*Peer Reviewed*

# Evaluation of Aerial Survey Methods for Dall's Sheep

**MARK S. UDEVITZ,**[1] *United States Geological Survey, Alaska Science Center, Anchorage, AK 99503, USA*
**BRAD S. SHULTS,** *United States National Park Service, Western Arctic National Parklands, Kotzebue, AK 99752, USA*
**LAYNE G. ADAMS,** *United States Geological Survey, Alaska Science Center, Anchorage, AK 99503, USA*
**CHRIS KLECKNER,**[2] *United States Geological Survey, Alaska Science Center, Anchorage, AK 99503, USA*

## Abstract

*Most Dall's sheep (*Ovis dalli dalli*) population-monitoring efforts use intensive aerial surveys with no attempt to estimate variance or adjust for potential sightability bias. We used radiocollared sheep to assess factors that could affect sightability of Dall's sheep in standard fixed-wing and helicopter surveys and to evaluate feasibility of methods that might account for sightability bias. Work was conducted in conjunction with annual aerial surveys of Dall's sheep in the western Baird Mountains, Alaska, USA, in 2000–2003. Overall sightability was relatively high compared with other aerial wildlife surveys, with 88% of the available, marked sheep detected in our fixed-wing surveys. Total counts from helicopter surveys were not consistently larger than counts from fixed-wing surveys of the same units, and detection probabilities did not differ for the 2 aircraft types. Our results suggest that total counts from helicopter surveys cannot be used to obtain reliable estimates of detection probabilities for fixed-wing surveys. Groups containing radiocollared sheep often changed in size and composition before they could be observed by a second crew in units that were double-surveyed. Double-observer methods that require determination of which groups were detected by each observer will be infeasible unless survey procedures can be modified so that groups remain more stable between observations. Mean group sizes increased during our study period, and our logistic regression sightability model indicated that detection probabilities increased with group size. Mark–resight estimates of annual population sizes were similar to sightability-model estimates, and confidence intervals overlapped broadly. We recommend the sightability-model approach as the most effective and feasible of the alternatives we considered for monitoring Dall's sheep populations.* (WILDLIFE SOCIETY BULLETIN 34(3):732–740; 2006)

## Key words

*aerial survey, Alaska, Dall's sheep, double observer, mark–recapture, mark–resight,* Ovis dalli*, population estimate, sightability model, visibility bias.*

Aerial surveys are commonly used to estimate population size for various species of ungulates (Thompson et al. 1998), and it is widely recognized that these surveys are subject to visibility bias (Caughley 1974, Pollock and Kendall 1987). Dall's sheep (*Ovis dalli dalli*) may be less subject to visibility bias than many species because Dall's sheep use relatively exposed habitats, and their white color often provides a strong contrast to their background when snow and ice are not present. Aerial surveys of mountain sheep are usually conducted by attempting to completely cover at least a sample of survey units and attempting to find every sheep in each of those units (McDonald et al. 1990, Neal et al. 1993, Bodie et al. 1995, Ayres 1996, Whitten 1997). Unadjusted total numbers of detected sheep are commonly used as either population estimates or indices for those units (Ayres 1996), under the assumption that either essentially all sheep are detected or that there is no systematic change in detectability between surveys (Udevitz 2002). However, detectability of mountain sheep can be influenced by size and composition of groups, activity, habitat, light conditions, position of sheep relative to aircraft, and topographic position (McDonald et al. 1990, Strickland et al. 1992, Bodie et al. 1995).

Four basic approaches have been used to account for visibility bias in aerial surveys of mountain sheep. In double-sampling methods (Gasaway et al. 1986, McDonald et al. 1990, Whitten 1997), units are first surveyed at some standard intensity, and then, a subsample from those units is resurveyed at a higher intensity. Detection probability is estimated from the ratio of the 2 total counts in the double-sampled areas. The sight–resight method (McDonald et al. 1990, Strickland et al. 1992) usually uses crews in 2 aircraft, with one crew surveying a unit as soon as possible after the other. Rather than using the total counts for each unit, however, this approach uses information on which groups of sheep are detected by each of the crews. Detection probabilities are estimated based on the proportion of groups detected by one crew that also are detected by the other crew. The mark–resight method (Neal et al. 1993, Douglas and Longshore 1995, Andrew et al. 1997) uses physical marks, such as collars, on the sheep and estimates detection probabilities based on the proportion of marked sheep detected during the survey. Finally, in the sightability-model approach (Samuel et al. 1987, Bodie and Oldenburg 1995, Bodie et al. 1995), logistic regression is used to model detection probabilities as functions of covariates, such as group size or vegetation cover, for groups that are detected versus those that are not detected in trial surveys with radiocollared sheep. The model is then used to estimate detection probabilities in future surveys based on covariates associated with detected groups in those surveys.

Estimates of Dall's sheep population sizes and age and sex compositions in the Baird Mountains, Alaska, USA, have been based on unadjusted counts from intensive fixed-wing surveys conducted annually since 1986 (Carroll 1993, Ayres 1996). These counts indicated precipitous population declines following a series of severe winters in the early 1990s. Based on these data, subsistence and sport-hunting seasons were closed by emergency order in 1991, resulting in substantial management controversy. Annual variation in subsequent counts from the Baird Mountains was not consistent with reasonable assumptions about reproduc-

---

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

[2] *Present address: Central Middle School of Science, Anchorage, AK 99501, USA*

tion and survival (B. S. Shults, United States National Park Service, unpublished data), raising questions about the accuracy and precision of the survey methodology used to monitor the population. Our objectives were to attempt to use each of the basic aerial survey methods to estimate the western Baird Mountains Dall's sheep population, to assess the validity of key assumptions required by the methods, and to evaluate the relative feasibility of using these methods as monitoring protocols.

## Study Area

We surveyed 1,828 km$^2$ in the western Baird Mountains, Alaska (including the Maiyumerak Mountains), USA. The survey area was partitioned into 18 count units that had been used in previous surveys to monitor the Dall's sheep population (Singer et al. 1983, Carroll 1993, Ayres 1996). Count units ranged from 60–130 km$^2$, and each could be surveyed in about 1–1.5 hours. The Baird Mountains separate the Noatak and Kobuk River valleys and extend east from the headwaters of the Agashashok and Eli Rivers to the Redstone and Ambler Rivers. Topography and vegetation is characterized by rolling tussock tundra interspersed with knolls and mountains reaching 700–915 m in elevation (Ayres 1986). Dall's sheep density was highest in the Maiyumerak Mountains, east to the Nakolik River, and south to the headwaters of the Squirrel River. Sheep also inhabited other areas in the Baird Mountains, but their distribution was patchy and at a lower density than in the survey area (National Park Service, unpublished data).

## Methods

We captured sheep by helicopter net gunning (Jessup et al. 1988) and fitted them with black radiocollars during 21–29 March 2000. We distributed captures throughout the winter range of the population and instrumented males and females to approximate the sex ratio observed in previous surveys. We radiocollared additional sheep during 21–30 March 2001 and 3–5 April 2002 to maintain a sample of approximately 30–45 radiocollared sheep in the study area. The project study plan was reviewed and approved by the Alaska Science Center Animal Care and Use Committee on 8 December 1999.

We conducted aerial surveys during 8–20 July 2000, 9–18 July 2001, 6–14 July 2002, and 9 July 2003, after sheep had formed postlambing aggregations. We did not conduct surveys in fog or if wind speed was >37 km/hour. Crews in 1–2 fixed-wing aircraft used radiotelemetry to locate every radiocollared sheep in each unit before it was surveyed. We used 3 additional Piper PA-18 Super Cubs (Piper Aircraft Corporation, Lock Haven, Pennsylvania) to survey units according to the standard protocol used in previous years for annual monitoring surveys. Survey crews consisted of a pilot and a secondary observer, seated behind the pilot. The same 3 aircraft and pilots were used each year. Pilots were highly experienced with Dall's sheep surveys and served as primary observers. Secondary observers and their experience levels varied among years. In addition to assisting with observations, secondary observers also operated the radiotelemetry receiver and recorded global positioning system locations and characteristics associated with each detected group of sheep.

We randomly assigned crews to units and surveyed units in a random order each year. The survey procedure generally consisted of flying along elevation contours at as low an altitude as practical and attempting to observe all sheep in the unit. Exact flight paths and air speeds were left to each pilot's discretion, subject to the requirement that all terrain in each unit be safely and thoroughly searched for sheep. When sheep were observed, we recorded the group location, size, composition, and other covariates related to habitat and behavior. Then, we determined whether any sheep in the group were radiocollared and used radiotelemetry to identify those individuals. After completing the survey of a unit, we used radiotelemetry to locate any radiocollared sheep in that unit that had not been observed and recorded their associated covariates.

For a randomly selected subset of the units surveyed each day in 2001 and 2002, we used a Robinson R-44 helicopter (Robinson Helicopter Company, Torrance, California) to resurvey the unit. The helicopter entered the unit immediately after the fixed-wing crew completed its survey and followed the same survey protocol: identifying any radiocollared sheep that were detected and using radiotelemetry to locate any missed sheep immediately after completion of the survey of the unit. The helicopter crew consisted of a pilot who was highly experienced with Dall's sheep surveys and a less-experienced, secondary observer, seated beside the pilot. We used the same helicopter and crew both years. At the end of each day, helicopter and fixed-wing crews met and attempted to determine which groups were detected by each of the crews based on the location, size, and composition of the detected groups.

After covering all the survey units each year, we continued surveying units as weather and resources allowed, obtaining additional sightability data. We surveyed units in random order, subject to the constraint that no crew could survey any unit more than once each year.

### *Data Analysis*

We used the paired fixed-wing and helicopter surveys of units containing $\geq 1$ detected sheep to evaluate the potential for using helicopter counts to adjust for any sightability bias in the fixed-wing counts. We assessed differences between the fixed-wing and helicopter surveys with paired *t*-tests.

We attempted to determine which groups were detected by each crew by comparing their mapped locations and the observations of group size and composition. Initial attempts did not use any information about radiocollared sheep, mimicking the procedure that would have to be used if there were no marked individuals. To assess our ability to determine which groups were detected by each crew, we used the groups containing radiocollared sheep detected during a fixed-wing survey of a unit and then relocated during the paired helicopter survey of that unit. We compared sizes and compositions of the groups associated with each of the radiocollared sheep at the 2 observation times. We used logistic regression with likelihood-ratio chi-square tests (Hosmer and Lemeshow 2000:12–16) to assess the relations of elapsed time and initial group size to the probability of a group changing in size between observation times. We log-transformed group sizes for this analysis. For those groups that changed in size, we used linear regression to assess the relation between elapsed time and the magnitude of the proportional change. We used a likelihood-ratio chi-square test to assess the difference in proportions of groups that increased versus those that decreased in size.

For mark–resight estimation of population size, we used the first fixed-wing survey of each unit, each year. Because of interruptions due to inclement weather, this included surveys of units on 8, 19, and 20 July in 2000; 9, 10, 11, and 14 July in 2001; and 6, 7, and 10 July in 2002. We used the bias-corrected Lincoln–Petersen estimator (Chapman 1951) to estimate the number of individuals in the population during each survey.

For comparison with our sightability-model confidence intervals (see below), we used a bootstrap procedure to estimate confidence intervals for the mark–resight population estimates. For a given year, we obtained a bootstrap sample of detected, marked individuals by random drawing with replacement from a hypergeometric ($N_i$, $M_i$, $n_i$) distribution, where $N_i$ was the original mark–resight estimate of population size, $M_i$ was the number of radiocollared sheep in the population, and $n_i$ was the number of sheep detected in the resight survey for the *i*th year. We reestimated the number of individuals in the population and the detection probability for each bootstrap sample ($n = 2{,}000$) and took the 2.5 and 97.5 percentiles of the resulting distributions as 95% confidence limits.

In 2001 and 2002, we were able to complete 1–2 additional fixed-wing surveys of some units. We used the data from all the resight surveys in those years to also calculate Bowden mark–resight estimates of population size, along with their log-based confidence intervals (Bowden and Kufeld 1995).

An essential assumption of the mark–resight method is that there is either a random distribution of marks or random detection during the survey (Seber 1982). One testable consequence of marks being randomly distributed is that the expected proportion of sheep with marks will be the same for all group sizes. We used the observed distributions of collars in detected groups to assess whether this consequence held for our surveys. We combined group sizes into categories so that expected numbers of marked individuals were $\geq 5$ for each category. We used the likelihood-ratio chi-square test to evaluate differences in marked proportions.

We used all observations of marked groups during 2000–2003 ($n = 254$) for developing a sightability model, but only 38 of these were undetected. With that sample size for undetected groups, our data could support reliable inference about effects on sightability for no more than 2 or 3 single-degree-of-freedom variables (Harrell 2001:60–61). Therefore, we used these data to formally test hypotheses only about the effects of group size and aircraft type (i.e., fixed-wing vs. helicopter). We used likelihood-ratio chi-square tests to assess relations of these variables to detection probabilities in a logistic-regression model (Hosmer and Lemeshow 2000) that contained both of these effects and their interaction. Group size was log-transformed. We removed variables with likelihood-ratio $P$-values $>0.05$ to obtain a final sightability model for estimating population size. We evaluated the fit of the final model with the Hosmer–Lemeshow goodness-of-fit test (Hosmer and Lemeshow 2000:147–150), the Osius and Rojek goodness-of-fit test (Hosmer and Lemeshow 2000:153–154), and the Stukel tests (Hosmer and Lemeshow 2000:154–155).

To estimate population size with the sightability model, we used the same survey data used for the Chapman mark–resight estimates (i.e., the first survey of each unit by a fixed-wing aircraft each year). We estimated population sizes by summing the size of each detected group, divided by its estimated probability of detection (Samuel et al. 1987). The number of groups was estimated by summing the reciprocal of the estimated detection probability for each detected group. We estimated mean group size for each survey as the estimated population size, divided by the estimated number of groups. For comparison with the mark–resight estimates of overall detection probabilities, we estimated the proportion of the population detected as the number of detected individuals, divided by the estimated population size.

We used a bootstrap procedure similar to that used by Cogan and Diefenbach (1998) to estimate confidence intervals for the sightability-model estimates of population size, mean group size, and the proportion of the population detected. For each year, we generated a bootstrap survey sample of detected groups by taking a simple random sample with replacement from the original survey data, until the number of individual sheep exceeded the number detected in the original survey. We then excluded the last-selected group from the sample if the resulting sample size discrepancy was smaller than what would be obtained if that group was retained. We obtained a bootstrap model-development sample by drawing a simple random sample with replacement from the original set of observations on marked groups used to develop the sightability model until we had the same number of observations as in the original data set. We then reestimated the parameters of the model with the bootstrap model-development sample and applied that to the bootstrap-survey sample of the detected groups to obtain bootstrap-sample estimates of population size, mean group size, and the proportion of the population detected. We repeated this 2,000 times and took the 2.5 and 97.5 percentiles of the resulting distribution of estimates as the 95% confidence limits.

## Results

We completed at least one fixed-wing survey of all 18 Baird Mountains survey units each year, except 2003, when we were only able to complete surveys of 9 units because of inclement weather. Additional fixed-wing surveys of some units were completed in 2001 and 2002. Numbers of radiocollared sheep in the study area during the survey periods ranged from 35–43, with 72–80% of these being females. Overall, we conducted 109 fixed-wing surveys of Baird Mountains units, with a mean survey rate of 0.82 minutes/km$^2$ (SE = 0.02). Unadjusted total counts of sheep detected during the first fixed-wing survey of each unit increased from 522 in 2000 to 682 in 2002 (Table 1).

We completed 30 helicopter surveys of Baird Mountains units in 2001 and 2002 with a mean survey rate of 0.75 minutes/km$^2$ (SE = 0.04). These included 24 paired surveys of units that contained sheep and could be used for direct comparison of counts by helicopter and fixed-wing crews. Most counts from the helicopter were greater than or equal to the corresponding counts from a fixed-wing aircraft, but the mean difference in the counts was negligible ($\bar{x}$ difference $= 0.25$ sheep, $t_{23} = 0.09$, $P = 0.93$), and 7/24 (29%) of the counts from fixed-wing aircraft were greater than the corresponding counts from the helicopter (Fig. 1). Helicopter survey rates on these units were similar to those of the fixed-wing aircraft ($\bar{x}$ difference $= 0.02$ min/km$^2$, SE $= 0.04$).

***Table 1.*** Chapman (1951) mark–resight estimates of detection probabilities and population sizes from fixed-wing surveys of Dall's sheep in the western Baird Mountains, Alaska, USA.

| Year | Marked sheep available | Marked sheep detected | Total sheep detected | Detection probability | | Population size | |
|---|---|---|---|---|---|---|---|
| | | | | Estimate | 95% CI | Estimate | 95% CI |
| 2000 | 30 | 27 | 522 | 0.90 | 0.80–1.00 | 578 | 522–648 |
| 2001 | 42 | 36 | 616 | 0.86 | 0.76–0.95 | 716 | 646–803 |
| 2002 | 41 | 36 | 682 | 0.88 | 0.78–0.98 | 774 | 699–868 |

In some cases, the determination of which groups were detected by each crew seemed relatively straightforward, based only on location and group characteristics. In other cases, however, even though both crews apparently saw similar total numbers of sheep in the same general area, almost none of the groups matched.

In units that were resurveyed with the helicopter, there were 41 groups observed by fixed-wing crews that contained radiocollared sheep in the initial survey. For 10 of those groups, by the time the radiocollared sheep were relocated by the helicopter, either radiocollared sheep initially in separate groups joined together or radiocollared sheep initially together in a single group split into different groups, so it was not possible to define groups by the radiocollared sheep they contained. Of the 31 groups that could be defined by the radiocollared sheep they contained, 17 had changed in size by time the helicopter located them. So, at most, only 14 (or 34%) of the original 41 groups detected in a fixed-wing survey existed in their original form by the time of the helicopter survey.

For the 17 identifiable groups that changed in size, the average change was 42% (SE = 0.05) of the original group size. Twelve (71%) of these groups became smaller ($\chi_1^2 = 2.88$, $P = 0.09$). The probability of a group changing in size between observation times increased with initial group size ($\chi_1^2 = 9.07$, $P < 0.01$; Fig. 2), but we did not detect any relation with elapsed time ($\chi_1^2 = 0.91$, $P = 0.34$). For those groups that changed in size, there was no apparent relation between the magnitude of the proportional change and the amount of elapsed time (slope = 0.00, $t_1 = 0.24$, $P = 0.81$).

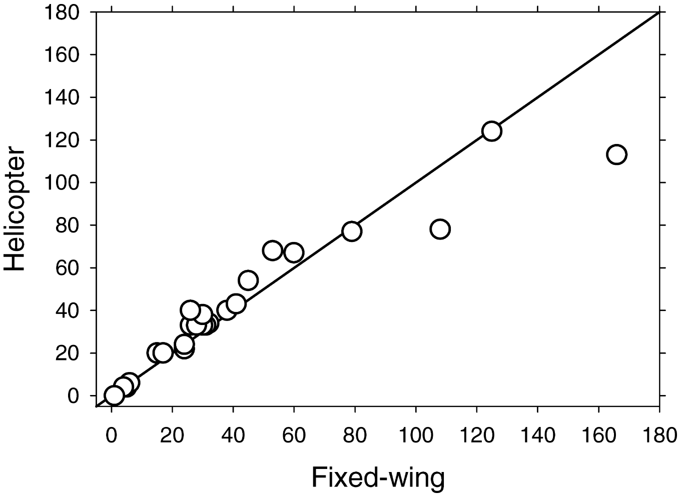

***Figure 1.*** Comparison of counts of Dall's sheep from a Robinson R44 helicopter and Piper PA-18 Super Cubs in double surveys of units in the western Baird Mountains, Alaska, USA, 2001–2002.

Proportions of radiocollared sheep detected in the initial fixed-wing surveys ranged from 0.86 to 0.90 (Table 1). Mark–resight population estimates increased each year during the study period, primarily because of increases in the numbers of detected sheep rather than differences in estimated detection probabilities (Table 1). Bowden estimates of population size (Table 2) were similar to Chapman estimates. The confidence interval was smaller for the Bowden than for the Chapman estimate in 2002, when we surveyed most units 3 times ($\bar{x} = 2.6$ surveys per unit), but not in 2001 (Tables 1, 2), when we surveyed most units twice ($\bar{x} = 2.2$ surveys per unit).

Larger groups tended to have smaller proportions of radiocollared sheep in 2000 ($\chi_3^2 = 9.58$, $P = 0.02$; Fig. 3). Radiocollared sheep were more evenly distributed relative to group sizes in 2002 ($\chi_3^2 = 5.22$, $P = 0.16$), but a residual tendency for the larger groups to have smaller proportions of radiocollared sheep was still evident (Fig. 3). We did not attempt this analysis for 2001 because group sizes were not recorded for many of the radiocollared sheep located during the initial surveys that year.

Overall, we detected 85% (216) of the 254 groups considered in sightability trials. We detected all groups ($n = 33$) of $\geq 19$ sheep (Fig. 4). Detection rates increased with log group size ($\chi_1^2 = 19.45$, $P < 0.01$), but the effects of aircraft type and its interaction with log group size were not significant ($\chi_2^2 = 1.91$, $P = 0.38$; Fig. 4). Therefore, we modeled sightability as a function of log group size only (Fig. 5). Neither the Hosmer–Lemeshow test ($\chi_7^2 = 8.51$, $P = 0.29$), Osius and Rojek test ($z = -0.64$, $P = 0.52$), or the Stukel test ($\chi_1^2 < 0.001$, $P = 1.00$) indicated any lack of fit for this final model.

Sightability-model estimates of population size were similar to the mark–resight estimates, but confidence intervals were smaller (Table 3). Mean group sizes tended to increase during the study period, but that did not have a large effect on the estimated overall proportion of sheep detected (Table 3).

## Discussion

Intensive aerial surveys are commonly used to monitor Dall's sheep populations in Alaska, USA (Whitten 1997), but the proportion of the population detected in these surveys has rarely been estimated. Based on proportions of marked sheep detected in several unpublished surveys, Whitten (1997) estimated that fixed-wing surveys in Alaska detect about 70–80% of the Dall's sheep present. Using previously published data for fixed-wing surveys of Dall's sheep in Alaska, we estimated overall detection rates ranging from 58–76% (Table 4). These rates are generally higher than those from aerial surveys of mountain sheep outside Alaska (Table 4). Our estimates of detection rates from previous Alaskan surveys (Table 4) may be somewhat high because they are based on

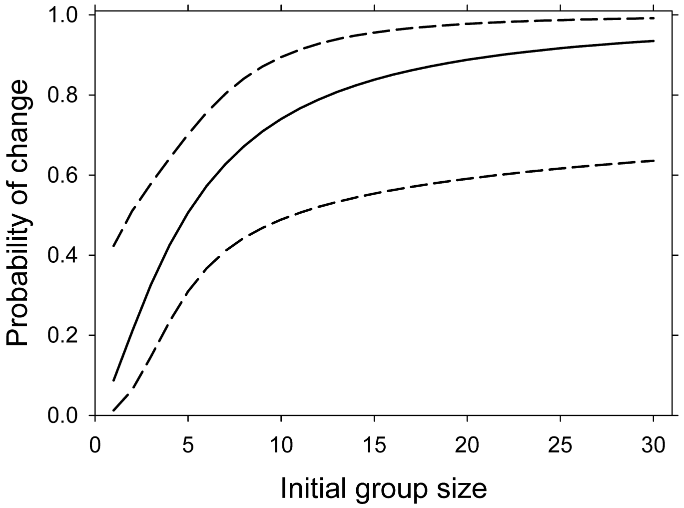

***Figure 2.*** Estimated probability of Dall's sheep groups changing in size between paired observations during aerial surveys in the western Baird Mountains, Alaska, USA, 2001–2002. Dashed lines indicate 95% confidence interval.

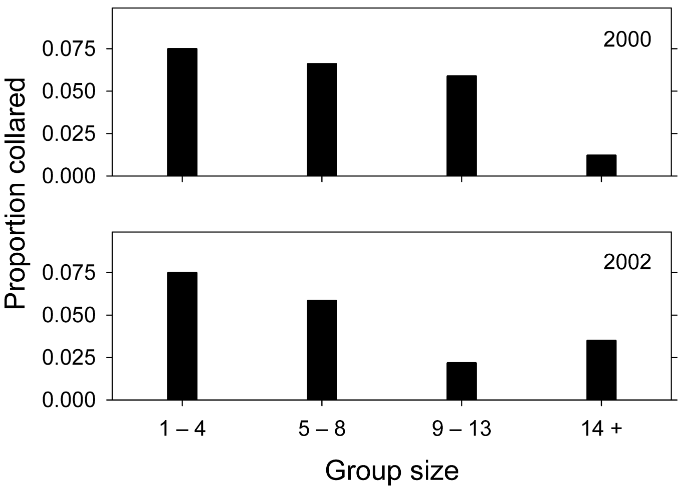

***Figure 3.*** Distribution of radiocollared sheep relative to group size during 2 years of fixed-wing surveys in the western Baird Mountains, Alaska, USA.

population sizes that are likely underestimated (McDonald et al. 1990, Strickland et al. 1992, 1994, Whitten 1997). However, sheep may tend to be more detectable in Alaska because Dall's sheep are white and there is a relative lack of trees and tall brush in most of their summer habitats (Whitten 1997). Detection rates in our surveys (Tables 1, 3) were even higher than the apparent rates for other Alaskan surveys (Table 4). Our surveys averaged more time per unit map area than some other Alaskan surveys (Table 4). This may at least partly explain our higher detection rates in those cases, but times per unit map area may not be fully comparable measures of intensity because they do not account for differences in terrain complexity (Whitten 1997). Sheep may also be more detectable in the Baird Mountains because the proportion of extremely rugged terrain is smaller than in many other Alaskan summer sheep ranges. Still, our fixed-wing surveys did not detect all of the sheep in surveyed units. The methodology of these surveys was equivalent to what has been used previously to monitor the Baird Mountains Dall's sheep population (Carroll 1993, Ayres 1996) and is commonly used to monitor other Alaskan Dall's sheep populations (Whitten 1997). Unadjusted counts from our fixed-wing surveys underestimated population sizes by about 10–15% (Tables 1–3).

Detectability of mountain sheep is generally considered to be higher from helicopters than fixed-wing aircraft (Strickland et al. 1992, Bodie et al. 1995, Whitten 1997). Helicopters provide greater visibility, slower flying speeds, and higher maneuverability relative to fixed-wing aircraft, but these advantages may only become important in the most rugged terrain (Whitten 1997). The most important factor affecting differences in detectabilities appears to be the relative intensity of the surveys. In 2 studies that provided comparisons of fixed-wing aircraft and helicopters in Dall's sheep surveys, detection rates were substantially higher for the helicopter surveys when their intensity averaged 1.42 (Whitten 1997) to 1.58 (Strickland et al. 1992) times the fixed-wing survey intensity. However, overall detection rates were only slightly higher for the helicopter survey when its intensity was only 1.21 times the fixed-wing survey intensity, and counts by the fixed-wing crews were higher than the helicopter in 12/34 (35%) of the units surveyed at those intensities (Whitten 1997). Average intensities in our surveys were about the same for helicopter and fixed-wing surveys. Counts from our helicopter surveys were not consistently higher than counts from our fixed-wing surveys of the same units (Fig. 1) and detection probabilities were similar for the

***Table 2.*** Bowden (Bowden and Kufeld 1995) mark–resight estimates of population sizes from fixed-wing surveys of Dall's sheep in the western Baird Mountains, Alaska, USA.

| Year | Marked sheep available | Detections of marked sheep | Total detections of sheep | Population size | |
|---|---|---|---|---|---|
| | | | | Estimate | 95% CI |
| 2001 | 43 | 80 | 1,289 | 690 | 602–779 |
| 2002 | 41 | 100 | 1,751 | 734 | 669–799 |

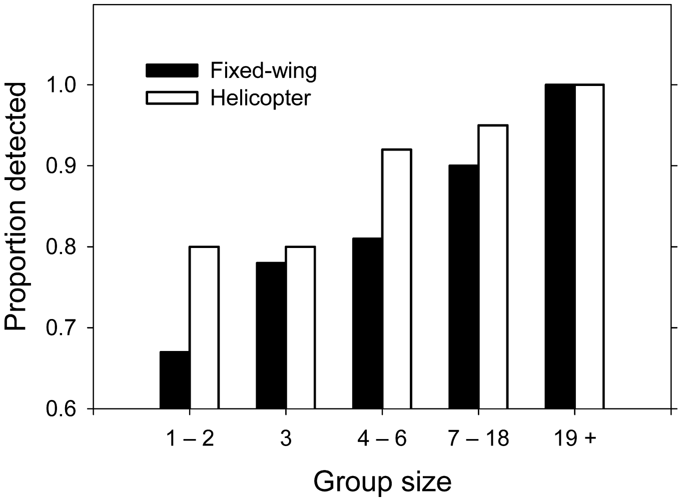

***Figure 4.*** Proportions of Dall's sheep groups detected by crews in a Robinson R44 helicopter and Piper PA-18 Super Cubs during sightability trials in the western Baird Mountains, Alaska, USA, 2000–2003.

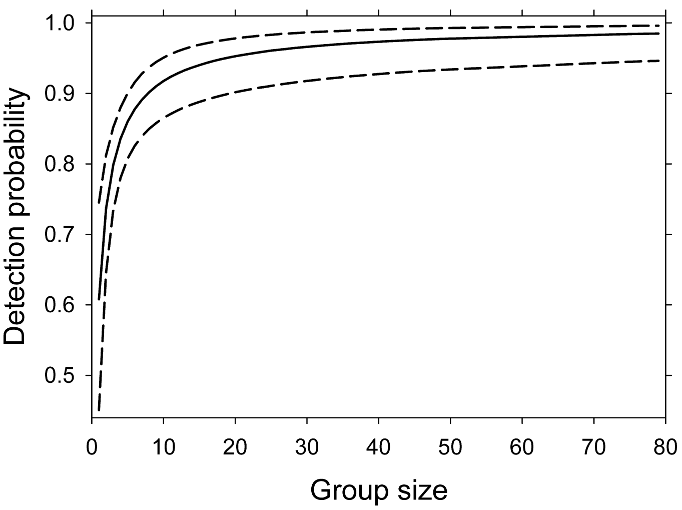

***Figure 5.*** Predicted detection probabilities from a logistic-regression sightability model for aerial surveys of Dall's sheep in the western Baird Mountains, Alaska, USA, 2000–2003. Dashed lines indicate 95% confidence interval.

2 aircraft types. Our detection probabilities were <1 for all but the largest group sizes (Fig. 5). In other cases where detection probabilities have been estimated, helicopter surveys have never detected all of the sheep present (Strickland et al. 1992, Neal et al. 1993, Bodie et al. 1995, Douglas and Longshore 1995).

Our average helicopter survey intensity was less than the average intensities of 0.97–1.34 minutes/km$^2$ for helicopter surveys of Dall's sheep by Whitten (1997) but greater than the intensity of 0.30 minutes/km$^2$ used by Strickland et al. (1992). Our survey intensities could have been increased, but our experienced pilots were already using all the time that they believed was necessary to thoroughly search all the terrain in each unit for sheep. Therefore, it is unlikely that additional search time would have substantially increased our detection probabilities. It appears that thorough searches of all terrain in Baird Mountains survey units require about the same amounts of time for experienced pilots in either fixed-wing aircraft or helicopters, and detection rates for these 2 aircraft types will be about the same at the required survey intensity.

Our attempt to implement the double-sampling approach was based on resurveying a subsample of the units with a helicopter, using the ratio of the 2 total counts in these units as an estimate of detection probability for the fixed-wing survey and using that estimate to adjust counts in the remaining units that were only surveyed with a fixed-wing aircraft. An essential assumption of this approach is that all the sheep in the double-sampled units were detected by the helicopter crew. This assumption has not generally been met by helicopter surveys (Strickland et al. 1992, Neal et al. 1993, Bodie et al. 1995, Douglas and Longshore 1995) and was not met by our helicopter surveys. Therefore, total counts from helicopter surveys cannot be used to provide reliable estimates of detection probabilities for fixed-wing surveys.

The sight–resight method potentially offers an approach that would also be based on the subsample of units resurveyed with the helicopter but would not require all the sheep in those units to be detected in the helicopter survey. However, one of the most basic requirements of the sight–resight method is that group characteristics remain unchanged between times when the first and second observers have opportunities to detect a group. We found that group characteristics were highly dynamic, and most groups had changed in some way by the time they could be observed by a second crew. Groups tended to be smaller when sheep were located by the second crew, and most changes occurred for the larger groups that often required more intensive circling to determine their composition. This suggests that, in many cases, the first aircraft may have disturbed the sheep, causing groups to fragment. The behavioral response of some sheep may have been increased because of their previous experience with the helicopter net gunning used for capture. It is also possible that some groups began fragmenting even before they were detected by the first approaching aircraft, but we did not have any data to address these potential issues.

In any case, the sight–resight method appears to be infeasible with our survey protocol because groups observed by the first aircraft often did not even exist by the time the second aircraft arrived. It might be possible to use this approach if the protocol was changed so that the first aircraft caused fewer disturbances, and the second aircraft followed more closely. With that change in protocol, only the second aircraft could be used to obtain detailed composition data, which are usually additional objectives in these surveys. Dall's sheep surveys conducted by McDonald et al. (1990) and Strickland et al. (1992, 1994) reduced the potential of disturbance in their implementations of a sight–resight approach by having the first crew fly at a higher altitude, with less circling of groups and a lower search intensity than the second crew. It should be noted that the increased margin of safety associated with higher-altitude flying also was part of the motivation for their approach (Strickland et al. 1992). Average search intensities for first crews were substantially lower than ours in the surveys by Strickland et al. (1992, 1994) but were higher than ours in the survey by McDonald et al. (1990; Table 4). Overall detection rates for the first crews were lower than ours in all of these surveys (Table 4).

Without any marked animals, McDonald et al. (1990) and Strickland et al. (1992, 1994) could not assess their accuracy in

***Table 3.*** Sightability-model estimates of mean group sizes, population sizes, and proportions of the population detected in fixed-wing surveys of Dall's sheep in the western Baird Mountains, Alaska, USA.

| | **Mean group size** | | **Proportion detected** | | **Population size** | |
|---|---|---|---|---|---|---|
| **Year** | **Estimate** | **95% CI** | **Estimate** | **95% CI** | **Estimate** | **95% CI** |
| 2000 | 4.45 | 3.42–6.02 | 0.87 | 0.82–0.92 | 600 | 568–642 |
| 2001 | 5.02 | 3.96–6.39 | 0.88 | 0.83–0.92 | 700 | 668–743 |
| 2002 | 6.63 | 5.34–8.23 | 0.90 | 0.86–0.94 | 758 | 728–797 |

***Table 4.*** Detection rates for aerial surveys of mountain sheep in North America.

| Location | Species | Aircraft type | Detection rate | Survey intensity (min/km$^2$) | Reference |
|---|---|---|---|---|---|
| California | Bighorn | Helicopter | 0.21[a,b] | n.r.[c] | Douglas and Longshore 1995 |
| Colorado | Bighorn | Helicopter | 0.57[a,b] | 2.60 | Neal et al. 1993 |
| Alaska | Dall's | Fixed-wing | 0.58[b,d] | 0.19 | Strickland et al. 1992 |
| Idaho | Bighorn | Helicopter | 0.66[b,d] | n.r. | Bodie et al. 1995 |
| Alaska | Dall's | Fixed-wing | 0.68[d] | 1.00 | McDonald et al. 1990 |
| Alaska | Dall's | Fixed-wing | 0.76[d] | 0.38 | Strickland et al. 1994 |
| Alaska | Dall's | Fixed-wing | 0.76[b,d,e] | 0.84[b,e] | Whitten 1997 |
| Alaska | Dall's | Fixed-wing | 0.88[a,b] | 0.82[b] | This study |

[a] Detection rate calculated as the proportion of marked sheep detected.

[b] Average of rates from multiple surveys.

[c] n.r. = value not reported.

[d] Detection rate calculated as the number detected, divided by the total estimated population size.

[e] Only includes surveys of Unit 20A and the Delta Controlled Use Area judged by Whitten (1997) to have adequate intensities for associated helicopter surveys.

identifying which groups were detected by each crew. Groups recorded in close proximity were often "pooled to account for movement, aggregation, and segregation between surveys based on deductive judgment of the survey crews" (McDonald et al. 1990:178). When there was doubt, crews used the conservative approach of assuming groups were seen in both surveys, which likely overestimated detection probability and, consequently, underestimated population size (McDonald et al. 1990, Strickland et al. 1992, 1994). The higher detection-probability estimates from our surveys are consistent with this assertion, although our detection probabilities may also have been higher due to less-rugged habitat or higher search intensities in at least some surveys. McDonald et al. (1990) and Strickland et al. (1992, 1994) concluded that a sight–resight approach with conservative rules for reconciling observations gave more reasonable population estimates than a double-sampling approach.

Mark–resight methods have been used for estimating mountain sheep populations in Colorado (Neal et al. 1993) and California (Douglas and Longshore 1995, Andrew et al. 1997). The precision of these methods is a function of the number of marked individuals and the resighting probability (Seber 1982). Mark–resight estimates from single resighting surveys typically have poor precision because the probability of resighting marked sheep in a single survey is usually low (Table 4). For example, confidence intervals for mark–resight estimates based on single resight surveys of mountain sheep by Douglas and Longshore (1995) and Andrew et al. (1997) were larger than the associated population estimates. The probability of resighting marked sheep and, consequently, the precision of the population estimate can be increased by conducting multiple resighting surveys (Neal et al. 1993). Neal et al. (1993) obtained a confidence interval that was only 20% of their population estimate by conducting 14 resighting surveys of a mountain sheep population in Colorado, USA. Our experience indicates that, because of the large area to be covered and infrequent suitable weather, it would not be possible to complete more than one resight survey of the Baird Mountains population in most years. However, the relatively high detection probabilities for this population made it possible for us to obtain estimates from single resight surveys that had precisions (Table 1) comparable to those of Neal et al. (1993).

In years when it is possible to complete additional resight surveys of at least some units, it is possible to use the information from the additional resights. Detection probabilities for individual sheep will depend on the number of times the units they are in are surveyed. Most sheep will tend to stay in the same unit throughout the survey period, so there will be substantial heterogeneity in detection probabilities if all units are not surveyed the same number of times. The Bowden estimator (Bowden and Kufeld 1995) accommodates this heterogeneity, and the precision of the mark–resight estimates can be improved by using this estimator to incorporate information from the additional resights. However, it appears that most units must be surveyed at least 3 times to obtain this improvement. This type of heterogeneity would not be a problem if the same number of additional surveys could be completed for all units. In that case, the joint hypergeometric estimator (Bartmann et al. 1987) would be appropriate and would likely provide a more substantial improvement in precision.

Detection during our resight surveys was not random because detection probabilities were a function of group size (Fig. 5). Consequently, our mark–resight estimates must rely on the assumption of randomly distributed marks (Seber 1982). We attempted to obtain a random distribution of marks by collaring male and female sheep in proportion to their representation in the population, collaring only one sheep per group, and distributing collars throughout the population before they moved to their summer range and formed the postlambing aggregations that existed during the resight surveys. In spite of these efforts, collars apparently were not randomly distributed by the time of our first resight survey in 2000, although their distribution may have become random in subsequent years (Fig. 3). If detection probability increased with group size and a smaller proportion of the individuals in larger groups were marked, this would tend to induce a positive bias in the mark–resight estimates. However, detection probabilities were large enough in our surveys that the magnitude of any such bias is likely to be small.

Detection probabilities in aerial surveys of animals that occur in groups commonly depend on group size (Samuel et al. 1987, Anderson et al. 1998, Cogan and Diefenbach 1998, McCorquodale 2001), although Bodie et al. (1995) did not find an effect of group size on detectability in their surveys of mountain sheep in Idaho. Bodie et al. (1995) suggested that the compactness of their groups and high overall sightability may have reduced group-size effects to a level that could not be detected with their sample size. Our study and others (McDonald et al. 1990, Strickland et al. 1992, 1994) have consistently found group size to be an important factor affecting detectability of Dall's sheep in Alaska, USA. The exact form of the detection function may vary with survey protocol and location, however. Our sightability function (Fig. 5) is higher than the function presented by McDonald et al. (1990) for group sizes <20 and lower than theirs for larger groups. Both functions estimate a detection probability of 0.95 for groups of 20 sheep. Our sightability function appears to be consistently higher than that of Strickland et al. (1992) who reported detection probabilities of 0.33 for single sheep and 0.90 for groups of 40.

A key assumption of the sightability-model approach is that the covariates included in the model can account for any changes in detectability that might occur between surveys. In addition to group size, factors related to habitat characteristics and animal activity have often been included in sightability models for other ungulate species (Samuel et al. 1987, Otten et al. 1993, Anderson and Lindzey 1996, Anderson et al. 1998, Cogan and Diefenbach 1998, McCorquodale 2001). Bodie et al. (1995) found habitat type and behavior were the most important factors affecting detectability of mountain sheep in Idaho. We did not attempt to evaluate factors other than group size and aircraft type because our sample size would not support reliable inference for additional variables (Rexstad et al. 1988, Harrell 2001). Furthermore, because overall detectability is so high, it would be difficult to obtain enough observations on undetected groups to support rigorous analysis of additional variables for the Baird Mountains Dall's sheep population. In 4 field seasons, we were able to accumulate on the order of twice the total number of observations (detected + undetected groups) typically used to develop sightability models (e.g., Bodie et al. 1995, Anderson and Lindzey 1996, Anderson et al. 1998, Cogan and Diefenbach 1998, McCorquodale 2001), but our sample of undetected groups was still relatively small.

However, our sightability-model estimates of population size were very close to our mark–resight estimates (Tables 1–3). This suggests that the effect of group size may have been enough to account for most of the changes in detectability between surveys. Simulation studies have shown that a simpler model, based only on group size, can outperform a more complex one, even when the more complex model includes the additional variables that are in the true sightability function (Cogan and Diefenbach 1998). The close agreement between our sightability-model and mark–resight estimates also suggests that any bias in the mark–resight estimates due to nonrandom distribution of marks may not have been too severe. Confidence intervals are smaller for the sightability model because each population estimate uses all the sightability data from all 4 years to estimate detection probability.

## Management Implications

The high cost of maintaining a marked population of sheep makes it unlikely that the mark–resight approach can be implemented as a long-term monitoring protocol for Dall's sheep populations in the Baird Mountains or most other locations. Of the other approaches that we considered, only the unadjusted counts and the sightability model were operationally feasible. The sightability model reasonably accounts for the effect of changes in group size on detectability. Mean group sizes did increase along with the population size, although the magnitude of this increase apparently was not enough to have much of an effect on overall detection rates during our study period. However, we expect that continued increases in mean group size would eventually result in increased overall detection probabilities, making this an important factor to take into account. Assuming that a marked population cannot be maintained, we recommend employing a sightability-model approach that accounts for group size rather than continued monitoring with unadjusted counts. However, it is important to recognize that a sightability model only accounts for variation within the range of characteristics (e.g., for aircraft, pilots, observers, and sheep) and conditions in which it was developed and that these may change over time. Thus, we also recommend continued work to refine the model and periodic verification that the model continues to provide reasonable estimates.

## Acknowledgments

We thank our survey pilots, S. Hamilton, G. Lee, J. Rood, and R. Swisher; telemetry pilots, J. Marten and M. Webb; and observers, J. Bryant, T. DeBruyn, W. Goodwin, J. Jackson, K. Joly, G. Roffler, M. Schnorr, and H. Weinstock. R. Swisher and B. Minor skillfully captured sheep for collaring. D. Spirtes and L. Dalle-Molle provided essential administrative and logistical support. We also appreciate helpful comments provided by K. Jenkins, W. Thompson, G. White, and 2 anonymous reviewers on earlier versions of this manuscript. Funding for this project was provided by the United States National Park Service and the United States Geological Survey.

## Literature Cited

Anderson, C. R., Jr., and F. G. Lindzey. 1996. Moose sightability model developed from helicopter surveys. Wildlife Society Bulletin 24:247–259.

Anderson, C. R., Jr., 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.

Andrew, N. G., V. C. Bleich, P. V. August, and S. G. Torres. 1997. Demography of mountain sheep in the east Chocolate Mountains, California. California Fish and Game 83:68–77.

Ayres, L. 1986. The movement patterns and foraging ecology of Dall sheep (*Ovis dalli dalli*) in the Noatak National Preserve, Alaska. Thesis, University of California, Berkeley, USA.

Ayres, L. A. 1996. Unit 23 & subunit 26A—western Brooks Range. Pages 129–147 *in* M. V. Hicks, editor. Dall sheep. Alaska Department of Fish and Game, Federal Aid in Wildlife Restoration Research Final Report, Juneau, USA.

Bartmann, R. M., G. C. White, L. H. Carpenter, and R. A. Garrott. 1987. Aerial mark–recapture estimates of confined mule deer in pinyon–juniper woodland. Journal of Wildlife Management 51:41–46.

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.

Bodie, W. L., and L. E. Oldenburg. 1995. A standardized technique for helicopter surveys of bighorn sheep. Biennial Symposium of the Northern Wild Sheep and Goat Council 9:65–68.

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

Carroll, G. 1993. Unit 23 & subunit 26A—western Brooks Range. Pages 124–138 *in* S. M. Abbott, editor. Dall sheep. Alaska Department of Fish and

Game, Federal Aid in Wildlife Restoration Research Final Report, Juneau, USA.

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

Chapman, D. G. 1951. Some properties of the hypergeometric distribution with applications to zoological censuses. University of California Publications in Statistics 1:131–160.

Cogan, R. D., and D. R. Diefenbach. 1998. Effect of undercounting and model selection on a sightability-adjustment estimator for elk. Journal of Wildlife Management 62:269–279.

Douglas, C. L., and K. Longshore. 1995. Costs and effectiveness of methods used for population estimates of desert bighorn sheep in Death Valley National Park. Desert Bighorn Council Transactions 39:1–9.

Gasaway, W. C., S. D. DuBois, D. J. Reed, and S. J. Harbo. 1986. Estimating moose population parameters from aerial surveys. University of Alaska Biological Papers 22, Fairbanks, USA.

Harrell, F. E., Jr. 2001. Regression modeling strategies. Springer-Verlag, New York, New York, USA.

Hosmer, D. W., and S. Lemeshow. 2000. Applied logistic regression. John Wiley & Sons, New York, New York, USA.

Jessup, D. A., W. E. Clark, and R. Mohr. 1988. The safety and cost-effectiveness of net-gun capture of desert bighorn sheep (*Ovis canadensis nelsoni*). Journal of Zoo and Wildlife Medicine 19:208–213.

McCorquodale, S. M. 2001. Sex-specific bias in helicopter surveys of elk: sightability and dispersion effects. Journal of Wildlife Management 65:216–225.

McDonald, L. L., H. B. Harvey, F. J. Mauer, and A. W. Brackney. 1990. Design of aerial surveys for Dall sheep in the Arctic National Wildlife Refuge, Alaska. Biennial Symposium of the Northern Wild Sheep and Goat Council 7:176–193.

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.

Otten, M. R. M., J. B. Haufler, S. R. Winterstein, and L. C. Bender. 1993. An aerial censusing procedure for elk in Michigan. Wildlife Society Bulletin 21: 73–80.

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.

Rexstad, E. A., D. D. Miller, C. H. Flather, E. M. Anderson, J. W. Hupp, and D. R. Anderson. 1988. Questionable multivariate statistical inference in wildlife habitat and community studies. Journal of Wildlife Management 52:794–798.

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

Seber, G. A. F. 1982. The estimation of animal abundance and related parameters. Second edition. Oxford University, New York, New York, USA.

Singer, F. J., D. Johnson, and R. Quimby. 1983. Dall sheep numbers and distribution in the Noatak National Preserve, 1983. Natural Resource Survey and Inventory Report, Alaska Region. U.S. National Park Service, Anchorage, Alaska, USA.

Strickland, D. L., L. L. McDonald, J. Kern, T. Spraker, and A. Loranger. 1994. Analysis of 1992 Dall's sheep and mountain goat survey data, Kenai National Wildlife Refuge. Biennial Symposium of the Northern Wild Sheep and Goat Council 9:35–42.

Strickland, D., L. L. McDonald, D. Taylor, K. Jenkins, and J. Kern. 1992. Estimation of Dall sheep numbers in the Wrangell–St. Elias National Park and Preserve. Biennial Symposium of the Northern Wild Sheep and Goat Council 8:231–255.

Thompson, W. L., G. C. White, and C. Gowan. 1998. Monitoring vertebrate populations. Academic, San Diego, California, USA.

Udevitz, M. S. 2002. Abundance: population size and density estimation. Pages 4–11 *in* A. H. El-Shaarawi and W. W. Piegorsch, editors. Encyclopedia of environmetrics. Volume 1. John Wiley & Sons, New York, New York, USA.

Whitten, K. R. 1997. Estimating population size and composition of Dall sheep in Alaska: assessment of previously used methods and experimental implementation of new techniques. Alaska Department of Fish and Game, Federal Aid in Wildlife Restoration Research Final Report, Juneau, USA.

***Mark S. Udevitz*** (left) is a research statistician with the United States Geologic Survey (USGS), Alaska Science Center in Anchorage, Alaska. He earned B.S. and M.S. degrees in Wildlife Biology and Management at Colorado State University and West Virginia University and a Ph.D. in Biomathematics and Statistics at North Carolina State University. His research focuses on methods for sampling, estimating demographic parameters, and modeling dynamics of wildlife populations. ***Brad S. Shults*** (not pictured) is a wildlife biologist and pilot for the United States National Park Service in Kotzebue, Alaska. He received his B.S. and M.S. in wildlife biology from the University of Alaska, Fairbanks. His primary interests are population dynamics and the management of harvested species in national park units within Alaska. ***Layne G. Adams*** (right) is a wildlife research biologist with the USGS Alaska Science Center. He received his B.S. and M.S. from Colorado State University and his Ph.D. from the University of Minnesota. He has conducted research in Alaska on population dynamics and predator–prey relationships of wolves and ungulates for more than 20 years. ***Christopher Kleckner*** (not pictured) is a former wildlife research biologist with the USGS Alaska Science Center. He received his B.A. from Boston University and his M.S. and M.A.T. from the University of Alaska, Anchorage. Currently he is an eighth-grade science teacher at Central Middle School of Science in Anchorage, Alaska.

*Associate Editor: White.*