ESTIMATING NUMBERS OF FEMALES WITH CUBS-OF-THE-YEAR IN THE YELLOWSTONE GRIZZLY BEAR POPULATION

KIM A. KEATING, U.S. Geological Survey, Northern Rocky Mountain Science Center, Montana State University, Bozeman, MT 59717, USA, email: kkeating@montana.edu

CHARLES C. SCHWARTZ, U.S. Geological Survey, Northern Rocky Mountain Science Center, Interagency Grizzly Bear Study Team, Montana State University, Bozeman, MT 59717, USA, email: chuck_schwartz@usgs.gov

MARK A. HAROLDSON, U.S. Geological Survey, Northern Rocky Mountain Science Center, Interagency Grizzly Bear Study Team, Montana State University, Bozeman, MT 59717, USA, email: mark_haroldson@usgs.gov

DAVID MOODY, Wyoming Game and Fish Department, 260 Buena Vista, Lander, WY 82520, USA, email: dmoody@missc.state.wy.us

Abstract: For grizzly bears (*Ursus arctos horribilis*) in the Greater Yellowstone Ecosystem (GYE), minimum population size and allowable numbers of human-caused mortalities have been calculated as a function of the number of unique females with cubs-of-the-year (F_{CUB}) seen during a 3-year period. This approach underestimates the total number of F_{CUB}, thereby biasing estimates of population size and sustainable mortality. Also, it does not permit calculation of valid confidence bounds. Many statistical methods can resolve or mitigate these problems, but there is no universal best method. Instead, relative performances of different methods can vary with population size, sample size, and degree of heterogeneity among sighting probabilities for individual animals. We compared 7 nonparametric estimators, using Monte Carlo techniques to assess performances over the range of sampling conditions deemed plausible for the Yellowstone population. Our goal was to estimate the number of F_{CUB} present in the population each year. Our evaluation differed from previous comparisons of such estimators by including sample coverage methods and by treating individual sightings, rather than sample periods, as the sample unit. Consequently, our conclusions also differ from earlier studies. Recommendations regarding estimators and necessary sample sizes are presented, together with estimates of annual numbers of F_{CUB} in the Yellowstone population with bootstrap confidence bounds.

Criteria for recovering the grizzly bear in the lower United States include annual limits on mortalities (U.S. Fish and Wildlife Service 1993). Since 1993, these limits have been calculated as a function of the number of F_{CUB} present in the population, as estimated during 6-year running periods. Currently, the number of F_{CUB} present each year (*N*) is estimated as the number of such animals actually observed ($\hat{N}_{Obs}$). To the extent that criteria for distinguishing family groups are conservative (see Knight et al. 1995), and because it is highly unlikely that all such animals are seen, $\hat{N}_{Obs}$ almost certainly underestimates *N*. This helps ensure that mortality limits are conservative, but precludes calculation of valid confidence bounds. Moreover, use of a biased estimator like $\hat{N}_{Obs}$ effectively removes decisions regarding the appropriate degree of conservatism from the purview of managers. This is not a trivial issue because the magnitudes of biases and uncertainties inherent in $\hat{N}_{Obs}$ may be biologically and managerially significant.

Efforts to calculate statistically sound estimates of *N* have focused on parametric approaches. Eberhardt and Knight (1996) applied the Peterson-type estimators of Chapman and Bailey (Seber 1982), and Boyce et al. (M.S. Boyce, D. MacKenzie, B.F.J. Manly, M.A. Haroldson, and D. Moody, 1999, Cumulative counts of unique individuals for estimating population size, U.S. Fish and Wildlife Service, Missoula, Montana, USA) recommended the maximum likelihood method of Lewontin and Prout (1956). These methods assume that each family group has an equal probability of being sighted. Because this assumption is untenable for the Yellowstone data (K.A. Keating, M.A. Haroldson, D. Moody, and C.C. Schwartz, 1999, Estimating the number of females with cubs-of-the-year in the Yellowstone grizzly bear population: are maximum-likelihood estimates that assume equal sightability conservative? U.S. Fish and Wildlife Service, Missoula, Montana, USA) estimates based on these methods will be negatively biased. Seeking a more robust approach, Boyce et al. (2001) recommended joint estimation of *N* over all years using an estimator derived from the zero-truncated negative binomial distribution. This estimator can be traced to Greenwood and Yule (1920), with early applications to wildlife population estimation by Tanton (1965, 1969) and Taylor (1966). The sampling model assumed by the negative binomial estimator allows for heterogeneous sighting probabilities among individuals and, thus, is equivalent to model M_h of Otis et al. (1978). Unfortunately, Boyce et al. (2001) found that the negative binomial estimator gave reasonable results only when the coefficient of variation among individual sighting probabilities (CV) was assumed to be constant over time. This assumption is difficult to justify for grizzly bears in Yellowstone, where year-to-year differences in distributions and abundances of foods affect bear movement patterns and, in turn, the likelihood of seeing particular bears (Picton et al. 1986). Such differences almost certainly affect heterogeneity among individual sighting probabilities, implying that CV varies among years. Also, because

the size, distribution, and behavior of bear populations may interact in ways that affect sightability (Keating 1986), CV likely changes with *N*. The claim of an increased bear population in Yellowstone (Boyce et al. 2001), therefore, is inconsistent with the assumption of a constant CV. The joint estimation procedure recommended by Boyce et al. (2001) suffers other drawbacks as well. Most seriously, estimates of *N* from previous years may change retrospectively as new data are added — a property that is justifiable only if CV is truly constant over time. Overall, problems with the parametric methods used to date argue for considering other alternatives.

Many nonparametric estimators might apply to this problem (e.g., Otis et al. 1978, Bunge and Fitzpatrick 1993, Lee and Chao 1994). Indeed, when estimating *N* under model M_h, many studies have favored non-parametric methods such as the jackknife (Burnham and Overton 1978, 1979), Chao (Chao 1984, 1989), and sample coverage estimators (Chao and Lee 1992, Lee and Chao 1994). Among the nonparametric methods available, however, there is no universal best choice, as relative performances can vary with *N*, CV, or sample size (Burnham and Overton 1979, Smith and van Belle 1984, Chao 1988). What we require is an estimator that is reasonably robust to variations in these parameters over the range of values experienced when sampling the Yellowstone grizzly bear population. To identify such an estimator, we used Monte Carlo methods to compare performances of 7 nonparametric methods when sampling from a range of conditions that encompassed those deemed plausible for observations of F_{CUB} in the GYE.

METHODS

General Problem and Notation

The sampling model we used approximates the true sampling scheme, in which reports of F_{CUB} come from observers using various sampling methods (ground-based observation, trapping, systematic fixed-wing observations, or fixed-wing observations made incidental to other work). Because the sampling period associated with each of these methods varies considerably (or, in some cases, is undefined) we used the sighting of an individual F_{CUB} as the sample unit. The problem of estimating population size from repeated sightings of unique individuals may then be phrased as a special case of the more general model in which multiple individuals may be sighted during a given sampling period (e.g., Otis et al. 1978).

Suppose that, during a given year, after recording *n* independent random sightings of individuals from a closed population of size *N* (where *N* is unknown), we observe *m* unique animals. The average probability that any particular sighting will be of the *i*th individual is p_i, and probabilities for all *N* individuals are given by $\mathbf{p} = (p_1, p_2, ..., p_N)$ where

$$\sum_{i=1}^{N} p_i = 1$$

Because the model allows for heterogeneous p_i values, temporal or spatial differences in habitat use or sampling effort are incorporated into **p**, as are differences in probabilities of reporting and recording sightings of particular animals. We assume all individuals are correctly identified (consequences of misidentification are considered below). In our sample, individuals were observed with frequency $\mathbf{n} = (n_1, n_2, ..., n_N)$, which is multinomially distributed with cell probabilities $(p_1, p_2, ..., p_N)$. However, we do not know the identities of the *N - m* animals for which $n_i = 0$. The number of different individuals observed exactly *j* times was f_j, and $\mathbf{f} = (f_0, f_1, f_2, ..., f_n)$ is fully observable except for f_0, the number of bears not observed in our sample. Important relationships include

$$n = \sum_{i=1}^{N} n_i = \sum_{j=1}^{N} j f_j$$
$$m = \sum_{j=1}^{N} f_j$$

and $N - m = f_0$. The problem is to estimate *N* (or, equivalently, f_0) using only the observable information in **f** and *n*.

In this idealized model, all information about population size is obtained from the *n* randomly sighted individuals. For the Yellowstone grizzly bear population, observations of radiomarked F_{CUB} made during radiorelocation flights provide additional information from non-randomly sighted individuals. In particular, observations of otherwise unobserved F_{CUB} may be added to *m* to improve the estimate of minimum population size, yielding $\hat{N}_{Obs} \geq m$. $\hat{N}_{Obs}$ provides a natural lower bound for estimating *N* and is the estimator that has been used previously to set annual mortality limits. Overall, we seek an estimator that improves upon $\hat{N}_{Obs}$ while minimizing the risk of overestimating *N*.

The Estimators

In addition to *m* and $\hat{N}_{Obs}$, which we included in our analyses for comparative purposes, we evaluated 7 non-parametric estimators (see Table 1 for example calculations). The first 5 methods we considered estimate *N* as $\hat{N} = m + \hat{f_0}$, where $\hat{f_0}$ is an estimate of the number of unobserved individuals.

We first examined Chao's (1984) estimator,

$$\hat{N}_{Chao1} = m + \frac{f_1^2}{2f_2} \tag{1}$$

In Eq. (1), $\hat{f_0} = f_1^2/(2f_2)$. Using $\hat{N}_{Chao1}$, the statistical expec-

Table 1. Example calculations for the 7 non-parametric estimators compared in this study, using 1997 grizzly bear sighting data from the Greater Yellowstone Ecosystem. For 1997, $n = 65$ sightings of females with cubs-of-the-year (F_{CUB}) were made via means other than radiotelemetry. Distinguishing individuals as per Knight et al. (1995), $m = 29$ unique animals were seen; 13 were seen once ($f_1 = 13$), 7 were seen twice ($f_2 = 7$), 4 were seen 3 times ($f_3 = 4$), 1 was seen 4 times ($f_4 = 1$), 3 were seen 5 times ($f_5 = 3$), and 1 was seen 7 times ($f_7 = 1$). Two additional and otherwise unobserved F_{CUB} were seen only as a result of using radiotelemetry. Because all calculations were carried out in double precision, rounding errors are evident in some of the examples.

Estimator	Example calculation
Unique F_{CUB} observed via random sightings	$m = 29$
Unique F_{CUB} observed via random sightings and radiotelemetry	$\hat{N}_{Obs} = 29 + 2 = 31$
Chao	$\hat{N}_{Chao1} = m + \dfrac{f_1^2}{2f_2} = 29 + \dfrac{13^2}{2(7)} \approx 41.1$
Bias-corrected Chao	$\hat{N}_{Chao2} = m + \dfrac{f_1^2 - f_1}{2(f_2 + 1)} = 29 + \dfrac{13^2 - 13}{2(7+1)} \approx 38.8$
First-order jackknife	$\hat{N}_{J1} = m + \left(\dfrac{n-1}{n}\right)f_1 = 29 + \left(\dfrac{65-1}{65}\right)13 = 41.8$
Second-order jackknife	$\hat{N}_{J2} = m + \left(\dfrac{2n-3}{n}\right)f_1 - \left(\dfrac{(n-2)^2}{n(n-1)}\right)f_2 = 29 + \left(\dfrac{2(65)-3}{65}\right)13 - \left(\dfrac{(65-2)^2}{65(65-1)}\right)7 \approx 47.7$
Best-order jackknife	$\hat{N}_{Jk1} = \hat{N}_{J1} = 41.8$ was selected because $T_1 = \dfrac{\hat{N}_{J2} - \hat{N}_{J1}}{[\hat{var}(\hat{N}_{J2} - \hat{N}_{J1} \mid m)]^{1/2}} \approx \dfrac{47.7 - 41.8}{[17.996]^{1/2}} \approx 1.396 < 1.960,$ where $\hat{var}(\hat{N}_{J2} - \hat{N}_{J1} \mid m) = \dfrac{m}{m-1}\left[\displaystyle\sum_{j=1}^{2}(b_j)^2 f_j - \dfrac{(\hat{N}_{J2} - \hat{N}_{J1})^2}{m}\right]$ $\approx \dfrac{29}{29-1}\left[\left(\dfrac{2(65)-3}{65} - \dfrac{65-1}{65}\right)^2 13 + \left(\dfrac{(65-2)^2}{65(65-1)}\right)^2 7 - \dfrac{(47.7-41.8)^2}{29}\right] \approx 17.996$
First-order sample coverage	$\hat{N}_{SC1} = \dfrac{m + f_1\hat{\gamma}^2}{\hat{C}_1} = \dfrac{29 + 13(0.325)}{0.800} \approx 41.5,$ where $\hat{C}_1 = 1 - \dfrac{f_1}{n} = 1 - \dfrac{13}{65} = 0.800$ and $\hat{\gamma}^2 = \max\left\{\dfrac{m}{\hat{C}_1}\displaystyle\sum_{j=1}^{n}\dfrac{j(j-1)f_j}{n(n-1)} - 1, \, 0\right\} = \max\left\{\dfrac{29}{0.800}\left(\dfrac{2(7)+6(4)+12(1)+20(3)+42(1)}{65(65-1)}\right) - 1, \, 0\right\} \approx 0.325$
Second-order sample coverage	$\hat{N}_{SC2} = \dfrac{m + f_1\hat{\gamma}^2}{\hat{C}_2} = \dfrac{29 + 13(0.319)}{0.803} \approx 41.3,$ where $\hat{C}_2 = 1 - \dfrac{f_1 - 2f_2/(n-1)}{n} = 1 - \dfrac{13 - 2(7)/(65-1)}{65} \approx 0.803$ and $\hat{\gamma}^2 = \max\left\{\dfrac{m}{\hat{C}_2}\displaystyle\sum_{j=1}^{n}\dfrac{j(j-1)f_j}{n(n-1)} - 1, \, 0\right\} = \max\left\{\dfrac{29}{0.803}\left(\dfrac{2(7)+6(4)+12(1)+20(3)+42(1)}{65(65-1)}\right) - 1, \, 0\right\} \approx 0.319$

tation for the estimate, $E(\hat{N})$, equals N only when sighting probabilities are the same for all animals; i.e., when CV= 0. Theoretically, when CV > 0, $E(\hat{N}) < N$ (Chao 1984). This does not ensure $\hat{N}_{Chao1} \le N$ in all cases, but does suggest that $\hat{N}_{Chao1}$ might provide an inherently conservative approach to estimating N. We also considered a similar bias-corrected form of this estimator, developed by Chao (1989). Where the sample unit is the individual animal, Chao's (1989) estimator is given by (Wilson and Collins 1992),

$$\hat{N}_{Chao2} = m + \frac{f_1^2 - f_1}{2(f_2 + 1)}$$

Here, $\hat{f}_0 = (f_1^2 - f_1)/[2(f_2 + 1)]$. Unlike $\hat{N}_{Chao1}$, $\hat{N}_{Chao2}$ will yield an estimate even when $f_2 = 0$.

Burnham and Overton (1978, 1979) devised a jackknife estimator ($\hat{N}_{Jk}$) of the general form

$$\hat{N}_{Jk} = m + \sum_{j=1}^{k} \alpha_{jk} f_j$$

where α_{jk} is a coefficient in terms of n, and $\alpha_{jk} = 0$ when $j > k$ (see Table 2). Here, f_0 is estimated as the series

$$\sum_{j=1}^{k} \alpha_{jk} f_j$$

Theoretically, jackknife estimates of order $k = 1$ to n could

be calculated, but variance increases rapidly with k so that, in practice, k is small (Burnham and Overton 1979). We considered the first- and second-order jackknife estimators ($\hat{N}_{J1}$ and $\hat{N}_{J2}$, respectively; Table 2), as well as a best kth-order jackknife estimator. Burnham and Overton (1979) suggested 2 methods for choosing a best value for k for a particular study. Because previous work showed little difference between them (K.A. Keating unpublished data), we considered only their first method, which evaluates estimates of order $k = 1$ to 5 (Table 2). The method is as follows. Beginning with $k = 1$ and proceeding to subsequently higher values of k, test the null hypothesis that $E(\hat{N}_{J, k+1} - \hat{N}_{Jk}) = 0$ versus the alternative hypothesis that $E(\hat{N}_{J, k+1} - \hat{N}_{Jk}) \ne 0$. If the observed difference is not significant, testing ends and $\hat{N}_{Jk}$ is taken as the best jackknife estimate. We reference the resulting kth-order estimate as $\hat{N}_{Jk1}$. The test is based on the statistic

$$T_k = \frac{\hat{N}_{J,k+1} - \hat{N}_{Jk}}{[\hat{var}(\hat{N}_{J,k+1} - \hat{N}_{Jk} \mid m)]^{1/2}}$$

where

$$\hat{var}(\hat{N}_{J,k+1} - \hat{N}_{Jk} \mid m) = \frac{m}{m-1}\left[\sum_{j=1}^{n}(b_j)^2 f_j - \frac{(\hat{N}_{J,k+1} - \hat{N}_{Jk})^2}{m}\right]$$

Table 2. Jackknife estimators of population size, $\hat{N}_{Jk}$, for order k = 1-5, where m is the number of unique individuals observed after n samples and f_i is the number of individuals observed exactly i times (after Burnham and Overton 1979).

$$\hat{N}_{J1} = m + \left(\frac{n-1}{n}\right)f_1$$

$$\hat{N}_{J2} = m + \left(\frac{2n-3}{n}\right)f_1 - \left(\frac{(n-2)^2}{n(n-1)}\right)f_2$$

$$\hat{N}_{J3} = m + \left(\frac{3n-6}{n}\right)f_1 - \left(\frac{3n^2 - 15n + 19}{n(n-1)}\right)f_2 + \left(\frac{(n-3)^3}{n(n-1)(n-2)}\right)f_3$$

$$\hat{N}_{J4} = m + \left(\frac{4n-10}{n}\right)f_1 - \left(\frac{6n^2 - 36n + 55}{n(n-1)}\right)f_2 + \left(\frac{4n^3 - 42n^2 + 148n - 175}{n(n-1)(n-2)}\right)f_3$$

$$- \left(\frac{(n-4)^4}{n(n-1)(n-2)(n-3)}\right)f_4$$

$$\hat{N}_{J5} = m + \left(\frac{5n-15}{n}\right)f_1 - \left(\frac{10n^2 - 70n + 125}{n(n-1)}\right)f_2 + \left(\frac{10n^3 - 120n^2 + 485n - 660}{n(n-1)(n-2)}\right)f_3$$

$$- \left(\frac{(n-4)^5 - (n-5)^5}{n(n-1)(n-2)(n-3)}\right)f_4 + \left(\frac{(n-5)^5}{n(n-1)(n-2)(n-3)(n-4)}\right)f_5$$

and $b_j = \alpha_{j,k+1} - \alpha_{jk}$. T_k was evaluated at $\alpha = 0.05$ using P values determined from the standard normal distribution.

Chao and Lee (1992) proposed an estimator based on sample coverage (C), where C is the sum of the p_i values for the m individuals actually observed in the sample. Lee and Chao (1994) offered 2 estimators of C that, in the notation of our sampling model, are given by

$$\hat{C}_1 = 1 - \frac{f_1}{n} \qquad (2)$$

and

$$\hat{C}_2 = 1 - \frac{f_1 - 2f_2/(n-1)}{n} \qquad (3)$$

In Eqs. (2) and (3), the quantities f_1/n and $[f_1 - 2f_2/(n-1)]/n$, respectively, estimate the sum of the p_i values for the f_0 unobserved animals. For our model (equivalent to model M_h of Otis et al. [1978]), Lee and Chao (1994) then estimated N as

$$\hat{N}_{SCj} = \frac{m}{\hat{C}_j} + \frac{f_1}{\hat{C}_j} \hat{\gamma}^2$$
$$= \frac{m + f_1 \hat{\gamma}^2}{\hat{C}_j} \qquad (4)$$

where $j = 1$ or 2, and γ is a measure of the coefficient of variation of the p_i's . Essentially, Eq. (4) begins with a Peterson-type estimator ($m/\hat{C}$) that assumes equal sightability (i.e., all $p_i = 1/N$; Darroch and Ratcliff 1980), then adds a bias correction term ($f_1 \hat{\gamma}^2/\hat{C}_j$) that increases with heterogeneity, as estimated by $\hat{\gamma}^2$. Put another way, the quantity $f_1 \hat{\gamma}^2$ estimates the number of additional individuals that would have been observed if **p** had, in fact, been homogeneous. Adding this to m then dividing by the estimated coverage estimates N. Where the sample unit is the sighting of an individual animal, $\hat{\gamma}^2$ is calculated as (Chao and Lee 1992),

$$\hat{\gamma}^2 = \max\left\{ \hat{N} \sum_{j=1}^{n} \frac{j(j-1)f_j}{n(n-1)} - 1, \ 0 \right\} \qquad (5)$$

Calculation of $\hat{\gamma}^2$ requires an initial estimate of N. Following Chao and Lee (1992), we used $\hat{N} = m/\hat{C}_j$. We considered but did not use the partitioned sample coverage estimator of Chao et al. (1993, 2000) because pre-

liminary Monte Carlo results showed the method offered no advantage over $\hat{N}_{SCj}$ when applied to our field data.

Monte Carlo Comparisons

Estimator performances were compared using Monte Carlo methods. Parameters for the Monte Carlo sampling were chosen to encompass the range of values deemed plausible when sampling F_{CUB} in the GYE. Overall, we simulated 15 populations, including all combinations of $N = 20$, 40, and 60 animals, where the coefficient of variation among the p_i values was set to CV = 0.0, 0.25, 0.50, 0.75, or 1.0. We calculated p_i as the integral of a standard beta distribution over the interval $(i - 1)/N$ to i/N; i.e.,

$$p_i = I_{i/N}(a,b) - I_{(i-1)/N}(a,b), \qquad (6)$$

where $I_x(a,b)$ is the incomplete beta function ratio with parameters a and b (Johnson et al. 1995). We used a downhill simplex (Press et al. 1992) to select values for a and b (Table 3) that gave the desired CV among the p_i values. We then sampled each population, with replacement, by generating n pseudorandom numbers from the specified beta distribution and tallying each as a sighting of the ith animal if it fell within the interval $(i - 1)/N$ to i/N. We chose n so that the number of sightings per individual in the population (n/N) was equal to 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, or 4.0. After each sampling bout, we estimated N using each of the estimators described above. This process was repeated 1,000 times for each parameterization of the model. For each parameterization and estimator, performance was summarized as the bias and root mean square error (RMSE) of the estimator, where

$$\text{RMSE} = \sqrt{\text{bias}^2 + \text{SD}^2}$$

In addition, 2 estimators ($\hat{N}_{SC1}$ and $\hat{N}_{SC2}$) yielded explicit estimates of CV, in the form of $\hat{\gamma}$ (Eq. 5).

Following the above analyses, the most promising estimator was selected. Confidence bounds for estimates based on the best method were calculated using the method of Boyce et al. (2001), in which bootstrap samples were drawn from the distribution of individual sighting frequencies implied by $\hat{N}$ (i.e., from the estimate of the vector **n**). Details are as follows. A model population with $\hat{N}$ indi-

Table 3. Values of the parameters (a, b) of the standard beta distributions used to model $\mathbf{p} = (p_1, p_2, ...p_N)$, where p_i is the probability that a particular sighting will be of the ith animal. Values are listed by size (N) of the model population and the coefficient of variation (CV) among the p_i values.

N	CV = 0.00	CV = 0.25	CV = 0.50	CV = 0.75	CV = 1.00
			(a, b)		
20	(1.000, 1.000)	(0.955, 1.270)	(0.791, 1.380)	(0.664, 1.446)	(0.589, 1.600)
40	(1.000, 1.000)	(1.084, 1.398)	(0.797, 1.382)	(0.686, 1.477)	(0.593, 1.512)
60	(1.000, 1.000)	(1.173, 1.449)	(0.794, 1.369)	(0.688, 1.462)	(0.611, 1.559)

viduals was constructed and the first m individuals were assigned sighting frequencies $\mathbf{n}^* = (n_1^*, n_2^*, ..., n_m^*)$, corresponding to the actual sighting frequencies (n_i values) for the m animals observed in the original sample. The remaining $\hat{N}$ - m individuals were assigned sighting frequencies of 0. A bootstrap sample of $\hat{N}$ (rounded to the nearest integer) individual sighting frequencies (n_i^* values) was then randomly drawn with replacement from $\mathbf{n}^*$. The number of samples for which $n_i^* = j$ was tabulated as f_j^*, giving the bootstrap sighting frequency vector $\mathbf{f} = (f_1^*, f_2^*, ..., f_n^*)$, and the bootstrap number of sightings

$$n^* = \sum_{j=1}^{n} j f_j^*$$

The estimate was then recalculated using the information in $\mathbf{f}^*$ and n^*. This procedure was repeated 1,000 times for each estimate. Confidence bounds were calculated using both the percentile and bias-corrected-and-accelerated (BCA) methods (Efron and Tibshirani 1993). We assessed performances of the 2 methods by comparing observed versus nominal coverages.

Although 90 or 95% confidence bounds are normal for scientific hypothesis testing, managers may appropriately choose a higher level of risk. Thus, we compared coverages for lower, 1-tailed 70, 80, 90, and 95% confidence bounds. Earlier studies reported 2-tailed confidence bounds (e.g., Eberhardt and Knight 1996, Boyce et al. 2001). However, we believe 2-tailed bounds are inappropriate for this problem because managers charged with recovering the Yellowstone grizzly bear population are concerned with possible overharvest, not underharvest. Thus, they seek assurance that the true population size is greater than or equal to the estimated size. It follows that lower, 1-tailed confidence bounds provide the appropriate measure of uncertainty.

Field Data

Sightings of F_{CUB} were examined for 1986–2001. We considered only sightings from within the grizzly bear recovery zone and the surrounding 10-mile buffer area because calculated mortality limits only apply to human-caused mortalities within this area. Boyce et al. (2001) considered sightings throughout the GYE. Consequently, sample sizes (n values) and numbers of unique, randomly observed F_{CUB} (m values) reported herein differ slightly from values reported by Boyce et al. (2001).

For each year, unique family groups were distinguished as per Knight et al. (1995). Observations of radiocollared animals made during radiolocation flights were included when calculating the minimum number of F_{CUB} known to exist in the population each year ($\hat{N}_{\text{Obs}}$), but were excluded from statistical estimates of N because such sightings were non-random. Sightings were summarized by year as the

number of unique family groups seen once, twice, etc. Total numbers of F_{CUB} for each year were then estimated using the method selected following our Monte Carlo comparisons. Lower, 1-tailed confidence bounds were calculated using the selected bootstrap procedure.

RESULTS

Monte Carlo Comparisons

Patterns of estimator performance varied little with population size. For brevity, therefore, we discuss only results for model populations with $N = 40$ individuals.

Population Estimates.—All estimates tended to converge toward N as relative sample size (n/N) increased, but rate of convergence and direction of bias at small to moderate sample sizes varied considerably among estimators and with CV (Fig. 1). Contrary to expectations, Chao's (1984) estimator, $\hat{N}_{\text{Chao1}}$, was postively biased when CV was small. This bias was especially pronounced when n/N also was small. However, $\hat{N}_{\text{Chao1}}$ was among the least biased estimators when CV was large, regardless of sample size. As predicted by theory (Chao 1989), $\hat{N}_{\text{Chao1}}$ was nearly unbiased when CV = 0, but became increasingly and negatively biased as CV increased. The jackknife estimators ($\hat{N}_{\text{J1}}$, $\hat{N}_{\text{J2}}$, and $\hat{N}_{\text{Jk1}}$) were all negatively biased when $n/N < 1.0$, but tended to overestimate N at sample sizes where $1.0 < n/N \leq 3.0$, particularly when CV was small. The jackknife estimators also did not converge toward N as quickly as other estimators as sample size increased. Patterns for the 2 sample coverage estimators were similar: both tended to overestimate N when n/N and CV were small, but converged relatively quickly toward N as n/N exceeded 1.0, particularly when $0.25 \leq$ CV ≤ 0.75.

With some methods, it was not always possible to estimate N. Over the full range of conditions modeled, $\hat{N}_{\text{Chao1}}$, $\hat{N}_{\text{Jk1}}$, $\hat{N}_{\text{SC1}}$, and $\hat{N}_{\text{SC2}}$ failed to yield estimates in 0.2% of the cases (range = 0.0–29.0% for $\hat{N}_{\text{Chao1}}$; range = 0.0–6.6% for $\hat{N}_{\text{Jk1}}$, $\hat{N}_{\text{SC1}}$, and $\hat{N}_{\text{SC2}}$). Reasons for failures varied. For $\hat{N}_{\text{Chao1}}$, no estimate is possible when $f_2 = 0$ because this leads to division by zero (Eq. 1). For $\hat{N}_{\text{Jk1}}$, the selection process was aborted if a best jackknife estimate was not selected from the estimates $\hat{N}_{\text{J1}}$- $\hat{N}_{\text{J5}}$. Using $\hat{N}_{\text{Jk1}}$, Burnham and Overton (1979) similarly failed to identify a best estimate in 3.7% of their trials. For $\hat{N}_{\text{SC1}}$ and $\hat{N}_{\text{SC2}}$, no population estimate is possible if the estimated sample coverage is zero, as this also leads to division by zero (Eq. 4). This occurs when individuals in the sample are seen only once each, so that $f_1 = n$ and $f_2 = 0$ (Eqs. 2 and 3). For all of these methods, failure rates declined as sample size and, hence, information content increased.
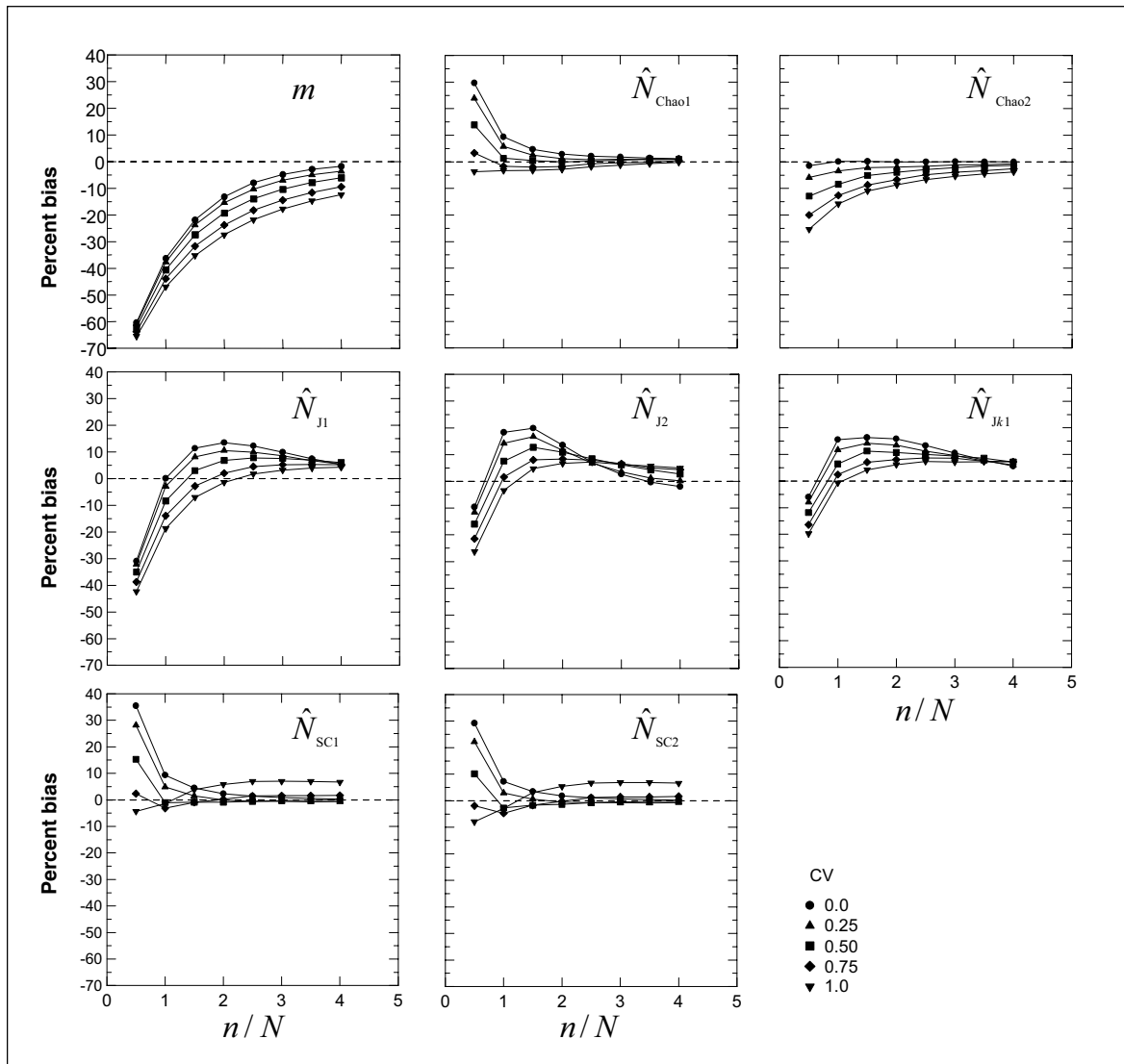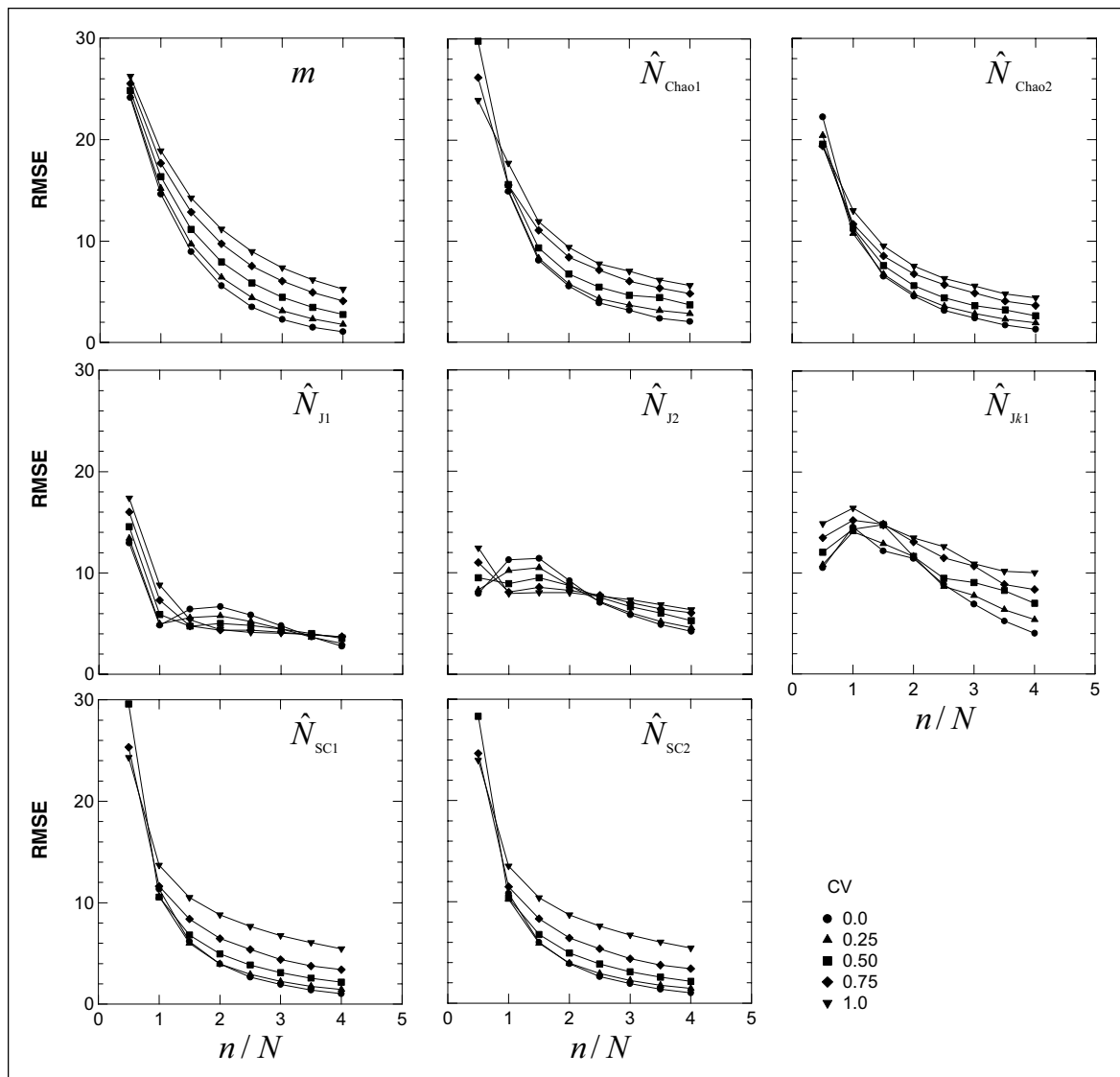
Fig. 1. Percent bias of population estimates calculated using the Chao ($\hat{N}_{\text{Chao1}}$), bias-corrected Chao($\hat{N}_{\text{Chao2}}$), first-order jackknife ($\hat{N}_{\text{J1}}$), second-order jackknife ($\hat{N}_{\text{J2}}$), best-order jackknife ($\hat{N}_{\text{Jk1}}$), first-order sample coverage ($\hat{N}_{\text{SC1}}$), and second-order sample coverage ($\hat{N}_{\text{SC2}}$) estimators. Number of unique individuals observed (*m*) is shown for comparison. Each point represents the mean of 1,000 Monte Carlo replicates; in each, calculations were based on *n* random sightings drawn from a model population with *N* = 40 individuals. CV gives the coefficient of variation among sighting probabilities for the 40 individuals. CV = 0.0 indicates equal sightability.

For $\hat{N}_{\text{Chao1}}$, $\hat{N}_{\text{Chao2}}$, $\hat{N}_{\text{SC1}}$, and $\hat{N}_{\text{SC2}}$, RMSE declined monotonically toward zero as *n/N* increased (Fig. 2). Patterns of decline were indistinguishable for $\hat{N}_{\text{SC1}}$ and $\hat{N}_{\text{SC2}}$, and RMSE converged more quickly toward zero for these estimators than for $\hat{N}_{\text{Chao1}}$ or $\hat{N}_{\text{Chao1}}$. Also for these 4 estimators, RMSE increased with CV when *n/N* ≥ 1. When *n/N* was small, $\hat{N}_{\text{J1}}$, $\hat{N}_{\text{J2}}$, and $\hat{N}_{\text{Jk1}}$ exhibited the lowest RMSEs of the estimators we evaluated. However, rate of convergence toward zero as sample size increased was slow compared to other methods; indeed, RMSE for the jackknife estimators often increased with sample size when 0.5 ≤ *n/N* ≤ 2.0. Also, relatively low RMSEs, especially

for $\hat{N}_{\text{J1}}$, often were due to low standard deviations overcompensating for high bias. This suggested that $\hat{N}_{\text{J1}}$ may yield narrow confidence bounds, but that those bounds will be centered around highly biased estimates, likely resulting in poor coverage.

Of the methods we compared, our overall choice was the second-order sample coverage estimator, $\hat{N}_{\text{SC2}}$ (see Discussion). Comparing observed versus nominal lower, 1-tailed confidence bounds for $\hat{N}_{\text{SC2}}$ showed that coverage was affected by *n/N* and CV, and by the method used to calculate confidence bounds (Figs. 3 and 4). Disparities between observed and nominal coverages generally

Fig. 2. Root mean square error (RMSE) of population estimates calculated using the Chao ($\hat{N}_{Chao1}$), bias-corrected Chao ($\hat{N}_{Chao2}$), first-order jackknife ($\hat{N}_{J1}$), second-order jackknife ($\hat{N}_{J2}$), best-order jackknife ($\hat{N}_{Jk1}$), first-order sample coverage ($\hat{N}_{SC1}$), and second-order sample coverage ($\hat{N}_{SC2}$) estimators. Number of unique individuals observed (m) is shown for comparison. Each data point represents the mean of 1,000 Monte Carlo replicates; in each, calculations were based on n random sightings drawn from a model population with $N = 40$ individuals. CV gives the coefficient of variation among sighting probabilities for the 40 individuals. CV = 0.0 indicates equal sightability.

increased with CV, but declined as the nominal confidence level increased. Results varied most noticeably with n/N when CV ≥ 0.75. Using the percentile bootstrap method, nominal values sometimes overstated the true coverage when CV = 0.0, but tended to either closely approximate or understate true coverage when 0.25 ≤ CV ≤ 1.0 (Fig. 3). Using the BCA bootstrap method, nominal values more closely approximated observed coverages when CV = 0.0, and tended to either approximate or understate true coverage when 0.25 ≤ CV ≤ 0.75. For CV = 1.0, however, nominal values tended to overstate true coverage by a large margin when n/N ≥ 2.0. Overall, we chose the

percentile bootstrap method for calculating confidence bounds because, with CV = 0.0 unlikely in natural populations, we believe that it better minimizes the risk of overestimating N.

Estimates of n/N *and CV.*—In our Monte Carlo study, n/N and CV were important determinants of performance for our estimator of choice, $\hat{N}_{SC2}$. Estimates of these values are given by $n/\hat{N}_{SC2}$ and $\hat{\gamma}$ (Eq. 5), respectively. Presumably, such estimates might be used to ask whether actual values of n/N and CV in our field studies were within the range of values in which $\hat{N}_{SC2}$ performed well. First, however, it is prudent to ask whether $n/\hat{N}_{SC2}$ and $\hat{\gamma}$ themselves provide
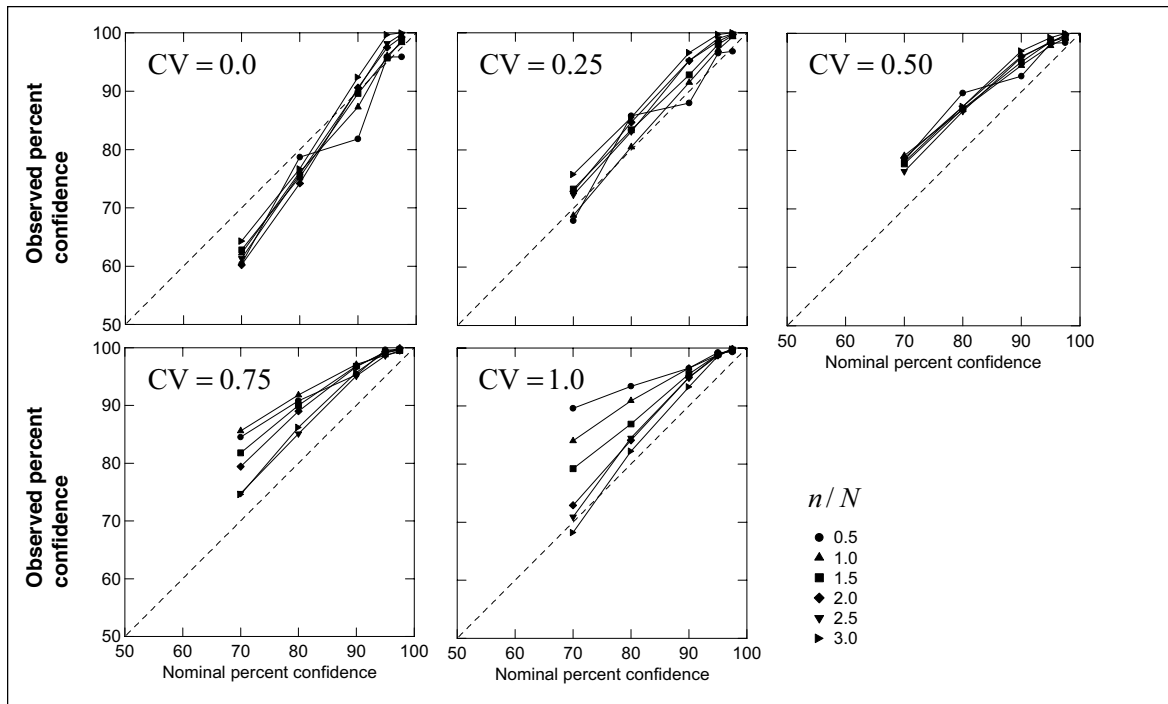
Fig. 3. Observed versus nominal coverages of lower, 1-tailed confidence bounds for second-order sample coverage estimates (N_{sc2}), calculated using the percentile bootstrap method (Efron and Tibshirani 1993). Points above the dashed line indicate that mean observed coverage was greater than nominal coverage, so confidence bounds tended to be conservative. Each data point represents the mean of 1,000 Monte Carlo replicates; in each, calculations were based on *n* random sightings drawn from a model population with *N* = 40 individuals. CV gives the coefficient of variation among sighting probabilities for the 40 individuals. CV = 0.0 indicates equal sightability.
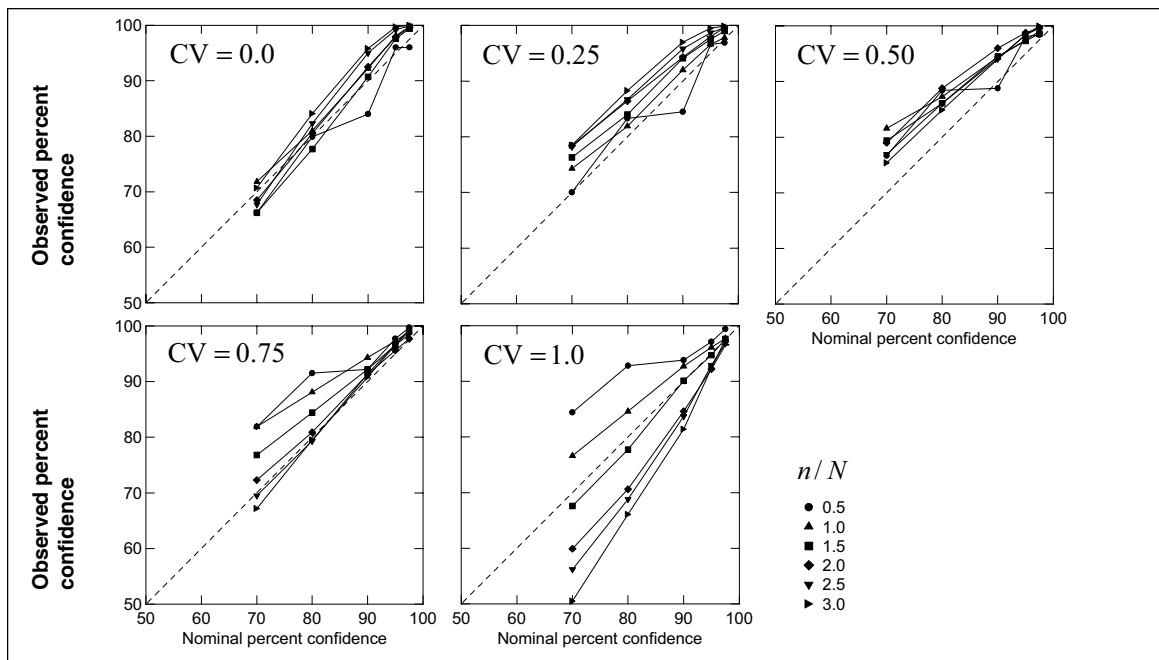
Fig. 4. Observed versus nominal coverages of lower, 1-tailed confidence bounds for second-order sample coverage estimates (N_{sc2}), calculated using the bias corrected and accelerated bootstrap method (Efron and Tibshirani 1993). Points above the dashed line indicate that mean observed coverage was greater than nominal coverage, so that confidence bounds tended to be conservative. Each data point represents the mean of 1,000 Monte Carlo replicates; in each, calculations were based on *n* random sightings drawn from a model population with *N* = 40 individuals. CV gives the coefficient of variation among sighting probabilities for the 40 individuals. CV = 0.0 indicates equal sightability.

good estimates. Comparisons showed that $n/\hat{N}_{SC2}$ provided nearly unbiased estimates of n/N throughout the range of conditions we modeled (Fig. 5a). However, $\hat{\gamma}$ was a biased estimator of CV, overestimating the true value when CV = 0.0 and underestimating in all other cases (Fig. 5b). The degree to which $\hat{\gamma}$ underestimated CV when CV ≥ 0.25 was influenced by relative sample size. When $n/N = 3.0$, $\hat{\gamma}$ tended to underestimate CV by about 0.07–0.14. When $n/N = 0.5$, $\hat{\gamma}$ tended to underestimate CV by about 0.10-0.59.

Field Data

Observation frequencies for F_{CUB} in Yellowstone's grizzly bear recovery area and the surrounding 10-mile buffer zone were tabulated for 1986–2001 (Table 4). Sample sizes ranged from 20 observations in 1987 to 94 in 1999. Using $\hat{N}_{SC2}$ and rounding to the nearest integer, estimated numbers of F_{CUB} in the Yellowstone population ranged from 20 animals in 1987 and 1989 to 60 in 2000 (Table 5). Estimated relative sample size $(n/\hat{N}_{SC2})$ averaged 1.5 and ranged from 0.5 in 1995 to 2.6 in 1986 and 1999, with $n/\hat{N}_{SC2} ≥ 1.0$ for 14 of the 16 years examined (Table 5). The estimated coefficient of variation among individual sighting probabilities $(\hat{\gamma})$ averaged 0.46 and ranged from 0.0 in 1990, 1993, and 1994 to 0.90 in 2000 (Table 5).

The total number of unique F_{CUB} actually observed $(\hat{N}_{Obs})$ ranged from 13 in 1987 to 42 in 2001 (Table 5). This included animals that would not have been detected without radiotelemetry. The number of unique F_{CUB} detected through random sightings alone (m) ranged from 12 in 1987 to 39 in 2001 (Table 5). On average, additional information provided by radiotelemetry increased

the number of unique F_{CUB} observed by 2.1 animals/year (range = 0–5 animals). For every year, $\hat{N}_{SC2}$ exceeded $\hat{N}_{Obs}$ (Table 5). However, when rounded to the nearest integer, the lower, 1-tailed 95 and 90% confidence bounds for $\hat{N}_{SC2}$ were less than $\hat{N}_{Obs}$ for 10 and 5 of the years, respectively (Table 5). Lower, 1-tailed 70 and 80% confidence bounds were $≥ \hat{N}_{Obs}$ for all years except 1990 (Table 5).

DISCUSSION

Whether Yellowstone's grizzly bears are removed from the threatened species list depends, in part, on whether human-caused mortalities are within calculated limits. Because mortality limits are computed as a function of the number of F_{CUB} present in the population, statistically sound estimates of annual numbers of F_{CUB} (N) are needed. Parametric methods proposed by Eberhardt and Knight (1996) and Boyce et al. (2001; unpublished report,1999) improved on the practice of basing mortality limits on a minimum estimate for N, determined as the number of unique F_{CUB} observed in a given year $(\hat{N}_{Obs})$. However, these methods require untenable assumptions about the form and constancy of distributions of individual sighting probabilities. At best, these assumptions leave unnecessary room for dispute, potentially undermining the credibility of results and diverting attention from other important issues. At worst, they can cause serious biases.

Nonparametric approaches are free of assumptions about distributions of sighting probabilities, but have not previously been applied to this problem. Nor should they be applied uncritically, as both absolute and relative performances of different estimators can vary with sampling conditions. In this study, we sought a nonparametric
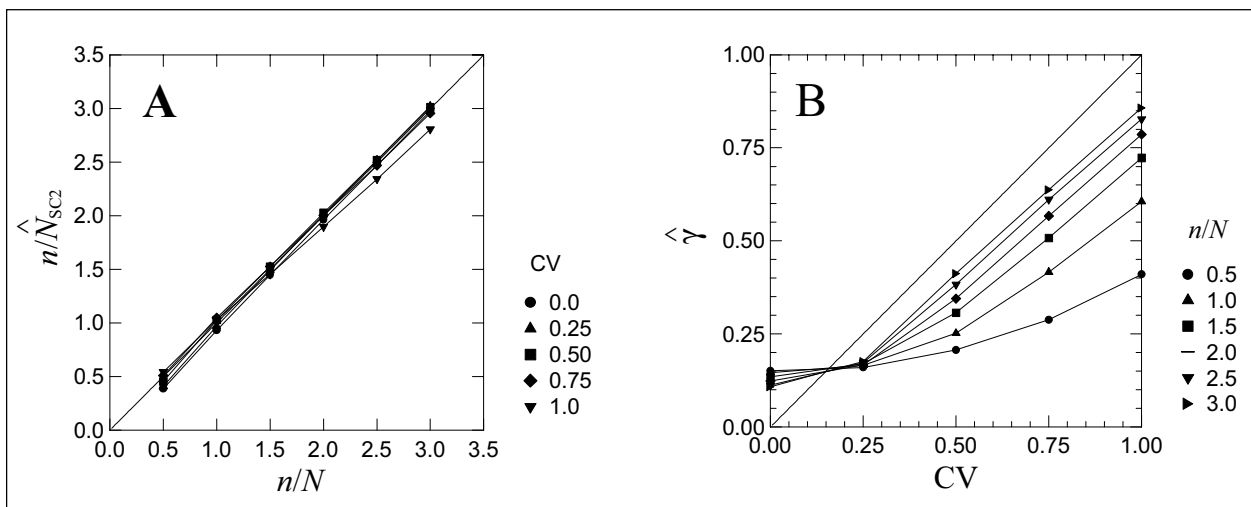
Fig. 5. Estimated ($n/\hat{N}_{SC2}$) versus observed (n/N) relative sample sizes (A), and estimated ($\hat{\gamma}$) versus observed (CV) values for the coefficient of variation among individual sighting probabilities (B). In both (A) and (B), each point represents the mean value, based on 1,000 Monte Carlo replicates; in each, calculations were based on n random sightings drawn from a model population with N = 40 individuals.

Table 4. Observation frequency (f_j) by year, where f_j is the number of unique females with cubs-of-the-year (F_{CUB}) that were seen exactly j times during that year. Total number of observations is given by $n = \sum_{j=1}^{\infty} jf_j$. Only observations made without the benefit of radiotelemetry and within or <10 miles of the designated grizzly bear recovery zone were included.

| | | Observation frequency | | | | | | | | | | | | | | |
Year	n	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	f_{10}	f_{11}	f_{12}	f_{13}	f_{14}	f_{15}
1986	82	7	5	6	1	1	0	1	2	0	0	0	0	0	0	1
1987	20	7	3	1	1	0	0	0	0	0	0	0	0	0	0	0
1988	36	7	4	4	1	1	0	0	0	0	0	0	0	0	0	0
1989	27	6	5	0	1	0	0	1	0	0	0	0	0	0	0	0
1990	49	7	6	7	1	1	0	0	0	0	0	0	0	0	0	0
1991	62	11	3	3	3	1	2	1	0	0	0	0	0	0	0	0
1992	37	15	5	1	1	1	0	0	0	0	0	0	0	0	0	0
1993	29	7	8	2	0	0	0	0	0	0	0	0	0	0	0	0
1994	29	9	7	2	0	0	0	0	0	0	0	0	0	0	0	0
1995	25	13	2	1	0	1	0	0	0	0	0	0	0	0	0	0
1996	45	15	10	2	1	0	0	0	0	0	0	0	0	0	0	0
1997	65	13	7	4	1	3	0	1	0	0	0	0	0	0	0	0
1998	75	11	13	5	1	1	0	2	0	0	0	0	0	0	0	0
1999	94	9	4	6	2	4	2	0	1	0	0	1	0	0	0	0
2000	72	17	8	1	2	1	0	2	0	1	0	0	0	0	0	0
2001	84	16	12	8	0	1	1	0	0	1	0	0	0	0	0	0

Table 5. Estimates of annual numbers ($\hat{N}_{Obs}$) of females with cubs-of-the-year (F_{CUB}) in the Yellowstone grizzly bear population, 1986–2001. $\hat{N}_{Obs}$ gives the number of unique F_{CUB} actually observed, including those located using radiotelemetry; m gives the number of unique F_{CUB} observed using random sightings only; and $\hat{N}_{SC2}$ gives the second-order sample coverage estimates, per Lee and Chao (1994; Eqs. 3–5). Lower, 1-tailed confidence bounds are for $\hat{N}_{SC2}$ and were calculated using Efron and Tibshirani's (1993) percentile bootstrap method. Also included are annual estimates of relative sample size ($n/\hat{N}_{SC2}$, where n is the total number of observations of F_{CUB}) and of the coefficient of variation among sighting probabilities for individual animals ($\hat{\gamma}$, Eq. 5).

| | | | | Lower 1-tailed confidence bounds | | | | | |
Year	$\hat{N}_{Obs}$	m	$\hat{N}_{SC2}$	70%	80%	90%	95%	$n/\hat{N}_{SC2}$	$\hat{\gamma}$
1986	25	24	31.9	28.4	27.0	25.1	23.5	2.6	0.86
1987	13	12	19.5	16.8	15.2	13.3	11.7	1.0	0.37
1988	19	17	21.5	20.1	19.1	17.7	16.7	1.7	0.25
1989	15	13	20.2	16.9	15.3	13.7	12.3	1.3	0.71
1990	25	22	25.5	24.4	23.5	22.2	21.3	1.9	0.00
1991	24	24	34.5	31.1	29.3	27.0	25.2	1.8	0.63
1992	25	23	47.6	40.0	36.4	32.1	28.9	0.8	0.61
1993	19	17	21.8	20.1	19.0	17.9	16.3	1.3	0.00
1994	20	18	25.5	23.4	21.8	19.9	18.8	1.1	0.00
1995	17	17	54.9	41.2	35.9	28.8	24.7	0.5	0.86
1996	33	28	41.4	38.7	36.6	34.0	31.8	1.1	0.00
1997	31	29	41.3	37.5	35.5	33.0	31.1	1.6	0.57
1998	35	33	40.9	38.4	37.1	35.1	33.7	1.8	0.44
1999	32	29	35.7	33.3	32.1	30.4	29.0	2.6	0.61
2000	35	32	59.7	51.8	48.2	43.8	40.3	1.2	0.90
2001	42	39	54.6	49.5	47.3	44.6	42.2	1.5	0.58

method that performs well over the range of sampling conditions deemed plausible for sightings of F_{CUB} in the GYE. Comparing 7 variations of the Chao (Chao 1984, 1989), jackknife (Burnham and Overton 1978, 1979), and sample coverage (Chao and Lee 1992, Lee and Chao 1994) methods, our provisional choice for estimating numbers of F_{CUB} in the Yellowstone population was the second-order sample coverage estimator, $\hat{N}_{SC2}$. Differences between $\hat{N}_{SC2}$ and the first-order sample coverage estimator, $\hat{N}_{SC1}$, were minor, with both methods converging more rapidly toward N as sample size increased than did other estimators. For both estimators, however, the coefficient

of variation among individual sighting probabilities (CV) affected performance. Over all CV values, $\hat{N}_{SC2}$ exhibited a slightly better balance than $\hat{N}_{SC1}$ between tendencies to overestimate and underestimate when relative sample size (n/N) was in the range of $1.0 < n/N \leq 2.0$ (Fig. 1). Performance under these conditions was seen as particularly important because estimates of n/N for our field study were within this range most years (Table 5).

Chao's (1984) estimator ($\hat{N}_{Chao1}$) showed a greater tendency toward positive bias and exhibited somewhat larger RMSEs than $\hat{N}_{SC2}$ (Figs. 1, 2), but otherwise performed well. Because the most serious biases were associated

with model populations where CV = 0 (an unlikely situation in nature), $\hat{N}_{\text{Chao1}}$ may be a suitable alternative to the sample coverage estimators. However, we cannot recommend the other methods we compared. Over all CV values, RMSEs for $\hat{N}_{\text{Chao2}}$ were lower than for $\hat{N}_{\text{SC2}}$ (Fig. 2), but $\hat{N}_{\text{Chao2}}$ became increasingly and negatively biased as CV increased (Fig. 1). Because individual animals clearly are not equally sightable, use of such an estimator would introduce a chronic, negative bias into estimates of population size and sustainable mortality. Jackknife estimates oscillated, being negatively biased when n/N was small, positively biased at moderate values of n/N, and converging toward N only as n/N increased beyond values observed in our field study (Fig. 1). Neither bias nor RMSE declined monotonically with sample size for any of the jackknife estimators. This suggested that, relative to the other methods examined, larger sample sizes would be needed to achieve comparably accurate estimates and that increased sample size might actually lead to increased bias in some situations. The latter problem was particularly pronounced in the range of $1.0 < n/N \leq 2.0$ (Figs. 1, 2).

In a similar analysis, Mowat and Strobeck (2000) evaluated nonparametric estimators available in the program CAPTURE (Otis et al. 1978, White et al. 1982, Rexstad and Burnham 1991). They selected Burnham and Overton's (1979) best-order jackknife method ($\hat{N}_{\text{Jk1}}$) for estimating numbers of grizzly bears in 2 Canadian populations that showed evidence of "relatively weak heterogeneity" among individual capture probabilities (Mowat and Strobeck 2000:191). Our study differed in important respects. First, all else being equal, the underlying distribution of sighting probabilities should be more heterogeneous in our study (i.e., CV should be larger) because our sample unit consisted of a single sighting rather than a sample period. Second, because our sampling universe included only F_{CUB}, population size appeared to be smaller than the 74 and 262 animals estimated by Mowat and Strobeck (2000). Although population size was not a major determinant of estimator performance in our study, we considered only a narrow range of values ($N = 20$, 40, and 60 animals). Over a larger range, N might emerge as a more important factor. Third, we considered sample coverage estimators (Chao and Lee 1992, Lee and Chao 1994) not available in CAPTURE. Fourth, Mowat and Strobeck (2000), apparently, did not vary sampling effort in a way that would have revealed the oscillatory pattern we observed for the jackknife estimators.

Like all estimators we examined, performance of $\hat{N}_{\text{SC2}}$ varied with n/N. As expected, the largest biases and RMSEs were associated with the smallest relative sample size, $n/N = 0.5$. Performance improved dramatically, however, with even modest increases in n/N, leading us to

recommend a minimum sample size of $n/N = 1$. A nearly unbiased estimate of n/N was $n/\hat{N}_{\text{SC2}}$ (Fig. 5a). Observed values for $n/\hat{N}_{\text{SC2}}$ met or exceeded our recommended minimum for all but 2 years during 1986–2001 (Table 5). This suggested that observed sample sizes were large enough in most years to support fairly good estimates of N (Fig. 1). At this minimal level of sampling effort, however, confidence bounds were sometimes undesirably broad (Table 5). To narrow confidence bounds, we suggest that $n/N = 2$ is a reasonable and achievable goal. Based on estimates of N for 1996–2001 (Table 5), such a goal would translate into target sample sizes of about 80–120 independent random sightings of F_{CUB} per year. This compares with observed sample sizes of 45–94 sightings/year during that same period and indicates a need for increased support for this aspect of the Yellowstone grizzly bear monitoring effort.

Performance of $\hat{N}_{\text{SC2}}$ also varied with the degree of heterogeneity among individual sighting probabilities, as measured by CV. However, such variation was dramatic only when $n/N = 0.5$. When $n/N \geq 1$, $\hat{N}_{\text{SC2}}$ was fairly robust to variations in CV, especially in the range of $0.0 \leq CV \leq 0.75$ (Fig. 1). Even when CV = 1.0, bias was <10%, regardless of n/N (Fig. 1). An advantage of $\hat{N}_{\text{SC2}}$ is that CV is estimated ($\hat{\gamma}$, Eq. 5) as part of the calculation. For 1986–2001, $\hat{\gamma}$ averaged 0.46 and ranged from 0.0–0.9, suggesting that actual CVs were within the range of values in which $\hat{N}_{\text{SC2}}$ performs well. Our Monte Carlo study demonstrated, however, that $\hat{\gamma}$ was negatively biased when CV ≥ 0.25, particularly when n/N is small (Fig. 5). Using calculated values for $n/\hat{N}_{\text{SC2}}$ and $\hat{\gamma}$ (Table 5), rough corrections for such biases can be inferred from Fig. 5. For example, when $n/N = 1.0$ and CV = 0.4, $\hat{\gamma}$ tended to underestimate CV by about 0.2 (Fig. 5). Given $n/\hat{N}_{\text{SC2}} = 1.5$ and $\hat{\gamma} = 0.58$ for 2001 (Table 5), this suggests an unbiased estimate for CV of about 0.85 for that year. Similar inferences for other years yielded a maximum estimated CV of around 1.3 in 2000, but suggested that, overall, CV rarely was much greater than 1. Thus, we believe that actual CVs for sighting probabilities of F_{CUB} in the Yellowstone population typically are within the range of values in which $\hat{N}_{\text{SC2}}$ performs well.

Regardless of method, there is an inherent risk of overestimating N that, in turn, could lead to setting mortality limits at unsustainably high levels. To minimize this risk, we believe it is prudent to base management on some lower, 1-tailed confidence bound. This would provide a specified level of assurance that the population of F_{CUB} is at least as large as estimated. For example, calculated confidence bounds indicated that we can be 95% certain there were at least 42 F_{CUB} in the Yellowstone grizzly bear population in 2001, and 80% certain there were at least 47 (Table 5). To determine whether such bounds accu-

rately depict the risk of overestimating N, we compared nominal versus observed sample coverages using both the BCA and percentile bootstrap methods (Efron and Tibshirani 1993). The BCA method, theoretically, is superior to the percentile method (Efron and Tibshirani 1993). Nonetheless, we recommend the percentile method for this application because the BCA method substantially overstated true coverage under conditions that might reasonably occur in field studies; i.e., when CV= 1.0 and $n/N \geq 2.0$ (see Table 5). Such an error would cause us to understate the true risk of overestimating N. Although the percentile method overstated true coverage when CV = 0.0 and nomimal coverage was 70 or 80%, we view this as less serious because it is not reasonable to expect that CV = 0.0 for natural populations.

In general, we believe $\hat{N}_{SC2}$ is superior to $\hat{N}_{Obs}$ as a basis for calculating mortality limits for Yellowstone's grizzly bears, particularly if lower, 1-tailed confidence bounds are used to minimize the risk of overestimation. In some years, however, depending on the confidence level that is chosen, $\hat{N}_{Obs}$ may be the better alternative. For example, $\hat{N}_{Obs}$ equaled or exceeded the lower, 1-tailed 90% confidence bound for $\hat{N}_{SC2}$ (rounded to the nearest integer) in 8 of the 16 years examined (1986–90, 1993, 1994, 1998, and 1999; Table 5), yet is unburdened by the same risk of overestimation. Thus, it offers a superior estimate of a lower bound for N for those years. This situation occurs largely because $\hat{N}_{Obs}$ incorporates additional information from non-random sightings of radiocollared animals; information that cannot legitimately be used when calculating $\hat{N}_{SC2}$ or its confidence bounds.

Overall, we sought a reliable statistical method for estimating numbers of F_{CUB} because such estimates are essential for setting mortality limits for grizzly bears in the GYE. Given recommended sample sizes, we believe $\hat{N}_{SC2}$ is a reasonable choice for this purpose and that it improves on earlier approaches. We emphasize, however, that knowledge of the number of F_{CUB} is not, by itself, sufficient for setting mortality limits. Other calculations and assumptions are involved that merit additional and comparable scrutiny. Thus, we have refrained from using estimates generated in this study to project total population size or infer acceptable levels of mortality, believing that the remaining issues should be addressed first. An important issue is the assumption that every sighting was correctly identified to individual. Misidentifications undoubtedly occurred, leading to errors of Type I (sightings of the same animal mistakenly classified as sightings of different animals) or Type II (sightings of different animals mistakenly classified as sightings of the same animal). Our experience in applying the rule set of Knight et al. (1995) suggests that Type II errors are much more likely. Such a bias would cause a tendency to undercount the number of unique animals actually seen (m), while also inflating sighting frequencies (n_i values) for the $\hat{m}$ animals estimated to have been seen. In turn, this would lead to estimates of N that are more negatively biased than depicted in our Monte Carlo results, regardless of the estimator that is used. Such a bias, although undesirable, is not by itself inconsistent with our goal of improving on $\hat{N}_{Obs}$ while minimizing the risk of overestimating N. Effects of misidentification on precision are less clear, however. Misidentification introduces uncertainty in sighting frequencies and, thus, would increase uncertainty in estimates based on those frequencies. Our lower, 1-tailed confidence bounds did not incorporate this additional uncertainty and, thus, were probably higher than they would have been if effects of misidentification had been fully accounted for. The tendency toward positive bias in the lower confidence bound would have been countered to some degree by 2 factors. First, any negative bias in $\hat{N}$ resulting from misidentification would necessarily have been accompanied by a similar bias in the confidence bounds surrounding $\hat{N}$. Second, our lower, 1-tailed confidence bounds already were biased low within the range of conditions most often experienced in this study (Fig. 3). Overall, effects of misidentifications on precision would be mitigated, but to an unknown degree. Additional work to better define the nature, magnitude, and consequences of identification errors is needed and has been undertaken. In the meantime, we offer this work as the first in what we hope will be a series of refinements that better ensure reliable estimates of allowable mortality, while minimizing the risk of error.

LITERATURE CITED

BOYCE, M.S., D. MACKENZIE, B.F.J. MANLY, M.A. HAROLDSON, AND D. MOODY. 2001. Negative binomial models for abundance estimation of multiple closed populations. Journal of Wildlife Management 65:498–509.

BUNGE, J., AND M. FITZPATRICK. 1993. Estimating the number of species: a review. Journal of the American Statistical Association 88:364–373.

BURNHAM, K.P., AND W.S. OVERTON. 1978. Estimation of the size of a closed population when capture probabilities vary among animals. Biometrika 65:625–633.

———, AND ———. 1979. Robust estimation of population size when capture probabilities vary among animals. Ecology 60:927–936.

CHAO, A. 1984. Nonparametric estimation of the number of classes in a population. Scandinavian Journal of Statistics 11:265–270.

———. 1988. Estimating animal abundance with capture frequency data. Journal of Wildlife Management 52:295–300.

———. 1989. Estimating population size for sparse data in capture–recapture experiments. Biometrics 45:427–438.

———, W.-H. HWANG, Y.-C. CHEN, AND C.-Y. KUO. 2000. Estimating the number of shared species in two communities. Statistica Sinica 10:227–246.

———, AND S.-M. LEE. 1992. Estimating the number of classes via sample coverage. Journal of the American Statistical Association 87:210–217.

———, M.-C. MA, AND M.C.K. YANG. 1993. Stopping rules and estimation for recapture debugging with unequal failure rates. Biometrika 80:193–201.

DARROCH, J.N., AND D. RATCLIFF. 1980. A note on capture–recapture estimation. Biometrics 36:149–153.

EBERHARDT, L.L., AND R.R. KNIGHT. 1996. How many grizzlies in Yellowstone? Journal of Wildlife Management 60:416–421.

EFRON, B., AND R.J. TIBSHIRANI. 1993. An introduction to the bootstrap. Chapman and Hall, New York, New York, USA.

GREENWOOD, M., AND G.U. YULE. 1920. An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents. Journal of the Royal Statistical Society, Series A 83:255–279.

JOHNSON, N.L., S. KOTZ, AND N. BALAKRISHNAN. 1995. Continuous univariate distributions. Volume 2. Second edition. John Wiley and Sons, New York, New York, USA.

KEATING, K.A. 1986. Historical grizzly bears population trends in Glacier National Park, Montana. Wildlife Society Bulletin 14:83–87.

KNIGHT, R.R., B.M. BLANCHARD, AND L.L. EBERHARDT. 1995. Appraising status of the Yellowstone grizzly bear population by counting females with cubs-of-the-year. Wildlife Society Bulletin 23:245–248.

LEE, S.-M., AND A. CHAO. 1994. Estimating population size via sample coverage for closed capture–recapture models. Biometrics 50:88–97.

LEWONTIN, R.C., AND T. PROUT. 1956. Estimation of the number of different classes in a population. Biometrics 12:211–223.

MOWAT, G., AND C. STROBECK. 2000. Estimating population size of grizzly bears using hair capture, DNA profiling, and mark–recapture analysis. Journal of Wildlife Management 64:183–193.

OTIS, D.L., K.P. BURNHAM, G.C. WHITE, AND D.R. ANDERSON. 1978. Statistical inference from capture data on closed animal populations. Wildlife Monographs 62.

PICTON, H.L., D.M. MATTSON, B.M. BLANCHARD, AND R.R. KNIGHT. 1986. Climate, carrying capacity, and the Yellowstone grizzly bear. Pages 129–135 *in* G. Contreras and K. Evans, compilers. Proceedings of the grizzly bear habitat symposium. U.S. Forest Service General Technical Report INT–207.

PRESS, W.H., S.A. TEUKOLSKY, W.T. VETTERLING, AND B.P. FLANNERY. 1992. Numerical recipes in FORTRAN, the art of scientific computing. Second edition. Cambridge University Press, New York, New York, USA.

REXSTAD, E., AND K. BURNHAM. 1991. User's guide for interactive program CAPTURE. Colorado Cooperative Fish and Wildlife Research Unit, Fort Collins, Colorado, USA.

SEBER, G.A.F. 1982. The estimation of animal abundance. Macmillan, New York, New York, USA.

SMITH, E.P., AND G. VAN BELLE. 1984. Nonparametric estimation of species richness. Biometrics 40:119–129.

TANTON, M.T. 1965. Problems of live-trapping and population estimation for the wood mouse, *Apodemus sylvaticus* (L.). Journal of Animal Ecology 34:1–22.

———. 1969. The estimation and biology of populations of the bank vole and wood mouse. Journal of Animal Ecology 38:511–529.

TAYLOR, S.M. 1966. Recent quantitative work on British bird populations: a review. The Statistician 16:119–170.

U.S. FISH AND WILDLIFE SERVICE. 1993. Grizzly bear recovery plan. U.S. Fish and Wildlife Service, Missoula, Montana, USA.

WHITE, G.C., D.R. ANDERSON, K.P. BURNHAM, AND D.L. OTIS. 1982. Capture–recapture and removal methods for sampling closed populations. Los Alamos National Laboratory LA–8787–NERP, Los Alamos, New Mexico, USA.

WILSON, R.M., AND M.F. COLLINS. 1992. Capture–recapture estimation with samples of size one using frequency data. Biometrika 79:543–553.