sun effect on observations from fixed-wing aircraft followed the same pattern, though with lesser magnitude (Fig. 4A). Model-averaged detection probability of a group of 3 horses in no vegetation cover declined as the observer looked toward the sun from 85.63% (95% CI = 76.76–91.49%) to 79.02% (95% CI = 67.60–87.17%), with a similar rate in tree cover declining from 81.58% (95% CI = 71.05–88.88%) to 73.67% (95% CI = 60.79–83.47%).

Fatigue was an important factor in helicopter surveys (odds ratio = 0.278, 95% CI = 0.144–0.537), but was not contained in any supported fixed-wing models. In helicopter surveys, model-averaged detection probability for a 3-horse group in good conditions was 98.62% (95% CI = 95.61–99.58%) during the first 3 hr of each flight and 95.22% (95% CI = 85.82–98.50%) during subsequent hours. In poor conditions, this probability declined from 66.91% (95% CI = 51.41–79.45%) during the first 3 hr to 35.99% (95% CI = 22.73–81.80%) during subsequent hours. No models of cumulative fatigue (successive days of aerial

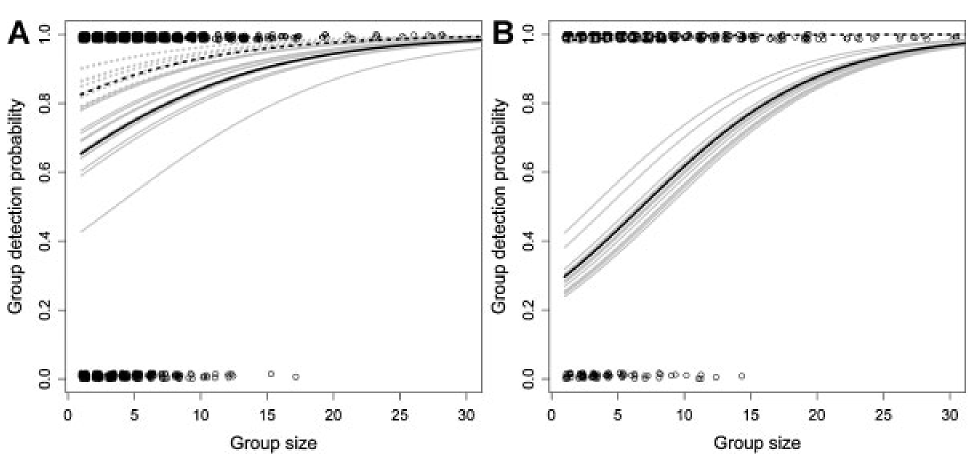

Figure 3. Model-averaged probabilities of group detection as a function of group size in 24 fixed-wing flights (A) and 16 helicopter flights (B) surveying feral horses (*Equus caballus*) in the western United States, 2004–2009. I show model estimates as dotted lines for best conditions (no vegetation, looking away from the sun, distance <0.8 km [fixed-wing] or open topography [helicopter], fatigue <4 hr) and as solid lines for worst conditions (tree cover [fixed-wing] or shrub cover [helicopter], looking toward the sun, distance >0.8 km [fixed-wing] or complex topography [helicopter], fatigue >3 hr). I show the random effects of observer|population in gray. I present observations (○) as 0 for missed groups and 1 for detected groups and offset values by the addition or subtraction of a small random value in order to help distinguish multiple observations at identical values.

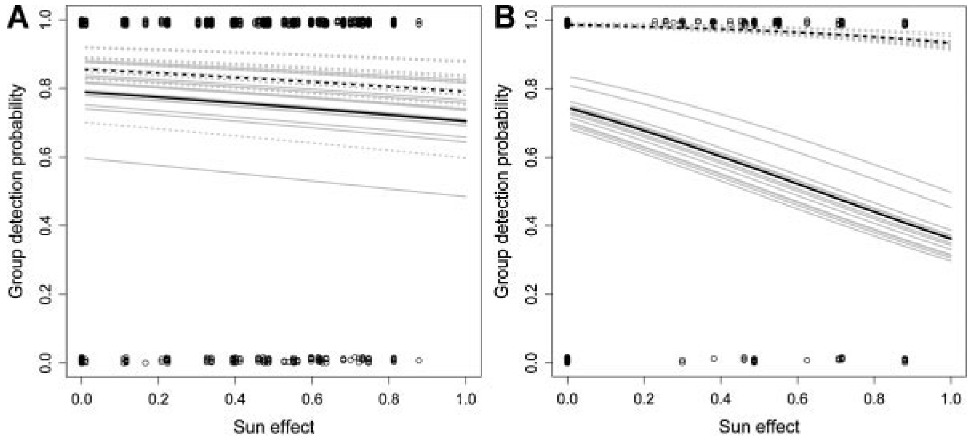

Figure 4. Model-averaged probabilities of detecting a 3-horse group as a function of sun effect, based on 24 fixed-wing (A) and 16 helicopter (B) flights in the western United States, 2004–2009 (fixed-wing: $n = 1,792$ observations, helicopter: $n = 725$ observations). I show model coefficients as dotted lines for best conditions (no vegetation, looking away from the sun, distance <0.8 km [fixed-wing] or open topography [helicopter], fatigue <4 hr) and as solid lines for worst conditions (tree cover [fixed-wing] or shrub cover [helicopter], looking toward the sun, distance >0.8 km [fixed-wing] or complex topography [helicopter], fatigue >3 hr), model estimates in bold, and observer|population (random effects on the model intercept) in gray. Observers are looking increasingly toward the sun as sun effect increases from 0 to 1. I show observations (○) as 0 for missed groups and 1 for detected groups and offset values by the addition or subtraction of a small random value in order to help distinguish multiple observations at identical values.

survey) received support for either fixed-wing or helicopter surveys.

Vegetation type was a factor in supported models for both fixed-wing and helicopter detection probability, but the magnitude of effect was small in all models (Table 2). The model-averaged estimate for vegetation type in fixed-wing surveys estimated detection probability of a group of 3 horses in good conditions declining with increasing vegetation height, though it should be noted that the range of observer performance by vegetation type included the poorest observer in all vegetation types performing worse than the model-averaged estimate for any individual vegetation type. The same small effect occurred in helicopter surveys, but horse groups in shrub and tree cover were detected equally and slightly less often than horses in no vegetation cover. The difference in detection probability for a 3-horse group in good conditions and no vegetation cover as compared to good conditions but in shrub or tree cover was only 2.03%. In poor conditions, this same probability declined 24.45%.

The best model for helicopter surveys included topography, but there was no difference in model-averaged detection probabilities between open and complex terrain (odds ratio = 0.395, 95% CI = 0.120–1.303). I omitted this factor from fixed-wing models because 97.67% of observations from fixed-wing aircraft were made over uniform terrain, leaving the data too homogeneous to be informative. Likewise, I omitted the effect of distance from helicopter surveys because aircraft presence artificially displaced horses. Distance was present in 2 of the most-supported fixed-wing models, but there was no difference in the model-averaged detection probabilities (odds ratio = 0.846, 95% CI = 0.465–1.539).

No models containing observer seat position or the effect of clouds creating shadow mosaics on the landscape were plausible for either fixed-wing or helicopter surveys. Percent snow cover was present in 1 weakly supported fixed-wing model ($\Delta_i = 3.646$), with 83.49% (95% CI = 76.28–88.84%) of 3-horse groups being detected in 0% snow and 91.95% (95% CI = 87.89–94.73%) being detected in 100% snow. The lowest detection probability for a 3-horse group (72.06%, 95% CI = 46.10–88.60%) occurred when snow cover was recorded as 50%, creating a patchwork of snow and dry ground around horse groups.

DISCUSSION

Detection probability of feral horse groups on the landscape was subject to internal, external, and environmental factors in aerial surveys across the western United States. Group size and sun effect were the most important influences on the visibility of horse groups during both fixed-wing and helicopter surveys, but variation among observers produced an observer effect that also contributed to detection probability. In the best fixed-wing survey conditions (looking away from the sun, no vegetation, distance <0.8 km, fatigue <4 hr), there was a 22.90% difference in estimated detection probability of 3-horse groups between the best and worst observer. In the best helicopter survey conditions (looking away from the sun, no vegetation, open topography, fatigue <4 hr), this difference was only 1.02%, but in the worst conditions (looking toward the sun, shrub cover, complex topography, fatigue >3 hr) the difference was 20.01%. Group size and sun effect, as well as observer effect, for fixed-wing surveys estimated a 15.06% difference between detection of 3-horse groups in best versus worst conditions. Likewise, group size, sun effect, vegetation type, topography, and fatigue, as well as observer effect, in helicopter surveys estimated a 31.71% difference between detection of 3-horse groups in best versus worst conditions.

The gregarious nature of some species is known to cause bias in aerial surveys because large groups are easier to detect than small groups, resulting in negatively skewed population estimates (Samuel and Pollock 1981). Feral horses are gregarious and results from both types of aircraft in my study indicate a strong relationship between detection probability and group size. This finding concurs with Graham and Bell (1989) who conducted fixed-wing aerial surveys of Australian feral horses and found a positive linear relationship between detection probability and group size. This trend was also found in helicopter surveys of elk, and similar to my data, single animals were only sighted 22% of the time and groups >15 elk were all detected (Samuel et al. 1987). Anderson et al. (1998) found that detection of elk groups increased with group size regardless of vegetation type or whether animals were active or bedded. Unfortunately, the effect of group size on visibility cannot be mitigated by simply altering survey design, but it can be accounted for using statistical corrections. For example, weighting the observed group size by the inverse of the probability that groups of that size would be observed at least once during the survey creates a bias-adjusted group size estimate that can be multiplied by the estimated number of groups to derive a population estimate corrected for group-size bias (Lubow and Ransom 2009).

In contrast to the strong relationships detected between group size and visibility of horses and elk, helicopter aerial surveys of bighorn sheep revealed no effect of group size on group detection (groups of 1–40 sheep), whereas habitat type (topography) and light condition (sun, shade, overcast) were highly significant influences (Bodie et al. 1995). That study reported an 86% visibility rate for sheep located on flats or open slopes, declining to 62% on cliffs or talus. This concurs with my findings, but that effect is attributed only to helicopter surveys in my study because they are often used specifically for areas of complex topography. The same effect could occur in fixed-wing surveys but may be confounded by other effects, such as altitude and maneuverability constraints that apply to most fixed-wing aircraft. The bias attributed to topography could be mitigated by arranging flight transects closer together in areas of complex terrain and using a statistical sampling method that allows for robust estimation when the observation of known groups is expected to be infrequent (multiple occasion mark–recapture, for example).

Sun effect helped explain considerable variation in aerial detection probability in my study, but surprisingly, few studies have investigated this important source of bias. Bodie et al. (1995) reported lighting condition at the location of bighorn sheep groups (sun or shade), but did not address the angle of the sun in relation to the observer. Short and Bayliss (1985) also reported on visibility bias in sunny versus overcast lighting in aerial surveys of kangaroos (*Macropus* sp.), with similar results as Bodie et al. (1995). Surprisingly, I could find no studies that examined the direct role of sun direction on detection probability during aerial surveys. There is considerable variation in the sun effect data for fixed-wing surveys in my study, which is likely due to the overhead wing structure of the Cessna 210 blocking the sun from direct influence on the observers. The sun still caused shading and illumination effects on objects located on the landscape when viewed from this aircraft, which appears to have resulted in the relationship detected. In contrast, helicopters allowed unobstructed skyward views and direct influence of sun on observer vision; consequently, helicopter sun effect data were considerably less variable. Observers did not collect data on the potential bias created by pelage color of animals, which is heterogeneous in most feral horse populations and may be confounded with sun effect. Illuminated white or gray horses can be relatively easy to detect, but the same color horses can be no more visible than black horses when backlit. In contrast, black horses can be no easier to detect in illuminated conditions versus backlit conditions (see Fleming and Tracey [2008] for a detailed description of this effect on aerial observations). Regardless of aircraft, bias introduced by sun effect during aerial surveys can be partially mitigated in many cases by pre-determining linear transects that are oriented parallel to the path of the sun. In areas where transects must follow topographic contours, this effect may be more difficult to mitigate but such topographic features may also block direct sun effect in many circumstances. Regardless, estimating sun direction and including it in candidate models may reduce the amount of unexplained variation in aerial survey data.

Vegetation type and percent cover have been examined in several studies of large mammals, and as in those, feral horses are generally more difficult to detect in tree cover than in areas with no cover or shrub cover. Bayliss and Yeomans (1989) found this relationship with buffalo (*Bubalus bubalis*), as did Samuel et al. (1987) with elk, and Short and Bayliss (1985) and Choquenot (1995) with kangaroos. Choquenot (1995) reported an observation bias by vegetation gradient similar to that in my study, with 23% of kangaroos missed in grass plains, 70% missed in shrubs, and 89% missed in riverine habitat. Unfortunately, avoiding the bias created by vegetation cover is difficult in aerial surveys. The best approach to mitigating this effect may be similar to that suggested for topographic bias, in addition to directly considering this factor in analytical models.

I did not detect an effect of distance from the aircraft, which is surprising and inconsistent with many other studies. Manly et al. (1996), Walter and Hone (2003), and Melville et al. (2008) all report the expected effect that detection decreases as groups increase in distance from the aircraft; likewise, this relationship forms one of the principle foundations of distance sampling (Buckland et al. 1993). I suspect that either the distance categories used were too broad to detect this effect or transects were sufficiently close together for observers to be able to detect most groups. Nonetheless, this effect is likely important and could contribute to bias in surveys that involve wider-spaced transects, as well as in populations that consist of large numbers of small groups, where group size may interact with distance to create negative bias (Caughley 1974).

The effect of snow cover was weakly supported in fixed-wing surveys and followed the trend of more known animals being missed in mixed snow conditions versus 0% or 100% snow. Samuel et al. (1987) reported a similar (though also non-significant) trend in elk surveys with 90% of groups in 0–19% snow cover being detected and only 56% of groups in

20–99% snow cover being detected. Detection of elk in 100% snow cover did not increase. Horse pelage color may also confound the effect of snow on group visibility since many populations have pinto-colored animals that form the same patchy white and brown visual as landscape with partial snow cover, making them more cryptic to aerial observers. The bias of snow effect can most easily be prevented by conducting surveys when snow is not present.

Fleming and Tracey (2008) suggest that psychological and physiological limitations of human observers can potentially lead to observation bias in aerial surveys. Observer perception and brain function can be highly variable and, paired with the physical constraints of various aircraft types and survey parameters (transect width, airspeed, and altitude), can present considerable unmeasured bias (Caughley 1974, Fleming and Tracey 2008). My study held airspeed, AGL, and transect width, as well as aircraft type for fixed-wing and helicopter surveys, relatively constant in order to help control these factors. Use of the individual observer, given the population, as a random effect also helped limit these biases in my analyses; however, the effect of fatigue on observers could not be controlled and influenced group detection during helicopter surveys. Fleming and Tracey (2008) suggest that some survey efforts attempt to minimize fatigue by limiting survey time or taking regular breaks, and the data presented here concur that such strategies may be important. It does present the particular problem of planning aerial surveys with enough hours to complete the survey, but not so many hours as to create the bias presented by fatigue. It may be warranted to replace observers after 3 hr of survey time, though given the variation in observer performance, any change of personnel should be directly addressed in modeling efforts. Surveys designed to optimize detection based on crepuscular wildlife behavior have the added benefit of mitigating observer fatigue by using short flight times interrupted by a long midday break (see Choquenot 1995, Cairns et al. 2008, for example).

The models analyzed in this study produced somewhat disparate results between aircraft types, though the best models for both fixed-wing and helicopter surveys included group size and sun effect. The differing results are likely a product of the behavioral influence of helicopters on horses, the lower AGL of helicopter surveys, greater distance of horses from helicopters at first observation, or confounding effects produced by any combination of factors specific to the physical structure of aircraft types. The behavioral effect found in my helicopter surveys concurs with Linklater and Cameron (2002) and that animals are often easier to detect while moving may help explain the disparity in detection probability between aircraft types during good conditions (Figs. 3 and 4). However, this advantage quickly disappears when conditions become poor and other factors are operating. In addition, such behavioral response may positively bias estimates due to individuals being counted repeatedly (Linklater and Cameron 2002). It does appear from my study that the stationary horse groups, higher altitude observations, and physical structure of the aircraft experienced during fixed-wing surveys likely resulted in fewer overall biases on group detection for animals that were available to both observers.

Estimated detection probabilities are often used to form sightability bias correction models for adjusting population estimates (Caughley 1974, Unsworth et al. 1990, Anderson et al. 1998, Barker 2008). However, such models are dynamic and specific to the time and location of a given survey and may not accurately reflect conditions during future surveys or at other locations (Samuel et al. 1987). The data I present here should not be considered as the basis for a feral horse sightability bias correction model, but rather as a tool for planning future surveys and to help mitigate sources of bias often encountered. It should also be noted that feral horse ranges are diverse in habitat, and some areas may be subject to sources of bias that I investigated whereas others may not. Regardless, valid statistical sampling techniques must be employed in order to help quantify error and correct for various sources of detection heterogeneity that may be present. Graham and Bell (1989), Bayliss and Yeomans (1989), Dawson and Miller (2008), Laake et al. (2008), and Lubow and Ransom (2009) have all applied sampling techniques to feral horse populations and continued expansion, adaptation, and use of these techniques is encouraged.

MANAGEMENT IMPLICATIONS

The common use of raw count, or so called census, aerial surveys can lead to inaccurate and negatively biased abundance estimates when detection of animals on the landscape is influenced by heterogeneous observation conditions. Sun effect, percent snow cover, and observer fatigue, can be minimized or eliminated by carefully planning surveys with transect orientation that allows observers to view animals parallel to the sun's path rather than perpendicular to it, conducting surveys when snow is not present, and limiting helicopter observer time to <4 hr per flight. Other important factors impacting detection, such as group size and topography, cannot be completely overcome with planning, but increasing survey effort by flying more transects with closer spacing may help increase the number of detected groups in the sampled area. Employing aerial survey methodologies that apply statistical sampling techniques is critical to addressing many of the negative biases created by heterogeneity of these factors. Use of such techniques can produce population estimates with quantified errors that will provide better information for decision making and may help bridge gaps between management and the various public stakeholders, politicians, and advocacy groups that frequently take an interest in feral horse management decisions.

ACKNOWLEDGMENTS

I thank the U.S. Bureau of Land Management for cooperation and assistance with special-use aviation needs over the federal lands they manage. B. Cade provided considerable statistical guidance and insight, as well as thoughtful and sometimes disturbing discussion on mixed-effect modeling. B. Lubow contributed greatly to design of aerial surveys. Thank you to D. Bolstad, B. Cade, S. Germaine, A. Kane,

J. Roelle, A. Shepherd, and the anonymous referees who provided critical reviews and comments on earlier drafts of this manuscript. Research was funded by the U.S. Geological Survey Wildlife Program. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

LITERATURE CITED

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

Anderson, C. R., D. S. Moody, B. L. Smith, F. G. Lindzey, and R. P. Lanka. 1998. Development and evaluation of sightability models for summer elk surveys. Journal of Wildlife Management 62:1055–1066.

Barker, R. 2008. Theory and application of mark–recapture and related techniques to aerial surveys of wildlife. Wildlife Research 35:268–274.

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.

Beever, E. A., and J. E. Herrick. 2006. Effects of feral horses in Great Basin landscapes on soil and ants—direct and indirect mechanisms. Journal of Arid Environments 66:96–112.

Beever, E. A., R. J. Tausch, and W. E. Thogmartin. 2008. Multi-scale responses of vegetation to removal of horse grazing from Great Basin (USA) mountain ranges. Plant Ecology 196:163–184.

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 62:372–378.

Buckland, S. T., D. R. Anderson, K. P. Burnham, and J. L. Laake. 1993. Distance sampling: estimating abundance of biological populations. First edition. Chapman and Hall, New York, New York, USA.

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.

Cairns, S. C., G. W. Lollback, and N. Payne. 2008. Design of aerial surveys for population estimation and the management of macropods in the Northern Tablelands of New South Wales, Australia. Wildlife Research 35:331–339.

Caughley, G. 1974. Bias in aerial survey. Journal of Wildlife Management 38:921–933.

Caughley, G., and D. Grice. 1982. A correction factor for counting emus from the air, and its application to counts in Western Australia. Australian Wildlife Research 9:253–259.

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

Choquenot, D. 1995. Species- and habitat-related visibility bias in helicopter counts of kangaroos. Wildlife Society Bulletin 23:175–179.

Dawson, M., and C. Miller. 2008. Aerial mark–recapture estimates of wild horses using natural markings. Wildlife Research 35:365–370.

Department of Commerce. 2010. National Oceanic and Atmospheric Administration Earth System Research Lab Sunrise/Sunset Calculator. <http://www.srrb.noaa.gov/highlights/sunrise/sunrise.html>. Accessed 24 Aug 2010.

Eberhardt, L. L., A. K. Majorowicz, and J.A. Wilcox. 1982. Apparent rates of increase for two feral horse herds. Journal of Wildlife Management 46:367–374.

Fahnestock, J. T., and J. K. Detling. 1999. The influence of herbivory on plant cover and species composition in the Pryor Mountain Wild Horse Range, USA. Plant Ecology 144:145–157.

Fleming, P., and J. Tracey. 2008. Some human, aircraft and animal factors affecting aerial surveys—how to enumerate animals from the air. Wildlife Research 35:258–267.

Garrott, R. A., D. B. Siniff, and L. L. Eberhardt. 1991. Growth rates of feral horse populations. Journal of Wildlife Management 55:641–648.

Gelman, A., and J. Hill. 2007. Data analysis using regression and multilevel/hierarchical models. Cambridge University Press, New York, New York, USA.

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

Harville, D. A. 1977. Maximum likelihood approaches to variance component estimation and to related problems. Journal of the American Statistical Association 72:320–338.

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

Levin, P. S., J. Ellis, R. Petrik, and M. E. Hay. 2002. Indirect effects of feral horses on estuarine communities. Conservation Biology 16:1364–1371.

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. 2009. Validating aerial photographic mark–recapture for naturally marked feral horses. Journal of Wildlife Management 73:1420–1429.

Manly, B. F. J., L. McDonald, and G. W. Garner. 1996. Maximum likelihood estimation for the double-count method with independent observers. Journal of Agricultural, Biological, and Environmental Statistics 1:170–189.

Melville, G. J., J. P. Tracey, P. J. S. Fleming, and B. S. Lukins. 2008. Aerial surveys of multiple species: critical assumptions and sources of bias in distance and mark–recapture estimators. Wildlife Research 35:310–319.

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.

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.

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

Schwarz, C. J., and G. A. F. Seber. 1999. Estimating animal abundance—review III. Statistical Science 14:427–456.

Short, J., and P. Bayliss. 1985. Bias in aerial survey estimates of kangaroo density. Journal of Applied Ecology 22:415–422.

Smith, M. A. 1986. Impacts of feral horse grazing on rangelands—an overview. Journal of Equine Veterinary Science 6:236–239.

Tracey, J. P., P. J. S. Fleming, and G. J. Melville. 2008. Accuracy of some aerial survey estimators—contrasts with known numbers. Wildlife Research 35:377–384.

Unsworth, J. W., L. Kuck, and E. O. Garton. 1990. Elk sightability model validation at the National Bison Range, Montana. Wildlife Society Bulletin 18:113–115.

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. Academic Press, San Diego, California, USA.

Associate Editor: David Forsyth.