

A META-ANALYSIS OF GRIZZLY BEAR DNA MARK-RECAPTURE PROJECTS IN BRITISH COLUMBIA, CANADA

INVITED PAPER

JOHN BOULANGER, Integrated Ecological Research, 924 Innes Street, Nelson, BC V1L 5T2, Canada, email: boulange@ecological.bc.ca

GARY C. WHITE, Department of Fishery and Wildlife Biology, Colorado State University, Fort Collins, CO 80523, USA, email: gwhite@picea.cfnr.colostate.edu

BRUCE N. McLELLAN, British Columbia Ministry of Forests, Box 9358, RPO #3, Revelstoke, BC V0E 3K0, Canada, email: Bruce.McLellan@gems9.gov.bc.ca

JOHN WOODS, Mount Revelstoke and Glacier National Parks, Box 350, Revelstoke, BC V0E 2S0, Canada, email: john_woods@pch.gc.ca

MIKE PROCTOR, Department of Biological Sciences, University of Calgary, 2500 University Avenue NW, Calgary, AB T2N 1N4, Canada, email: mproctor@netidea.com

STEFAN HIMMER, Arctos Wildlife Services, P.O. Box 562, Hagensborg, BC V0T 1H0, Canada, email: shimmer@laketown.net

Abstract: This study evaluates 7 mark-recapture projects using grizzly bear (*Ursus arctos*) DNA from bait stations with hair snags in British Columbia, 1996–98. The emphasis of these projects was to estimate population size. We evaluate how well this objective was met in terms of mark-recapture assumptions and ultimate precision and bias of estimates. We used the Pradel model in program MARK to compare smaller grids and cells with more intensive hair-snag coverage to larger grids and cells. Assuming demographic closure, the Pradel model estimates fidelity, recapture rate, and immigration rate from each of the grids. Results from this analysis suggest (1) closure violation was more likely in smaller grids than larger grids, and (2) recapture rate decreased as cell area increased (and sample intensity decreased). Sex-specific differences were suggested but not verified due to large standard errors of Pradel parameter estimates. We used newer mixture models in program MARK to evaluate heterogeneity with each sampling design and produced meta-analysis population estimates. These estimates were compared with corresponding estimates from the program CAPTURE. Results suggest that heterogeneity is evident with all sampling designs and the newer mixture models provide higher precision than CAPTURE heterogeneity models in most cases. This study highlights the tradeoff in study design faced when designing DNA mark-recapture studies. Namely, intensive sampling using smaller cells is needed to detect and model individual heterogeneity, but smaller grids are most vulnerable to closure violations. Strategies to confront this issue, such as use of radiocollared bears to index movements, are suggested. We discuss applications of recently developed mark-recapture models to DNA, radiotelemetry, and other sources of data that estimate apparent survival, population rate of change, and movements of bears in sampling areas. We argue that use of these newer methods is the most advantageous step in the application of the DNA mark-recapture technique to bear populations.

Ursus 13:137–152 (2002)

Key words: British Columbia, capture-recapture, closure violation, DNA, grizzly bear, heterogeneity variation, mark-recapture, MARK software, mixture models, Pradel models, *Ursus arctos*

The use of individual identification from DNA to produce mark-recapture estimates of grizzly bear population size has been used in Canada (Woods et al., Mowat and Strobeck 2000, Poole et al. 2001) and recently the United States (K. Kendall, U.S. Geological Survey, West Glacier, Montana, USA, personal communication, 2001). This technique shows promise in providing previously unobtainable estimates. However, there has been little study of sampling issues and optimal study design to achieve adequate bias and precision of estimates.

There are 3 principal concerns when applying mark-recapture methods to grizzly bear populations. First, widespread movements of grizzly bears relative to most sampling areas usually violates the assumption of population closure, and, therefore, naive population estimates correspond to the superpopulation (N^* in the immigration-emigration estimator of White [1996]) in the grid and surrounding area (Miller et al. 1997, Kendall 1999). The desired result is rarely the superpopulation but usually the average number of bears on the sampling area or bear density (Miller et al. 1997). Second, bears appar-

ently exhibit heterogeneous capture probability by age and sex classes (Woods et al. 1999) and other factors (Boulangier and McLellan 2001). Finally, obtaining adequate sample sizes and recapture rates to achieve adequate precision can be difficult with typical bear densities.

The main impact of heterogeneous capture probabilities is a negative bias in both population estimates and associated variances in models not robust to heterogeneity variation, leading to biased but apparently precise estimates, which can give a false sense of security when interpreting population estimates (White et al. 1982). To confront heterogeneity variation with sparse data, more complex mark-recapture models can be used, such as the M_h Chao (Chao 1989) heterogeneity estimator (which is theoretically robust to heterogeneity variation when data are sparse) in program CAPTURE (Otis et al. 1978, Rexstad and Burnham 1990). However, these models usually require large sample sizes to obtain precise population estimates (Otis et al. 1978, Chao 1989). Recently, likelihood-based heterogeneity estimators (Norris and Pollock 1996, Pledger and Efford 1998, Pledger 2000)

have become available that may be more efficient than the CAPTURE heterogeneity estimators, and, therefore, may provide more precise population estimates.

The fundamental challenge when estimating grizzly bear populations with DNA-based mark–recapture methods is designing studies to simultaneously minimize closure violation and minimize capture probability variation while obtaining adequate sample sizes. Since 1996, 11 DNA mark–recapture projects have been attempted in British Columbia, Canada, to estimate population size. These projects have used a variety of sampling designs to meet design objectives and budget limitations. We compare these studies in terms of closure violation, capture probability variation, and ultimate precision and bias of estimates. In addition, we apply the M_h models of Norris and Pollock (1996), Pledger and Efford (1998), and Pledger (2000) and evaluate relative gains in precision compared to the Chao M_h models. We also comment on the most appropriate implementation of DNA methodology, in light of the results of the study presented in this report and other similar studies.

METHODS

DNA Mark–Recapture Project Methods and Study Designs

The DNA mark–recapture method attracts grizzly bears to a lure, such as rancid meat, which is usually hung between 2 trees out of the reach of bears. A single-stranded barbed wire corral around the lure snags the hair of approaching bears. Usually 1 hair snag is placed in a cell of a specified area (e.g., 8 x 8 km). Many contiguous cells make up a sampling grid. The hair snags in the entire DNA grid are repeatedly sampled for 4–5 sessions. Hair from each session is used to identify individual bears and record capture and recapture data that are analyzed with statistical models for population estimates (Woods et al. 1999, Mowat and Strobeck 2000, Poole et al. 2001).

A variety of grid and cell areas and design strategies were used in projects in British Columbia. Grid areas

ranged from 1650 (25-km² cell area) to 8527 km² (81 km² cell area; Table 1). For the smaller 25-km² cells, hair snags (the lure and barbed wire) were not moved between sessions. For larger cell areas (>25 km²), hair snags were moved for each session in an attempt to evenly sample cells. When hair snags were moved among sessions, the cumulative sampling coverage of a cell was determined by both cell area and number of sessions. For example, the cumulative spatial hair snag coverage of the Upper Columbia River 1996 project (64-km² cell area/4 sessions = 1 hair snag per 16 km²) and the Prophet project (81-km² cell area/5 sessions = 1 hair snag per 16.2 km²) are similar despite different cell areas. There was a close relationship between cell area and grid area due to the logistical and financial constraints of DNA sampling that limited the number of hair snags (Fig. 1).

Of the 11 projects implemented in British Columbia (BC), 7 are used in this analysis (Table 1). The other 4 projects had inadequate sample designs or field implementation, or used designs that made a meta-analysis comparison problematic. More information about field sampling methods for DNA mark–recapture and other bio-

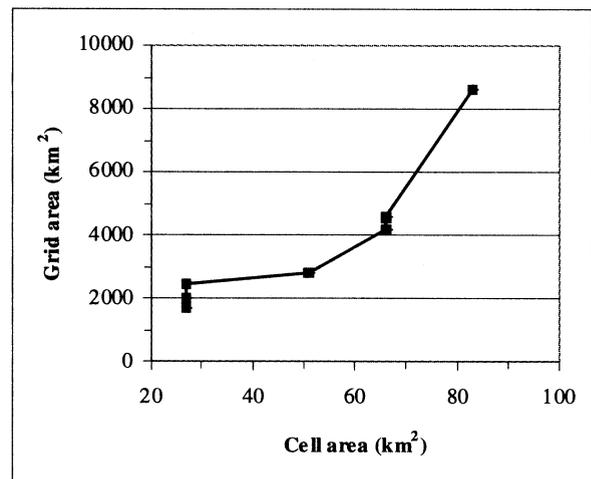


Fig 1. The relationship between grid area and cell area for DNA mark–recapture projects in British Columbia, Canada, 1996–1998.

Table 1. Design specifications for DNA mark–recapture projects in British Columbia, Canada, 1996–98.

Project	Year	Number of cells	Cell area (km ²)	Sessions sampled	Snags moved?	DNA grid area (km ²)	Reference
Jumbo	1998	66	25	4	No	1650	(Strom et al. 1999)
U. Columbia ^a	1997	76	25	5	No	1900	(Boulanger 2001)
U. Columbia	1998	94	25	5	No	2350	(Boulanger 2001)
Kingcome	1997	49	49	5	Yes	2401	(Boulanger and Himmer 2000)
U. Columbia	1996	64	64	4	Yes	4096	(Boulanger 2001)
Granby Kettle	1997	70	64	5	Yes	4480	(Boulanger 2000)
Prophet	1998	103	81	5	Yes	8527	(Poole et al. 2001)

^a Upper Columbia River Bear Research Project (formerly West Slopes Project).

logical findings of studies can be found in Woods et al. (1999), Mowat and Strobeck (2000), and Poole et al. (2001).

Genetic Analysis

Hair samples were analyzed by the Wildlife Genetics International Lab, the University of Alberta in Edmonton, or both. Standard procedures (Paetkau and Strobeck 1994, Paetkau et al. 1995, Woods et al. 1999) were used to identify individual bears from hair samples. An equation described in Woods et al. (1999) was used to estimate the conditional probability that a given individual shared the same genotype as a sibling (P_{sib}). Matches were accepted when P_{sib} was <0.05 . Sex of bears was initially determined using the SRY-ZFX/ZFY system as described in Woods et al. (1999) and Taberlet et al. (1993). In 1998, projects switched to the Amelogenin system to determine sex (Ennis and Gallagher 1994).

An assumption of this analysis is that the genetic identification of individuals is error-free. The sources of error related to our genetic data fall into 2 categories. First, errors in genotyping, scoring, and amplification (PCR) tend to create false individuals and thus positively bias estimates. Second, 2 individuals could have identical genotypes. They would erroneously be considered 1 individual, which would negatively bias estimates (termed the shadow effect by Mills et al. [2000]).

To deal with false individuals, all highly similar genotypes, single mismatches, and genotypes represented by only 1 sample were evaluated for scoring errors by reviewing original gels and amplification errors (i.e., allelic dropout or nonspecific bands on gel misinterpreted as alleles; Gagneux et al. 1997, Goosens et al. 1998, Taberlet et al. 1996, Taberlet et al. 1999), followed by reamplification (PCR) and electrophoresis. Each project's single-mismatch rate was compared against a set of genotypes from a population of bears that were genotyped from tissue, and thus had a low amplification error rate, and were captured and handled by researchers so the number of individuals was known. The reduction of biases in population estimates through error checking or filtering procedures such as employed in this study has been studied extensively by D. Roon (University of Idaho, Moscow, Idaho, USA, personal communication, 2001), who found that bias could be substantially reduced to acceptable levels if proper error checking procedures were used.

To deal with shadow effects, we used a threshold of 0.05 probability that any individual's genotype might match a full sibling (P_{sib} ; Woods et al. 1999). Individuals with a genotype that had a $P_{sib} > 0.05$ were removed from the analysis. The total probability of a match with other individuals within a population is difficult to calculate. The proportion of various relatives in a population is never

known, given that it would be necessary to obtain the weighted sum of probabilities of all potential relationships in the population. We therefore turned to an empirical example to see how common it would be for 2 individuals to have matching genotypes. D. Paetkau (Wildlife Genetics International, Nelson, BC, Canada) used 2 populations (of 134 and 119 individuals) of grizzly bears where relationships were ascertained over many years and microsatellite genotypes were determined. From these 2 populations, the rate of 2 individuals matching at the same 6 loci we used was 1 in 16,000 pairs of individuals compared. Furthermore, siblings were the predominant source of paired genotypes that mismatched by only 1 allele. We concluded that the rate of 2 individuals matching was most likely to occur with a full sibling (thus the P_{sib} criteria) and was diminishingly low as a bias in population estimation.

Comparison of Study Designs

Closure Violation and Recapture Rates.—The Pradel (1996) model as incorporated in program MARK (White and Burnham 1999), which estimates apparent survival (ϕ), recruitment (f), and recapture probability (p), was used to investigate the tradeoff between cell area and grid area in terms of recapture rates and violation of grid closure. We assumed that the population of bears was demographically closed for this analysis. The duration of sampling was approximately 2 months, and because bear mortality rates are low (McLellan et al. 1999), this assumption was reasonable. Apparent survival equaled true survival (S ; due to mortality) times the fidelity of bears (F) to the sampling grid ($\phi = SF$). Because the population was demographically closed, we assumed that true survival equaled 1 ($S = 1$) and therefore relative changes in F reflected bear fidelity to the sampling grid rather than actual mortalities (i.e., $\phi = F$). The Pradel recruitment rate estimates the number of new individuals in the population at time $j + 1$ per individual at time j . Because there were no births during sampling, f measures recruitment in the form of immigration onto the sampling grid or the rate of additions of bears into the sampling grid. Here, f is a rate and not an absolute measure of additions, such as the birth rate estimator (B_i) of the Jolly-Seber model (Seber 1982; i.e., $f_i = B_i/N_i$). Both fidelity (F) and recruitment rate (f) are measures of movement that pertain to 1 exit or entry from the grid per sampling session and are less sensitive to temporary movement across grid boundaries and only describe 1 component of closure violation. Therefore, this test should be considered an approximation of grid permeability to closure; if fidelity is low and additions high, then there are likely also high levels of temporary movement from the grid. Models were constrained to estimate fidelity, recapture rate, and rates of addition as a function

of sex, grid area, cell area, and whether hair snags were moved between sessions. *A priori*, we suspected differences in estimates of F (lower for males) and f (greater for males) because of larger home range sizes of males (McLellan 1989). Further, with increasing grid area, the ratio of edge to area becomes smaller, so we expected that increasing grid area would result in higher F and lower f values. Decreasing cell area of the sampling grid should positively affect p by increasing trap encounter rate. Covariates were standardized by mean and standard deviation and entered into MARK design matrices to formulate models. A logit link was used for the analysis.

The fit of models was evaluated using the Akaike information criterion (AIC) index of model fit. The model with the lowest AIC_c score was considered the most parsimonious, thus minimizing estimate bias and optimizing precision (Burnham and Anderson 1998). Change in AIC_c (Δ AIC_c) values was also used to evaluate the fit of models when AIC_c scores were close. In general, any model with a Δ AIC_c score of <2 was worthy of consideration. AIC_c weights (abbreviated as w_i) were calculated to determine the proportional support for each of the candidate models. Parameter estimates were averaged (termed model averaging) based on their support by the data (as indexed by AIC_c weights) to further account for model selection uncertainty (Burnham and Anderson 1998).

An assumption of the Pradel analysis was that capture, survival, and movements are independent. In addition, it is assumed that all individuals within a group have similar apparent survival rates and similar values of other model parameters. If individuals are not independent, the multinomial variances from the models become inflated or overdispersed, which causes underestimation of parameter variances and overfitting of models. Various goodness-of-fit tests are available to test and estimate the degree of overdispersion in the data set. Because lack of fit in the Pradel models can only be assessed for the recapture por-

tion of the encounter history, the goodness-of-fit test in Program RELEASE (Burnham et al. 1987) was used to assess goodness-of-fit. As suggested by White et al. (2002), we also used a bootstrap goodness-of-fit test for the Cormack-Jolly-Seber (Seber 1982) model to estimate overdispersion for the Pradel model under the assumption that lack of fit was due to recaptures of previously marked animals. If overdispersion was detected (as indicated by $\hat{c} > 1$), we used QAIC_c instead of AIC_c model selection criterion to select optimal models (Burnham and Anderson 1998, White et al. 2002).

Genetic tests to determine sex failed on a small segment of each project data set and these bears were excluded from the analysis (Table 2). Only the first 4 sessions of each project were considered in the Pradel analysis. This was to accommodate projects that had 4 sessions and to circumvent decreasing capture probabilities (time variation) that occurred on the fifth session of many of the mark-recapture projects. Time-specific parameterization was not considered because the average fidelity and rates of addition were of interest, and therefore a time-specific formulation complicated comparisons.

Sample Sizes to Detect Capture Probability Variation.—Studies were compared based on sample size (marked bears and capture probabilities) criteria defined by simulation studies, to determine if the studies had high enough sample sizes to detect likely forms of capture probability variation. White et al. (1982) recommended capture probabilities >0.3 for populations <100 and >0.2 for populations of 100–200 for adequate power to detect capture probability variation with program CAPTURE. Simulation studies based on the likely heterogeneity bias with bear populations further verified that power to detect heterogeneity and other forms of capture probability variation is low using the program CAPTURE model selection routine (M_h vs. M_o test) when capture probabilities are <0.2 and population size is <200 (Boulanger 2001).

Table 2. Summary statistics and program CAPTURE estimates for DNA mark-recapture projects conducted in British Columbia, Canada, 1996–98.

Project	Bears identified ($M_{t,t}$)			Model	Superpopulation estimates				
	males	females	unk ^a		\hat{N}	SE	CV	\hat{P}	Coverage ^b
Jumbo	10	18	5	M_h (Chao)	45	7.1	15.8%	0.26	73.3%
U. Columbia 97	12	23	5	M_h (Chao)	55	9.5	17.3%	0.20	72.7%
U. Columbia 98	22	13	5	M_t (Chao)	92	29.8	32.4%	0.12	43.5%
Kingcome	36	16	6	M_h (Chao)	102	20.7	20.3%	0.20	56.9%
U. Columbia 96	30	25	0	M_h (Chao)	108	23.8	22.0%	0.16	50.0%
Granby Kettle	7	12	3	M_t (Chao)	46	16.4	35.7%	0.13	47.8%
Prophet	47	46	5	M_h	166	26.2	15.8%	0.17	59.0%

^a Unknown sex due to genetic sex test failure.

^b Percent of the estimated superpopulation sampled ($M_{t,t}/\hat{N}$)

Analysis of Heterogeneity Variation and Its Effect on Population Estimates

The newer M_h mixture models of Norris and Pollock (1996), Pledger and Efford (1998) and Pledger (2000) were applied to the data, and these results were compared with the CAPTURE models. CAPTURE models were selected based on probable heterogeneity of bear capture rates, results of the CAPTURE model selection tests, and post hoc simulation tests performed on each of the data sets. The Chao M_h and M_t estimators were chosen over the M_h (jackknife) and M_t (Darroch) estimators due to reduced performance of these estimators with sparse data (Otis et al. 1978, Chao 1988, Chao 1989, Chao and Jeng 1992, Rosenberg et al. 1995).

M_h mixture models use a mixture of ≥ 2 capture probabilities to model heterogeneity of a single capture probability. This allows bimodal or multimodal distributions that may arise from heterogeneity of capture probabilities to be modeled. For example, the overall capture probability for an encounter history where a mixture of A distributions is used is

$$\sum_{i=1}^A \pi_i \theta_i^v (1 - \theta_i)^{t-v}$$

where v equals the number of captures of the animal for t occasions, π_i is the probability the animal has capture probability θ_i , with the sum of the π_i forced to equal 1. Thus, for $A = 2$, $\pi_2 = 1 - \pi_1$. From Carothers (1973) the mean capture probability ($\bar{\theta}$) (based on 2 mixture distributions) and coefficient of variation for the mean capture probability (CV ($\bar{\theta}$)) were estimated as

$$\bar{\theta} = \pi_1 \theta_1 + (1 - \pi_1) \theta_2$$

and

$$CV(\bar{\theta}) = \left[\sqrt{\pi_1(1-\pi_1)} \cdot |\theta_1 - \theta_2| \right] / \bar{\theta}$$

The coefficient of variation of $\bar{\theta}$ was used as an index of heterogeneity variation. A higher CV ($\bar{\theta}$) would indicate a greater degree of dispersion in capture probabilities, suggesting a higher degree of heterogeneity variation.

Both 4- and 5-session data sets were used in the analysis. Capture probability parameters for the fifth session of 4-session data sets were fixed at 0 to reflect lack of data for session 5. Sexes were pooled for the analysis due to unknown-sex bears in most of the data sets and to allow comparison with CAPTURE estimates, which also pooled sexes (Table 2). M_h mixture models, plus the M_t and M_o likelihood models, were used in Program MARK where capture probabilities were different for each of the 7 areas and constant for all areas. In addition, the cell area for each of the projects was entered as a covariate in the analysis. It was hypothesized that projects with smaller cells should exhibit higher capture probabilities but reduced heterogeneity (due to less difference in trap encoun-

ter rates between bears) than projects with larger cells. In addition, the degree of heterogeneity observed, as indexed by the probability of mixture, π_i , would be greater in smaller cells due to higher capture probabilities and subsequent increased power to discern heterogeneity variation. An alternative hypothesis was that larger grid areas (and cell areas) would exhibit decreased edge effects (bears on the edge of grids only being vulnerable to capture for a proportion of sampling periods) that would reduce individual heterogeneity at larger cell areas. As with the Pradel analysis, models were evaluated using AIC_c model selection methods. Population estimates were averaged across both heterogeneous and non-heterogeneous models. Mixture probabilities were averaged across similar distribution (M_{h2}) mixture model formulations.

RESULTS

Comparison of Study Designs: Pradel Analysis of Closure Violation and Recapture Rates

The Program RELEASE goodness-of-fit test for model $\{\phi[\text{project} \times \text{time}] p[\text{project} \times \text{time}]\}$ suggested minimal overdispersion of multinomial likelihoods or

$$\hat{c} = 1 \quad (\chi^2 = 19.304, 38 \text{ df}, P = 0.995)$$

A bootstrap goodness-of-fit test with the Cormack-Jolly-Seber (CJS) model was also used to assess overdispersion in the data set. A slightly reduced CJS model was used ($\phi[\text{project}] p[\text{project}]$ in Table 3) due to high standard errors of parameter estimates for the full sex \times project model. An estimate of $\hat{c} = 1.19$ was generated by the bootstrap test suggesting moderate overdispersion. We decided to use $\hat{c} = 1$ for simplicity in model selection and because simulations of the bootstrap approach have shown the estimates to be biased compared to the Program RELEASE goodness-of-fit test (White 2002).

A variety of factors influenced fidelity and recapture rates as indicated by the large number of models that were supported by the data (with $\Delta AIC_c < 2$) (Table 3). The minimum AIC_c model (Model 1) suggests that fidelity varied as a function of grid area and sex, recapture rate was constant, while additions were a function of sex. Other supported models suggested that recapture was a function of cell area (Model 2) and immigration rate was a function of grid area (Model 3). As shown by the ΔAIC_c values, the model with constant fidelity, immigration, and recapture probabilities (Model 18) had little support, with $\Delta AIC_c = 3.37$, suggesting that variation in these parameters across study areas was occurring.

Table 3. Pradel model results for meta-analysis of British Columbia, Canada, DNA mark-recapture projects, 1996–98. Akaike Information Criteria (AIC_c), the difference in AIC_c values between the i th model and the model with the lowest AIC_c value (Δ_i), Akaike weights (w_i), and number of parameters (K) are presented.

Model	Fidelity	Recapture rate	Immigration	AIC_c	Δ_i	w_i	K	Deviance
1	sex + grid area ^a	(.) ^b	sex	1365.15	0.00	0.127	6	942.11
2	sex + grid area	cell area	sex	1365.70	0.55	0.097	7	940.60
3	sex + grid area	(.)	sex + grid area	1366.14	0.99	0.078	7	941.04
4	grid area	(.)	sex	1366.23	1.09	0.074	5	945.26
5	grid area	(.)	(.)	1366.30	1.16	0.071	4	947.38
6	grid area	cell area	(.)	1366.62	1.47	0.061	5	945.65
7	sex x grid area	sex	(.)	1366.68	1.53	0.059	7	941.58
8	sex + grid area	hair snag moved	sex	1366.95	1.80	0.052	7	941.85
9	grid area	(.)	grid area	1367.15	2.00	0.047	5	946.18
10	sex x grid area	(.)	(.)	1367.21	2.06	0.045	7	942.11
11	sex x grid area	sex + cell area	(.)	1367.38	2.23	0.042	8	940.20
12	grid area	sex x cell area	(.)	1367.57	2.42	0.038	7	942.47
13	sex + grid area	sex + cell area	(.)	1367.64	2.49	0.037	7	942.54
14	grid area	hair snag moved	(.)	1368.08	2.93	0.029	5	947.11
15	sex x grid area	cell area	(.)	1368.36	3.21	0.026	7	943.26
16	grid area	cell area	grid area	1368.44	3.29	0.024	6	945.41
17	sex	sex	(.)	1368.47	3.32	0.024	5	947.50
18	(.)	(.)	(.)	1368.52	3.37	0.024	3	951.63
19	sex x grid area	sex x cell area	(.)	1369.12	3.97	0.017	9	939.85
20	sex	(.)	sex	1369.88	4.73	0.012	5	948.91
22	(.)	hair snag moved	(.)	1370.51	5.36	0.009	4	951.59
23	sex	sex	sex	1370.53	5.38	0.009	6	947.50
24	project	project	project	1384.45	19.30	0.000	21	929.28
25	project x time	project x time	project x time	1393.53	28.38	0.000	42	889.02

^a Covariates were modeled as an interaction (symbolized by x term specific slope and intercept) or as additive (symbolized by + term specific slope but common intercept).

^b Parameter did not vary.

The best way to interpret the results of this analysis is through model averaged parameter estimates (Fig. 2–4). A relationship is evident between fidelity and grid area, with smaller grids showing lower levels of fidelity than larger grids (Fig. 2); however, the level of precision in estimates is not high, suggesting a weak trend. Sexes displayed similar trends, however, the fidelity of males remained low at larger grid areas when compared to females. Contrary to our hypothesis, fidelity of females was lower than males when grid area was small.

In contrast to fidelity, recapture rates were highest in smaller grids with smaller cells and decreased as grid area increased (Fig. 3). Probable individual heterogeneity of capture probabilities causes Pradel model recapture rates to be positively biased, and therefore the actual recapture rates are artificially high. Immigration rates were relatively constant for all grid areas. Males displayed lower immigration rates than females, a result contrary to what we expected (Fig. 4). Estimates of immigration will also be sensitive to heterogeneity, which may explain the counterintuitive results. The difference in trend in fidelity, immigration rates, and recapture rates was negligible between sexes, potentially due to large standard errors of sex-specific estimates and the dilution of sex-specific recapture rates by age class.

Detection of Capture Probability Variation.—All projects were below the minimal capture probability level of 0.3 (as estimated by Chao models) to ensure adequate

power to detect capture probability variation (Table 3). Heterogeneity variation was detected in the Jumbo (M_{ibh} , chosen by model selection routine) (Strom et al. 1999) and Kingcome (M_h chosen; Boulanger and Himmer 2000) data sets. Time variation was detected with the Prophet (M_t chosen; Poole et al. 2001), Upper Columbia River 98 (M_t chosen), and Granby Kettle (M_t chosen; Boulanger 2000) projects. Behavior variation was detected in the Jumbo data set, which may have been due to closure violation (Strom et al. 1999). No forms of capture probability variation were detected in the Upper Columbia 1996 and 1997 projects using the CAPTURE model selection routine.

Comparison of CAPTURE and MARK Heterogeneity Estimators

CAPTURE Estimates.—The M_h , M_t , or M_{ih} models (Chao 1988, Chao 1989, Chao and Jeng 1992) were used to estimate population size using CAPTURE. For some data sets (Granby Kettle, Upper Columbia River 1998), time variation, low capture probabilities, and low sample sizes of marked bears compromised the performance of the M_h Chao and jackknife estimators, so M_t Chao estimators were used (Boulanger 2001). In the Prophet data set, time and heterogeneity variation were evident, so the M_{ih} estimator was used (Boulanger and McLellan 2001; Table 2). Pollock et al. (1990) states that the CV should be less than 20% for using estimates for management pur-

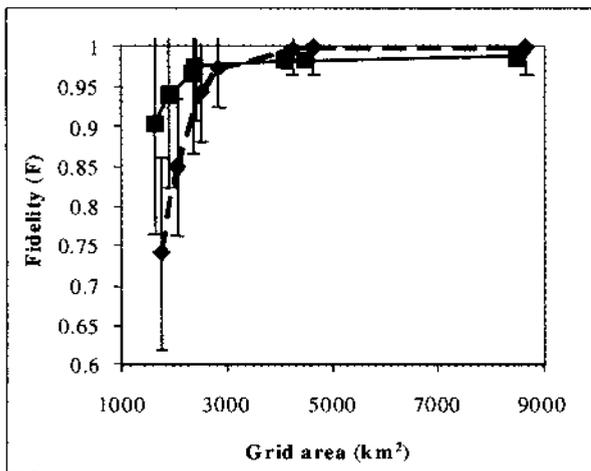


Fig. 2. Model averaged estimates of grid fidelity for male (solid line) and female (broken line) grizzly bears as a function of grid area from Pradel analysis for DNA capture-recapture studies in British Columbia, Canada, 1996–98.

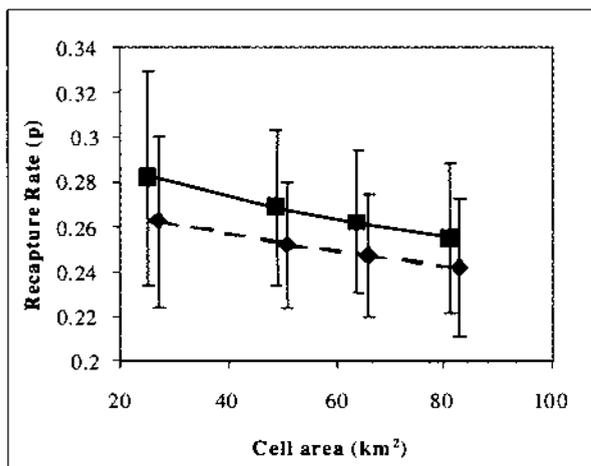


Fig. 3. Model averaged estimates of recapture rate for male (solid line) and female (broken line) grizzly bears from Pradel analysis as a function of cell area and sex for DNA capture-recapture studies in British Columbia, Canada, 1996–98. These estimates will be larger than closed model capture probability estimates (Table 2) due to heterogeneity variation, and because the Pradel model accounts for grid fidelity and additions which reduce closed model recapture rates.

poses. Four of the projects (Jumbo, Upper Columbia River 1997, Kingcome, and Prophet) met this criterion when the Chao estimators were used.

A general relationship between capture probability and estimate precision was evident from CAPTURE model estimates (Fig. 5). Only 3 studies (Jumbo and Upper Columbia 1997 with 25-km² cell size and Kingcome with 49-km² cell size) with capture probabilities ≥ 0.2 achieved coefficients of variation of $<20\%$ when population size was less than 108 bears. The Prophet study (81-km² cells

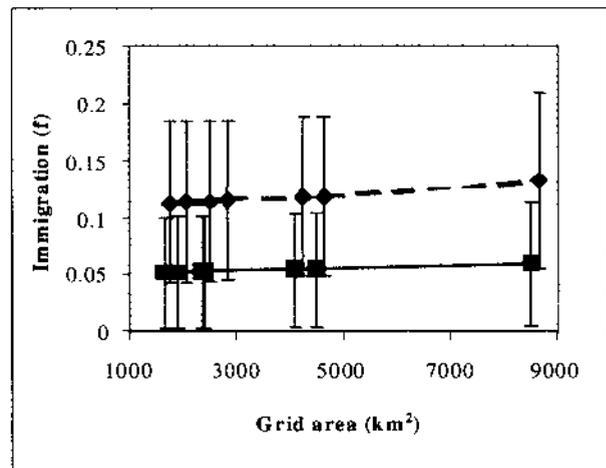


Fig. 4. Model averaged estimates of immigration rate for male (solid line) and female (broken line) grizzly bears from the Pradel analysis as a function of cell area and sex for DNA capture-recapture studies in British Columbia, Canada, 1996–98. Error bars are unconditional model averaged standard errors.

sampled for 5 sessions) also displayed a coefficient of variation of $<20\%$ but had a larger population size of 166 bears and 5 sample sessions to compensate for lower capture probabilities (Table 2).

MARK M_h Mixture Models.—The minimum AIC_c model for the closed capture analysis of the 7 study areas was a mixture of 2 distributions used for all 7 areas with a quadratic relationship between cell area and probability of mixture (Table 4, Model 1). In addition, models with a linear (Model 2) and quadratic relationship between mean capture probabilities for the 2 mixture distributions and grid cell area were also supported. The equivalent non-heterogeneous model with only 1 distribution (Model 14; M_0 of CAPTURE with quadratic cell covariates) was 15.2 AIC_c units lower, and the more complex 3-mixture distribution (Model 10) was 10.06 AIC_c units lower. Models incorporating time effects (Model 11) were 11.03 AIC_c units lower, suggesting that individual heterogeneity was a more important source of variation in capture probabilities than time-specific variation.

Observation of mean mixture capture probabilities (θ) and coefficients of variation of mixture capture probabilities ($CV(\theta)$) suggests a decrease in capture probability and a slight decrease in detected heterogeneity with increasing cell area (Tables 1, 5). One exception is the Kingcome project, which exhibited a higher $CV(\theta)$ than other projects. Also, θ and $CV(\theta)$ could be linearly related, which would suggest that heterogeneity variation is more detectable at higher capture probabilities rather than differences in heterogeneity with cell area. Coefficient of variations of mixture model population estimates met the criteria of 20% for use of estimates in manage-

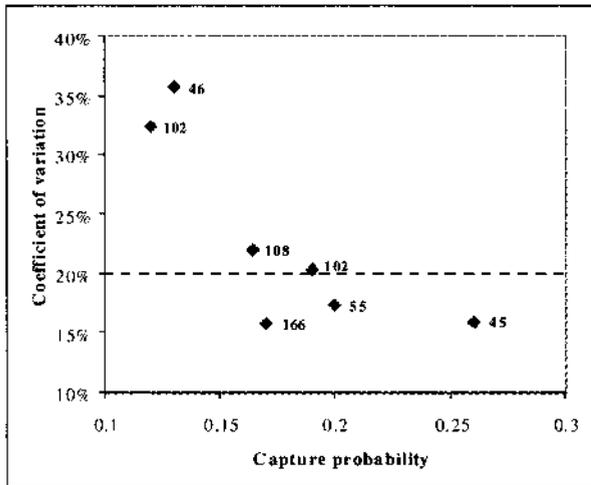


Fig. 5. The relationship between capture probability, estimated superpopulation size, and coefficient of variation for DNA capture–recapture studies of grizzly bears in British Columbia, Canada, 1996–98. Numbers beside data points are CAPTURE superpopulation estimates (Table 5). The horizontal reference line denotes the acceptable coefficient of variation level (20%).

ment in 4 of the 7 projects. In 5 of the 7 projects, the CV from mixture models was lower than that of CAPTURE estimates, suggesting improved precision. The population estimates were similar, with confidence intervals overlapping point estimates for both the CAPTURE and

mixture models. Further simulation modeling is needed to verify mixture model population estimates.

DISCUSSION

Comparison of Designs

Closure Violation.—This meta-analysis illustrates the challenges of estimating grizzly bear population size using DNA and other methods. Suitable estimates require intensive sampling using smaller cell areas than have been commonly used to achieve higher recapture rates for optimal performance of program CAPTURE. However, smaller grids with smaller cells are most prone to closure violation (Fig. 2) and may have substantially reduced numbers of bears at risk of capture.

The relationship between grid area and bear fidelity was not always clear, as many competing models were supported by the data (Table 3). This inconsistent trend was likely due to some studies having relatively little closure violation due to topography. For example, the boundaries of the Jumbo, Kingcome, and Upper Columbia River (1997) grids were likely partially closed by ice fields, highways, and large lakes, which would have reduced movement within the 1–2 month duration of these projects. The fidelity of bears to these areas was therefore higher than other projects with similar designs but more open topography (Upper Columbia 1998), leading to a large degree of variance in fidelity estimates for smaller grid areas (Fig. 2).

Table 4. Program MARK closed model selection results for meta-analysis of DNA mark–recapture projects in British Columbia, Canada, 1996–98. Akaike Information Criteria (AIC_c), the difference in AIC_c values between the *i*th model and the model with the lowest AIC_c value (Δ_i), Akaike weights (w_i), number of parameters (K), and deviance are presented.

Model	Model	AIC _c	Δ _i	w _i	K	Deviance
1	$M_h(2) \pi + gc, gc^{3a}, \theta_1 \& \theta_2 (\cdot)^b$	-225.57	0.000	0.237	12	232.33
2	$M_h(2) \pi + gc, gc^2, \theta_1 \& \theta_2 (+gc)$	-225.06	0.504	0.184	13	230.80
3	$M_h(2) \pi + gc, gc^2, \theta_1 \& \theta_2 (+gc, gc^2)$	-224.84	0.726	0.165	14	228.99
4	$M_h(2) \pi + gc, gc^2, \theta_1 \& \theta_2 (xgc)$	-224.74	0.829	0.157	14	229.09
5	$M_h(2) \pi + gc, \theta_1 \& \theta_2 (xgc)$	-224.42	1.150	0.134	13	231.44
6	$M_h(2) \pi + gc, gc^2, gc^3, \theta_1 \& \theta_2 (\cdot)$	-223.54	2.023	0.086	13	232.32
7	$M_h(2) \pi + gc, \theta_1 \& \theta_2 (\cdot)$	-220.15	5.413	0.016	11	239.77
8	$M_h(2)$	-219.55	6.021	0.012	10	242.40
9	$M_h(2) \pi \times \text{project}^c, \theta_1 \& \theta_2 (\cdot)$	-217.99	7.574	0.005	16	231.76
10	$M_h(3)$	-215.50	10.064	0.002	12	242.39
11	project* M_i	-214.55	11.013	0.001	40	185.58
12	project* M_o	-213.85	11.716	0.001	14	239.98
13	$M_n, p+gc$	-211.14	14.425	0.000	9	252.83
14	M_n	-210.41	15.153	0.000	8	255.58
15	$M_n, p+gc, gc^2$	-209.96	15.603	0.000	10	251.98
16	project* $M_h(2)$	-209.96	15.607	0.000	28	215.16
17	$M_h(2) \pi(\cdot), \theta_1 \& \theta_2 (\text{project})$	-209.96	15.608	0.000	22	227.52
18	$M_{ij}(2)$	-181.30	44.270	0.000	84	124.09
19	M_{ij}	-168.55	57.021	0.000	9	295.42

^a Cell area (gc) covariates were incorporated as an additive term (different intercept, same slope) or multiplicative (different intercepts and slopes).

^b Parameter was constant.

^c Project-specific parameterization of model.

Table 5. Comparison of MARK M_h mixture model and CAPTURE estimates for DNA mark-recapture projects, British Columbia, Canada, 1996–98. All coefficients of variation (CV) are expressed as a percentage.

Project	M_h mixture model averaged estimates					CAPTURE estimates							
	\hat{N}	SE	CV	CI		$(\bar{\theta})$	CV ($\bar{\theta}$)	Model ^a	\hat{N}	SE	CV	CI	
Jumbo	68	22.41	33	23	111	0.22	73	M_h	45	7.1	16	37	68
U.C.R. ^b 1997	72	23.03	32	27	117	0.22	73	M_h	55	9.5	17	48	107
U.C.R. 1998	72	17.47	24	22	122	0.22	73	M_h	92	29.8	32	71	301
Kingcome	101	16.9	17	60	143	0.20	82	M_h	102	20.7	20	77	163
U.C.R. 1996	111	19.56	18	73	150	0.16	44	M_h	108	23.8	22	78	177
Granby	38	7.57	20	23	53	0.16	44	M_h	46	16.4	36	30	102
Prophet	154	15.4	10	123	183	0.19	19	M_h	166	26.2	16	132	239

^a Chao (1989) estimators were used for all models.

^b Upper Columbia River Bear Research Project.

The Pradel analysis is similar to the closure test of Stanley and Burnham (1999) in that estimates of rates of addition (immigration) and dilution (fidelity) are used to assess closure violation. However, this analysis utilizes changes in actual parameter estimates as a function of study design parameters, whereas the test of Stanley and Burnham (1999) provides a hypothesis testing framework for testing closure violation which is less flexible in application. However, the approach of Stanley and Burnham (1999) can statistically detect permanent versus temporary movement across grid boundaries. Boulanger and McLellan (2001) demonstrated that the Pradel model can be used to emulate the tests of Stanley and Burnham (1999) by fixing model parameters in Program MARK.

The net result of closure violation is that the population numbers, as estimated from closed models, correspond to the superpopulation of bears in the grid and surrounding area if movement across grid boundaries is temporary and random (Kendall 1999). If movement across boundaries is not temporary and random, a variety of biases can occur that depend on the type of movement (Kendall 1999). All forms of movement (immigration only, emigration only, Markovian temporary emigration, random temporary emigration, one entry one exit) discussed by Kendall (1999) might occur during a DNA project. However, in most cases, temporary movement was more likely given the short time scale of the project and the timing; in the spring, home ranges of bears are relatively stable. In addition, the pattern of results from the test of Stanley and Burnham (1999) applied to the projects in this study suggested temporary rather than permanent movement (Boulanger, unpublished data). Therefore, application of one method of estimation geared toward a specific pattern of movement as suggested by Kendall (1999) is problematic, and the assumption of random temporary emigration (and the use of closed models) is the most parsimonious for this work.

When closure is violated the open Jolly-Seber model will not give an unbiased superpopulation estimate unless all movement corresponds to permanent "1 entry 1 exit" movement during a study and no temporary move-

ment occurs (Kendall 1999). A situation in which only permanent movement occurs is unlikely with short period studies such as the projects in this study. Permanent movement is more likely with longer term sampling studies such as those on polar bears (*Ursus maritimus*) in which a yearly sampling was employed (Amstrup et al. 2001). In addition, the Jolly-Seber model is not robust to heterogeneity of capture probabilities (Gilbert 1973), although recent development of the model to allow covariates (McDonald 2001) might improve its performance.

New methods may allow more robust estimates of the superpopulation. Boulanger and McLellan (2001) used the mean distance of capture from the grid edge of individual bears as a covariate with the Huggins (1991) closed model in program MARK to obtain a robust superpopulation estimate for the Prophet River study (Poole et al. 2001). This analysis revealed that bear capture probabilities were lowest near grid edges, presumably because bears located there spent less time on the sampling grid. This form of heterogeneity variation was not detected by the CAPTURE model selection routine, presumably due to low power. The superpopulation estimation method of Boulanger and McLellan (2001) directly considers the influence of movement across grid boundaries on capture probability, and therefore is potentially less dependent on assumptions regarding the type of movement across grid boundaries.

The amount that superpopulation estimates vary from the actual average number of bears on the grid is difficult to determine from the DNA data alone. Various methods to approximate the degree of closure violation have been proposed. Boulanger and McLellan (2001) used Pradel models to assess the relationship between bear capture probability and the average distance from grid edge that it was captured. Using this relationship they defined a core population of bears to approximate the grid-based population size. This approach attempts to eliminate the edge portion of the population to reduce bias under the assumption that edge bears will most likely exhibit the highest temporary and permanent movement during the study. This technique relies on even densities of bears in the sam-

pling grid so that the core population can be extrapolated to the entire grid area, and high enough sample sizes to allow fitting of Pradel (1996) models. Using this method, Boulanger and McLellan (2001) estimated a 25% difference (using CAPTURE model M_{th} for estimates) between superpopulation and average N estimates for the 8,527 km² Prophet River study, suggesting that closure violation cannot be mitigated by increasing grid size alone. Poole et al. (2001) assessed differences in male and female population estimates from the Prophet River study, estimating a 6.8% difference between average N and superpopulation estimates under the assumption of reduced closure violation by females and an equal sex ratio of bears. Poole et al. (2001) suggest that their method may underestimate closure violation, a conclusion supported by Boulanger and McLellan (2001).

A similar method to scale superpopulation estimates is the nested subgrid routine found in program CAPTURE; however, simulation studies suggest that sample sizes required for this routine are much larger than found with most bear studies (Wilson and Anderson 1985a). Wilson and Anderson (1985b) use the mean distance between successive captures of individuals to estimate the effective sampling area. This technique assumes that the population is completely open with no barriers to movement on the grid edges. This assumption probably does not apply to the British Columbia studies in which a portion of each sampling grid was topographically closed.

The most reliable method to correct for closure violation is to monitor a radiocollared segment of the population to index movements across grid boundaries (Miller et al. 1997, Powell et al. 2000, White and Shenk 2001). Capturing and collaring bears adds cost and is intrusive, but better estimates the average number of bears on the sampling area. Radiocollared bears were present in the Upper Columbia River study areas, with results suggesting closure violation with all grid designs that allowed scaling of superpopulation estimates to grid based average N estimates and density (Boulanger 2001).

Sample Size, Grid Area, Cell Area, and Detection of Capture Probability Variation.—This study highlights the challenge of obtaining adequate sample sizes to ensure power to detect capture probability variation and choose appropriate models. Many studies have shown that bears exhibit heterogeneous capture probabilities as a function of sex (Mace et al. 1994) or where the bear was captured relative to the grid edge (Boulanger and McLellan 2001) and have suggested differences due to age (Woods et al. 1999). Results of the M_h mixture analysis suggest that heterogeneity was evident in all projects as indicated by model selection results and point estimates of coefficient of variation. However, heterogeneity was only detected in the Jumbo and Kingcome studies (by the program

CAPTURE model selection routine), which also exhibited higher recapture rates than other studies. Heterogeneity was probably evident in other studies, but the power of statistical tests was insufficient to detect it. Newer tests for heterogeneity (MacKenzie and Manly 2001) may improve test power, but we conclude that CAPTURE tests will fail to detect heterogeneity with most bear data sets.

The assessment of heterogeneity as a function of study design is challenging because there are a variety of plausible relationships between cell area, grid area, and heterogeneity variation. First, larger cell areas should create greater heterogeneity by reducing coverage and subsequently reducing trap encounter rates by females with smaller home ranges. Simulation studies suggest this effect will be reduced by moving hair snags for each session in larger cell designs (Boulanger, unpublished data). Unfortunately, the possible increase in heterogeneity with increased cell area is difficult to detect because power to detect heterogeneity decreases with the reduced capture probabilities that accompany larger cell areas. Second, heterogeneity may decrease with increasing grid area due to lesser edge effects or edge bears only encountering a small subset of traps compared to resident bears (Kendall 1999, Boulanger and McLellan 2001). Grid area and cell area are directly correlated and confounded (Fig. 1) in this study, so it is impossible to separate these effects. Therefore, a full description of potential heterogeneity as a function of grid and cell area is not possible using these results.

Other potential sources of heterogeneity besides those caused by study design include sex and age-based variation and past physical capture history of bears (Boulanger 2001). Results of the Pradel analysis suggest that there is minimal difference between capture probabilities of sexes after accounting for fidelity and immigration to grid areas. This result is contrary to previous studies (Mace et al. 1994). However, the expected variation in capture probabilities of bears probably occurs as an interaction of age and sex. For example, adult females may show higher capture probabilities than male cubs or yearlings. Because age is not identifiable from DNA, comparing sexes alone does not provide a clear understanding of heterogeneity in capture probability variation. For this reason, estimators that model unidentifiable heterogeneity (i.e., age-based heterogeneity), such as the mixture M_h models or M_h (Chao) models, should be considered with bear data.

The problem of low power to detect capture probability variation does not negate the use of mark-recapture models for estimation of population size. It does, however, necessitate a cautious approach to model selection that incorporates the biology of bears and the results of other studies. This approach should involve simulation studies of likely forms of capture probability variation to

evaluate the performance of various models with each unique data set. Model performance is difficult to assess when data are sparse, and therefore simulation evaluation is an essential part of any analysis. Programs CAPTURE and MARK have built-in simulation modules that allow assessment of model performance with various forms of capture probability variation.

The Importance of Study Design

The Pradel analysis demonstrates that closure violation is possible regardless of grid area, and the degree of closure bias varies as a function of grid area, local topography, and other factors. Smaller grids risk a greater degree of closure violation, and therefore designs with smaller grids should only be considered if the area is topographically closed or if radiocollars are used to index bear movements. Studies must also be designed to provide high enough recapture rates to allow adequate modeling of individual heterogeneity variation. For mark-recapture estimation, the recapture rate in the population is as important or more important than the actual number of bears sampled. The optimal design of a study will depend on the expected number of bears in the study area and the degree in which the grid can be closed topographically. For example, the Jumbo and Upper Columbia 1997 (25-km² grid cell size) and Kingcome (49-km² cell size) achieved recapture rates of ≥ 0.2 and adequate precision presumably because the grid areas were reasonably closed by topography (Table 2, Fig. 5). In contrast, Upper Columbia River 1998, also with a 25-km² grid cell size, exhibited low recapture rates and lower precision because topographic closure was not met and capture probabilities were reduced (Boulanger 2001). Due to the nature of the area, the Prophet River grid area could not be closed topographically. In that area, Poole et al. (2001) used large cell sizes (81 km²), which reduced recapture rates; however, this reduction was compensated by the large population in the grid being sampled, and, using 5 sampling sessions, adequate levels of precision were achieved. In conclusion, both the larger cell size Prophet grid and smaller Jumbo, Upper Columbia River, and Kingcome grids differed, but, due to the different nature of the areas, all studies were successful in terms of obtaining precise estimates. Programs CAPTURE and MARK provide further simulation modules which are excellent tools for designing studies. We suggest that when planning studies, biologists use the estimated capture probability and study design parameters from this study to customize simulations to particular study designs.

Comparison of New M_h Mixture with CAPTURE Models

Comparison of the precision of MARK mixture mod-

els and CAPTURE estimates suggests that the mixture models are more precise than CAPTURE in most cases. This increase in precision is likely due to the meta-analysis method of data analysis and subsequent pooling from many projects, as well as the increased efficiency of the mixture model over the Chao heterogeneity model (Chao 1989). The mixture models were developed recently and need more testing to verify their efficiency and robustness. In particular, simulation studies need to address the minimum capture probability levels required for mixture models to discern multiple capture probability distributions. In addition, in some studies (Jumbo, Granby Kettle, Prophet, and Upper Columbia River 1998) time variation was present, and therefore the non-time based M_h mixture models might be biased. The consideration of time variation in terms of the meta-analysis is problematic due to arbitrarily pooling time-specific parameters between projects. A random-effects time-variation model as proposed by Pledger (2000) that is not implemented in MARK is one potential method to incorporate time variation into a meta-analysis. Finally, an assumption of the meta-analysis is that variation in mixture probabilities and individual group probabilities are explained adequately by cell area covariates. In particular, simulation studies need to address whether decreases in mixture probabilities with larger cells are due to decreased heterogeneity or are an artifact of reduced capture probabilities with larger cell areas and subsequent reduced power for the mixture model to discern multiple capture probability distributions. If power to discern multiple capture probability distributions is low at larger cell sizes, then both coefficient of variation of capture probability and associated population estimates will be biased low for the larger cell area projects. Assumptions regarding the relationship between cell area and mixture probability are somewhat relaxed by averaging results over various models, each of which have distinct assumptions regarding the relationship between cell and mixture model parameters. These results do suggest the general meta-analysis strategy is a potentially powerful tool to confront issues with sparse data (see below).

Genetic Issues

The data used for these analyses are the result of scrutiny by many individuals in repeated efforts since original field sampling. The genetic analysis of large numbers of samples over the past 5 years has greatly increased lab efficiency and decreased error rate; however, there still may be errors in identification. In addition, declaring individuals using the P_{sib} equation of Woods et al. (1999) is conservative but may still be biased in certain circumstances (Mills et al. 2000, Waits and Leberg 2000, Waits et al. 2001). Genetic identification errors remain a distinct issue if due care is not given in terms of number and

variability of loci sampled of the target population. In this study, we had the luxury of a known population of bears to reference our procedures against (D. Paetkau, Wildlife Genetics International, Nelson, BC, Canada) and therefore could minimize genetic errors.

Future methods may directly incorporate genetic uncertainty into parameter estimates (Schwarz and Stobo 1999). In particular, methods need to be developed that incorporate the probability of each type of error (i.e., classifying 2 or more animals as 1, and classifying 1 animal as 2 or more) into the capture-recapture model to ensure that unbiased population estimates with appropriate measures of precision are produced. The problem of unidentified gender will probably be overcome when more exact DNA primers for bears are developed. Current sex tests use primers developed for other species (Ennis and Gallagher 1994) and do not provide the highest resolution possible with bears.

New Methods to Confront Sparse Data

Program MARK introduces a new method to select models and incorporate uncertainty in model selection into final parameter estimates. The AIC approach to select a model acknowledges that more than one model might explain the data, especially if data are sparse. With model averaging, estimates of each model are considered and given weight based on how well the model is supported by the data. As exemplified in the Pradel analysis in this paper, robust inference into biological trends can be obtained by interpreting model-averaged parameter estimates and associated errors. This approach is more robust than past model selection routines, such as in program CAPTURE that select one model for a data set, or hypothesis testing approaches to model selection (Burnham and Anderson 1998, Johnson 1999).

An additional advantage of using program MARK to estimate population parameters is that sparse data can be pooled across times, areas, or other categories to achieve more robust estimates of parameters such as capture prob-

abilities. This can then result in more precise estimates of the parameters of interest (e.g., population size) with little cost in bias. For example, consider the estimation of population size with model M_0 for the 7 study areas using program CAPTURE. If each area combination were analyzed separately, 7 capture probability parameters must be estimated. In contrast, analyses can be conducted in MARK to examine whether 7 area-specific capture probabilities are needed, or whether differences in capture probabilities can be explained by cell area, or whether capture probabilities from all projects can be pooled. Further, MARK provides model selection criteria to select among the models used. Although Stanley and Burnham (1998) suggested that AIC did not perform well for the closed capture models, performance was enhanced when AIC criteria were used to average model estimates.

The meta-analysis approach to estimation using program MARK can potentially reduce problems with sparse data. If studies are conducted using similar methods so that the degree of capture probability variation is similar between projects, it is possible to pool data to estimate capture probability. In this study, many of the projects used different designs that compromised the overall efficiency of the meta-analysis based estimates. *A priori* standardization of methods is clearly an optimal step if a meta-analysis approach is to be effective.

Alternatives to Estimation of Population Size

Recently, a variety of flexible models have become available to estimate population trend and apparent survival. These quantities are less affected by closure violation and capture probability variation that challenges the estimation of population size (Table 6). These models do not attempt to estimate population size, but instead track the fates of marked animals within the population. The basic model is the Cormack-Jolly-Seber model (Seber 1992), which estimates capture probability and apparent survival (Lebreton et al. 1992). The Pradel model (Pradel

Table 6. MARK models for estimation of trend and other demographic parameters.

Model name	Parameters	References
Cormack-Jolly-Seber	ϕ — apparent survival (deaths and emigration) p — recapture rate of marked bears	Cormack 1964 Seber 1986 Lebreton et al. 1992
Pradel (1996)	ϕ — apparent survival (deaths and emigration) f — recruitment (births and immigration) λ — population rate of change p — recapture rate	Pradel 1996 Franklin 2002
Barker joint models	S — true survival (probability animal is alive) r — probability animal dies and is found (identified) R — probability animal is resighted alive F — probability that animal is still in sampling area	Barker 1997, 1999 Barker and White 2002
Multi-strata models	S — apparent survival (probability animal is alive and remains in one of the strata) Ψ — probability of movement between areas (or strata) p — recapture rate of marked animals	Brownie et al. 1993 Hestbeck et al. 1991 Powell et al. 2000

1996), used in this paper is a generalization of the Cormack-Jolly-Seber model that estimates recruitment and population rate of change (λ) over time (Pradel 1996, Franklin 2002). The Barker joint live and dead encounter model combines traditional capture data (i.e., DNA identification) with sighting data (via radio telemetry, hunter harvest, or other methods) and can estimate true survival if model assumptions are met (Barker 1997, 1999; Barker and White 2002). Multi-strata models (Hestbeck et al. 1991, Brownie et al. 1993) allow survival estimates and estimates of movements between 2 or more areas (each of the areas must be sampled).

The combined use of DNA methods with other methods, such as radio telemetry, holds promise to mitigate issues with both techniques. Historically, radio telemetry studies that attempted to estimate density were compromised by financial, logistic, and intrusiveness constraints of radiocollaring and tracking large numbers of bears. Therefore, some study areas were small to minimize the constraints of tracking many animals over larger areas. This resulted in a negative relationship between study area size and estimated density of many carnivores (Smallwood and Schonewald 1996). DNA sampling in conjunction with radio telemetry or other data sources allows larger areas to be sampled and a reduction of scale issues. For example the joint mark-recapture methods of Barker (1997, 1999) and Barker and White (2002) allows information from harvest, sightings, radio telemetry, and DNA sampling to be combined for estimates of survival. The multi-strata method of Powell et al. (2000) uses mark-recapture sampling and radio telemetry to estimate movements of animals in and out of study areas, and estimates true survival. Their analysis can be done efficiently in program MARK. Radio tracking is combined with live trapping of small mammals to estimate the proportion of the trapped population inhabiting the trapping grid (White and Shenk 2001). The combination of both types of data results in more robust and reliable estimates than either method alone.

The main constraint with models that estimate survival or population rate of change is that they need at least 3 years of data to estimate most parameters, although the robust design model (Kendall et al. 1995, 1997; Kendall 1999) can estimate survival with just 2 years of captures. However, the demography of bears demands a long-term approach to estimate most population parameters and therefore a longer period is needed to obtain results regardless of methods used. Program MARK provides a simulation module that allows assessment of model performance given likely sample sizes and parameter values. Use of simulations to evaluate likely monitoring plans is an essential first step in planning any monitoring program. Finally, successful use of program MARK requires

training in the theoretical background of the estimators used. Online manuals are available for program MARK (Cooch and White 2001) and provide a good start toward understanding these mark-recapture models.

Designing Long Term Monitoring Studies

Long-term monitoring studies should be designed to detect some level of change with a specified power. The degree of change to be detected will depend on the amount of temporal and spatial process variance in population size (i.e., extra variance in population size about a fixed trend effect). Power of the proposed design is affected by the sampling variance of the estimates of population size, the temporal and spatial process variance of the population, and the number of population estimates (Thompson et al. 1998). Previous research and software (Gibbs 1999, 2000; Gerrodette 1987) to evaluate the power of a monitoring program to detect a trend has not considered process variance of the population directly. Computer simulations that directly model process variance are the only valid approach currently available to determine power of a proposed design.

CONCLUSION

Our results highlight the potential advantages and disadvantages of application of DNA mark-recapture methods to bear populations. A few general conclusions can be made from this effort. First, the extreme importance of proper study design cannot be overemphasized. It could be argued that the ultimate success of a project is based upon how well it was designed, and the mark-recapture modeling procedures can only complement what has been done in the field and in the genetics lab. Second, genetics issues such as incorporation of genetic classification error rates into mark-recapture estimates are yet to be formulated and therefore an uncertain bias exists in estimates, especially if samples are of poor quality. Rigorous lab procedure is critical. Third, of the issues with population estimation using DNA mark-recapture, the most difficult to estimate is the effect of closure violation on population estimates if radiocollared animals are not used in sampling efforts. Fourth, an area of great promise is the combination of radiocollaring, DNA methods, and other forms of information to monitor and study bear populations using the newer mark-recapture models available in program MARK.

ACKNOWLEDGMENTS

The research efforts detailed in this paper were the collaborative effort of many individuals and funding agencies. K. Poole (Aurora Wildlife Research) and G. Mowat

(Aurora Wildlife Research) designed and collected data for the Prophet River data set used in this study (Poole et al. 2001). K. Strom (Axys Consulting) aided in the interpretation of the Jumbo Pass Project. D. Paetkau (Wildlife Genetics International) and C. Strobeck (University of Alberta) were the principal geneticists for all projects. J. Bonneville, M. Paradon, and K. Stalker assisted in the genetic analysis. A.R. Hamilton (Ministry of Environment, Lands, and Parks, British Columbia) orchestrated funding for this project, Forest Renewal British Columbia; Ministry of Environment, Lands, and Parks (Wildlife Branch); Parks Canada; British Columbia Ministry of Forests; Columbia Basin Fish and Wildlife Compensation Program; National Science and Engineering Research Council of Canada; British Columbia Environmental Assessment Office; Canadian Forest Products; Pope and Talbot Ltd; Western Forest Products; Prophet Indian Band; University of Alberta; University of Calgary; University of British Columbia; the Colorado Division of Wildlife; Friends of Mount Revelstoke and Glacier National Parks; and the Southern British Columbia Guides and Guide Outfitters Association. We also thank the field technicians, helicopter and fixed-wing pilots, and laboratory technicians, who gathered and analyzed field specimens and data. J. Nichols, L.S. Mills, R. Harris, G. Mowat, and 1 anonymous reviewer provided useful comments on earlier versions of this manuscript.

LITERATURE CITED

- AMSTRUP, S.C., T.L. McDONALD, AND I. STERLING. 2001. Polar bears in the Beaufort Sea: A 30-year mark-recapture case history. *Journal of Agricultural, Biological, and Environmental Statistics* 6:221-234.
- BARKER, R.J. 1997. Joint modeling of live-recapture, tag-resight, and tag-recovery data. *Biometrics* 53:666-677.
- . 1999. Joint analysis of mark-recapture, resighting and ring-recovery data with age-dependence and marking-effect. *Bird Study* 46(Supplement):82-91.
- , AND G.C. WHITE. 2002. Joint analysis of live and dead encounter of marked animals. Pages 361-367 R. Fields, editor. *Integrating people and wildlife for a sustainable future: Proceedings of the Second International Wildlife Management Congress, Gödöllő, Hungary*. The Wildlife Society, Bethesda, Maryland, USA.
- BOULANGER, J. 2000. Granby-Kettle/Boundary Forest District 1997 grizzly bear DNA mark-recapture inventory project: Statistical analysis and population estimates. Ministry of Environment, Lands, and Parks, Penticton, British Columbia, Canada.
- . 2001. Estimates of the grizzly bear (*Ursus arctos*) population size using DNA based mark-recapture, Upper Columbia River, British Columbia, Canada 1996-1998. Parks Canada, Revelstoke, British Columbia, Canada.
- , AND S. HIMMER. 2000. Kingcome (1997) DNA mark-recapture grizzly bear inventory project final report. Ministry of Environment, Lands, and Parks, Nanaimo, British Columbia, Canada.
- , AND B. McLELLAN. 2001. Closure violation in DNA-based mark-recapture estimation of grizzly bear populations. *Canadian Journal of Zoology* 79:642-651.
- BROWNE C., J.E. HINES, J.D. NICHOLS, K.H. POLLOCK, AND J.B. HESTBECK. 1993. Capture-recapture studies for multiple strata including non-markovian transitions. *Biometrics* 49:1173-1187.
- BURNHAM, K.P., AND D.R. ANDERSON. 1998. Model selection and inference: A practical information theoretic approach. Springer-Verlag, New York, New York, USA.
- , ———, G.C. WHITE, C. BROWNE, AND K.H. POLLOCK. 1987. Design and analysis methods for fish survival experiments based on release-recapture. American Fisheries Society, Bethesda, Maryland, USA.
- CAROTHERS, A.D. 1973. The effects of unequal catchability on Jolly-Seber estimates. *Biometrics* 29:79-100.
- CHAO, A. 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.
- , AND S.L. JENG. 1992. Estimating population size for capture-recapture data when capture probabilities vary by time and individual animal. *Biometrics* 48:201-216.
- COOCH, E., AND G.C. WHITE. 2001. Analysis of encounter data from marked animal populations, Program MARK: A gentle introduction. Cornell University and Colorado State University, Ithaca, New York, and Fort Collins, Colorado, USA.
- CORMACK, R.M. 1964. Estimates of survival from the sighting of marked animals. *Biometrika* 51:429-438.
- ENNIS, S., AND T.F. GALLAGHER. 1994. PCR based sex determination assay in cattle based on the bovine Amelogenin locus. *Animal Genetics* 25:425-427.
- FRANKLIN, A.B. 2002. Exploring ecological relationships in survival and estimating rates of population change using program MARK. Pages 350-356 in R. Fields, editor. *Integrating people and wildlife for a sustainable future: Proceedings of the Second International Wildlife Management Congress, Gödöllő, Hungary*. Wildlife Society, Bethesda, Maryland, USA.
- GAGNEUX, P., C. BOESCH, AND D.S. WOODRUFF. 1997. Microsatellite scoring errors associated with non-invasive genotyping based on nuclear DNA amplified from shed hair. *Molecular Ecology* 6:861-868.
- GERRODETTE, T. 1987. A power analysis for detecting trends. *Ecology* 68:1364-1372.
- GIBBS, J.P. 1999. Effective monitoring for adaptive wildlife management: lessons from the Galápagos Islands. *Journal of Wildlife Management* 63:1055-1065.
- . 2000. Monitoring populations. Pages 213-252 in L. Boitani and T.K. Fuller, editors. *Research techniques in animal ecology: Controversies and consequences*. Columbia University Press, New York, New York, USA.
- GILBERT, R.O. 1973. Approximations of the bias in the Jolly-Seber capture-recapture model. *Biometrics* 29:501-526.

- GOOSENS, B., L.P. WAITS, AND P. TABERLET. 1998. Plucked hair samples as source of DNA: reliability of dinucleotide microsatellite genotyping. *Molecular Ecology* 7:1237-1241.
- HESTBECK, J.B., J.D. NICHOLS, AND R.A. MALECKI. 1991. Estimates of movement and site fidelity using mark-resight data of wintering Canada geese. *Ecology* 72:523-533.
- HUGGINS, R.M. 1991. Some practical aspects of a conditional likelihood approach to capture experiments. *Biometrics* 47:725-732.
- JOHNSON, D.H. 1999. The insignificance of statistical significance testing. *Journal of Wildlife Management* 63:763-772.
- KENDALL, W.L. 1999. Robustness of closed capture-recapture methods to violations of the closure assumption. *Ecology* 80:2517-2525.
- , J.D. NICHOLS, AND J.E. HINES. 1997. Estimating temporary emigration using capture-recapture data with Pollock's robust design. *Ecology* 78:563-578.
- , K.H. POLLOCK, AND C. BROWNIE. 1995. A likelihood-based approach to capture-recapture estimation of demographic parameters under the robust design. *Biometrics* 51:293-308.
- LEBRETON, J.D., K.P. BURNHAM, J. CLOBERT, AND D.R. ANDERSON. 1992. Modeling survival and testing biological hypotheses using marked animals: a unified approach with case studies. *Ecological Monographs* 62:67-118.
- MACE, R.D., S.C. MINTA, T.L. MANLEY, AND K.E. AUNE. 1994. Estimating grizzly bear population size using camera sightings. *Wildlife Society Bulletin* 22:74-83.
- MACKENZIE, D.I., AND B.J. MANLY. 2001. Randomization tests for time effects and heterogeneity in capture probabilities for closed populations. *Journal of Agricultural, Environmental, and Biological Statistics* 6:292-301.
- MCDONALD, T.L. 2001. Estimation of population size using open mark-recapture models. *Journal of Agricultural, Biological, and Environmental Statistics* 6:206-220.
- MCLELLAN, B.N. 1989. Dynamics of a grizzly bear population during a period of industrial resource extraction. I. Density and age-sex composition. *Canadian Journal Zoology* 67:1857-1868.
- , F. HOVEY, R.D. MACE, J. WOODS, D. CARNEY, M. GIBEAU, W. WAKKINEN, AND W. KASWORM. 1999. Rates and causes of grizzly bear mortality in the interior mountains of British Columbia, Alberta, Washington, and Idaho. *Journal of Wildlife Management* 63:911-920.
- MILLER, S.D., G.C. WHITE, R.A. SELLERS, H.V. REYNOLDS, J.W. SCHOEN, K. TITUS, V.G.J. BARNES, R.B. SMITH, R.R. NELSON, W.W. BALLARD, AND C.C. SCHWARTZ. 1997. Brown and black bear density estimation in Alaska using radiotelemetry and replicated mark-resight techniques. *Wildlife Monographs* 133.
- MILLS, L.S., J.J. CITTA, K.P. LAIR, M.K. SCHWARZ, AND D. TALLMON. 2000. Estimating animal abundance using non-invasive sampling: Promises and pitfalls. *Ecological Applications* 10:283-294.
- 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.
- NORRIS, J.L., AND K.H. POLLOCK. 1996. Nonparametric MLE under two closed capture-recapture models with heterogeneity. *Biometrics* 52:639-649.
- 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.
- PAETKAU, D., W. CALVERT, I. STIRLING, AND C. STROBECK. 1995. Microsatellite analysis of population structure in Canadian polar bears. *Molecular Ecology* 4:347-354.
- , AND C. STROBECK. 1994. Microsatellite analysis of genetic variation in black bear populations. *Molecular Ecology* 3:489-495.
- PLEDGER, S. 2000. Unified maximum likelihood estimates for closed models using mixtures. *Biometrics* 56:434-442.
- , AND M. EFFORD. 1998. Correction of bias due to heterogeneous capture probabilities in capture-recapture studies of open populations. *Biometrics* 54:888-898.
- POLLOCK, K.H., J.D. NICHOLS, C. BROWNIE, AND J.E. HINES. 1990. Statistical inference for capture-recapture experiments. *Wildlife Monographs* 107.
- POOLE, K.G., G. MOWAT, AND D.A. FEAR. 2001. DNA-based population estimate for grizzly bears *Ursus arctos* in northeastern British Columbia, Canada. *Wildlife Biology* 7:105-115.
- POWELL, L.A., M. CONROY, J.E. HINES, AND D.G. KREMENTZ. 2000. Simultaneous use of mark-recapture and radiotelemetry to estimate survival, movement, and capture rates. *Journal of Wildlife Management* 64:302-313.
- PRADEL, R. 1996. Utilization of mark-recapture for the study of recruitment and population growth rate. *Biometrics* 52:703-709.
- REXSTAD, E., AND K. BURNHAM. 1990. User's guide for interactive program CAPTURE. Colorado Cooperative Fish and Wildlife Research Unit, Fort Collins, Colorado, USA.
- ROSENBERG, D.K., W.S. OVERTON, AND R.G. ANTHONY. 1995. Estimation of animal abundance when capture probabilities are low and heterogeneous. *Journal of Wildlife Management* 59:252-261.
- SCHWARZ, C.J., AND W.T. STOBO. 1999. Estimation and effects of tag-misread rates in capture-recapture studies. *Canadian Journal of Fisheries and Aquatic Sciences* 56:551-559.
- SEBER, G.A.F. 1982. The estimation of animal abundance. Charles Griffin and Company, London, U.K.
- . 1986. A review of estimating animal abundance. *Biometrics* 42:267-292.
- . 1992. A review of estimating animal abundance II. *International Statistical Review* 60:129-166.
- SMALLWOOD, K.S., AND C. SCHONEWALD. 1996. Scaling population density and spatial pattern for terrestrial carnivores. *Oecologia* 105:329-335.
- STANLEY, T.R., AND K.P. BURNHAM. 1998. Estimator selection for closed-population capture-recapture. *Journal of Agricultural, Biological, and Environmental Statistics* 3:131-150.
- , AND ———. 1999. A closure test for time specific capture-recapture data. *Environmental and Ecological Statistics* 6:197-209.
- STROM, K., M. PROCTOR, AND J. BOULANGER. 1999. Grizzly bear population survey in the Central Purcell Mountains, British

- Columbia. Axys Consulting, Calgary, Alberta, University of Calgary, Calgary, Alberta, and Integrated Ecological Research, Nelson, British Columbia, Canada.
- TABERLET, P., S. GRIFFIN, B. GOOSSENS, S. QUESTIAU, V. MANCEAU, N. ESCARAVAGE, L.P. WAITS, AND J. BOUVET. 1996. Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Research* 24:3189–3194.
- , H. MATTOCK, C. BUBOIS-PAGANON, AND J. BOUVET. 1993. Sexing free-ranging brown bears (*Ursus arctos*) using hairs found in the field. *Molecular Ecology* 2:399–403.
- , L.P. WAITS, AND G. LUIKART. 1999. Noninvasive genetic sampling: look before you leap. *Trends in Evolution and Ecology* 14:323–327.
- THOMPSON, W.L., G.C. WHITE, AND C. GOWAN. 1998. Monitoring vertebrate populations. Academic Press, San Diego, California, USA.
- WAITS, J.L., AND P.L. LEBERG. 2000. Biases associated with population estimation using molecular tagging. *Animal Conservation* 3:191–199.
- , G. LUIKART, AND P. TABERLET. 2001. Estimating the probability of identity among genotypes in natural populations: cautions and guidelines. *Molecular Ecology* 10:249–256.
- WHITE, G.C. 1996. NOREMARK: population estimation from mark–resighting surveys. *Wildlife Society Bulletin* 24:50–52.
- . 2002. Discussant: The use of auxiliary variables in capture–recapture modeling: an overview. *Journal of Applied Statistics* 29:103–106.
- , D.R. ANDERSON, K.P. BURNHAM, AND D.L. OTIS. 1982. Capture–recapture and removal methods for sampling closed populations. Los Alamos National Laboratory, Los Alamos, New Mexico, USA.
- , AND K.P. BURNHAM. 1999. Program MARK: Survival estimation from populations of marked animals. *Bird Study* 46(Supplement):120–138.
- , ———, AND D.R. ANDERSON. 2002. Advanced features of program MARK Pages 368–377 in R. Fields, editor. Integrating people and wildlife for a sustainable future: Proceedings of the Second International Wildlife Management Congress, Gödöllő, Hungary The Wildlife Society, Bethesda, Maryland, USA.
- , AND T.M. SHENK. 2001. Population estimation with radio marked animals. Pages 329–350 in J.J. Millspaugh and J.M. Marzluff, editors. Design and analysis of radio telemetry studies. Academic Press, San Diego, California, USA.
- WILSON, K.R., AND D.R. ANDERSON. 1985a. Evaluation of a nested grid approach for estimating density. *Journal of Wildlife Management* 49:675–678.
- , AND ———. 1985b. Evaluation of two density estimators of small mammal population size. *Journal of Mammalogy* 66:13–21.
- WOODS, J.G., D. PAETKAU, D. LEWIS, B.N. McLELLAN, M. PROCTOR, AND C. STROBECK. 1999. Genetic tagging free-ranging black and brown bears. *Wildlife Society Bulletin* 27:616–627.

Received: 28 May 2001.

Accepted: 7 January 2002.

Associate Editor: Mills.