

# Sub-sampling genetic data to estimate black bear population size: a case study

Catherine A. Tredick<sup>1,4</sup>, Michael R. Vaughan<sup>2</sup>, Dean F. Stauffer<sup>1</sup>, Stephanie L. Simek<sup>3</sup>, and Thomas Eason<sup>3</sup>

<sup>1</sup>Department of Fisheries and Wildlife Sciences, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA

<sup>2</sup>US Geological Survey, Virginia Cooperative Fish and Wildlife Research Unit, Blacksburg, VA 24061, USA

<sup>3</sup>Florida Fish and Wildlife Conservation Commission, Tallahassee, FL 32399, USA

**Abstract:** Costs for genetic analysis of hair samples collected for individual identification of bears average approximately US\$50 [2004] per sample. This can easily exceed budgetary allowances for large-scale studies or studies of high-density bear populations. We used 2 genetic datasets from 2 areas in the southeastern United States to explore how reducing costs of analysis by sub-sampling affected precision and accuracy of resulting population estimates. We used several sub-sampling scenarios to create subsets of the full datasets and compared summary statistics, population estimates, and precision of estimates generated from these subsets to estimates generated from the complete datasets. Our results suggested that bias and precision of estimates improved as the proportion of total samples used increased, and heterogeneity models (e.g.,  $M_{H(CHAO)}$ ) were more robust to reduced sample sizes than other models (e.g., behavior models). We recommend that only high-quality samples (>5 hair follicles) be used when budgets are constrained, and efforts should be made to maximize capture and recapture rates in the field.

**Key words:** American black bear, budget constraints, noninvasive genetic sampling, population estimates, sub-sampling, *Ursus americanus*

*Ursus* 18(2):179–188 (2007)

Noninvasive genetic sampling is widely used for capture–mark–recapture (CMR) studies of black (*Ursus americanus*) and brown (*U. arctos*) bears (Woods et al. 1999, Mowat and Strobeck 2000, Poole et al. 2001). This method has numerous advantages over traditional CMR techniques (e.g., live-trapping and deploying radiocollars), including less intense field efforts, increased capture probabilities and sample sizes, and the ability to examine populations without handling animals (Mills et al. 2000, Mowat and Strobeck 2000, Waits and Leberg 2000). As these techniques developed, problems and sources of error were identified, particularly with regard to laboratory analysis of the genetic material collected (usually hair). Genotyping errors, for example, may identify too few or extra false individuals, biasing final population estimates (see Taberlet et al. 1999, Waits and Leberg 2000, and Mills et al. 2000 for discussion of allelic dropout and the shadow effect). Discussion of these problems and potential solutions have been widespread (Taberlet

et al. 1996, Mills et al. 2000, Waits et al. 2001, Miller et al. 2002, Paetkau 2003, Creel et al. 2003, McKelvey and Schwartz 2004, Roon et al. 2005).

In contrast, discussion of cost-effectiveness and financial feasibility of large-scale noninvasive studies has been scarce. Although they have not been compared in the literature, genetic techniques in general have been touted as more cost-effective and less labor intensive than traditional CMR approaches. This may not be the case, however, for large-scale noninvasive studies of high-density bear populations. While equipment and labor costs in the field may be small compared to more invasive CMR studies, costs for genetic analysis of samples can be considerable, particularly when many samples are collected. Researchers conducting large-scale noninvasive sampling or studying large, high-density bear populations, particularly in the southeastern US, collect hundreds or thousands of samples in one season (T. Eason, Florida Fish and Wildlife Conservation Commission, Tallahassee, Florida, USA; F. van Manen, US Geological Survey Southern

<sup>4</sup>ctredick@vt.edu

Appalachian Field Branch, Knoxville, Tennessee, USA; J. Clark, University of Tennessee, Knoxville, Tennessee, USA; personal communications, 2004). In most cases, it is not financially feasible or statistically necessary to analyze all samples, leaving the question of how to sub-sample data to meet financial constraints while maintaining accuracy and precision of CMR estimates. Thus, our objectives with this paper were to use field data to explore the effects of random and nonrandom sub-sampling of complete genetic datasets on population size estimates. We used the model selection routines generated in Program CAPTURE (White et al. 1982) and present scenarios that minimize costs while maintaining accuracy and precision acceptable to biologists or managers.

## Study areas

The 50-km<sup>2</sup> Pungo Unit of Pocosin Lakes National Wildlife Refuge (PLNWR) in northeastern North Carolina is characterized by low-lying pocosin wetlands interspersed with agricultural fields. This matrix of forest and agriculture supports a high bear density (L.E. Hinesley, 2000, Pocosin Lakes National Wildlife Refuge: forest habitat management plan, US Fish and Wildlife Service, Columbia, NC, USA). The St. Johns area, in northeastern Florida, supports low to moderate bear numbers in fragmented habitat of marginal quality (T. Eason, personal communication, 2004). It is 20 times larger than the Pungo Unit at 967 km<sup>2</sup> and also is characterized by low-lying wetland forests.

## Methods

### Hair collection

We collected black bear hair samples from barbed-wire hair traps as described by Woods et al. (1999). A single-strand enclosure was used at PLNWR. A second strand of wire, placed approximately 25 cm above the ground, was used in the St. Johns study area to improve sample quality on the upper wire. We baited enclosures with food (bakery products and corn) suspended 2–3 m above the ground and activated them for 8 weeks during the summer (Jun–Aug 2001 for St. Johns, Jun–Aug 2002 for PLNWR). We collected hair and rebaited traps approximately every 7 days during the 8-week sampling season. We based trap density on average female home range sizes from each area to ensure

that bears had access to at least 1 sampling site (Otis et al. 1978). Home range sizes were substantially larger in St. Johns than in PLNWR (approximately 24 km<sup>2</sup> versus 2 km<sup>2</sup>, respectively: Hellgren 1988, Allen 1999, McCown et al. 2001), yielding less dense trap spacing in St. Johns. We established 51 traps in St. Johns (~1 trap/19 km<sup>2</sup>) and 33 traps on the Pungo Unit (~1 trap/1.5 km<sup>2</sup>).

### Genetic analysis

Each sample collected during the sampling seasons at the 2 study sites was analyzed at the Wildlife Genetics International (WGI) laboratory in Nelson, British Columbia, Canada. Six microsatellite markers were used to identify individual bears (St. Johns: G1A, G10B, G1D, G10H, MU50, MU59; PLNWR: G1A, G1D, G10H, G10J, G10L, MU50; Paetkau and Strobeck 1994, Paetkau et al. 1995), and extraction and analysis followed standard laboratory methods (PCR, electrophoresis: Paetkau and Strobeck 1994, Paetkau et al. 1995, Paetkau 2003).

Strict error-checking protocols (see Paetkau 2003) were followed to minimize scoring and amplification errors in the lab. Although we assumed hair samples with identical genotypes came from the same individual, it was possible for  $\geq 2$  bears to share identical observed genotypes (Woods et al. 1999). We estimated the probability of this occurring by calculating the probability of identity (PI: Paetkau and Strobeck 1994) and a more conservative estimate of this probability,  $PI_{SIBS}$ , which accounts for populations containing many siblings or closely related individuals and thus represents the upper limit for the theoretical range of PI values (Taberlet and Luikart 1999). Any genotype with  $PI_{SIBS} > 0.01$  was excluded from analyses. All pairs of genotypes that were different at only 1, 2, or 3 markers were scrutinized to ensure the number of individuals was not overestimated due to allelic dropout (Taberlet et al. 1996, Gagneux et al. 1997). Pairs of samples were reanalyzed at the markers where they differed until differences were confirmed or allelic dropout was confirmed and corrected (Paetkau 2003).

### Population estimation and sub-sampling scenarios

We generated population estimates from the complete datasets for each area (all samples collected during the 8-week seasons) using CAPTURE (White

**Table 1. List of scenarios used for sub-sampling hair samples of black bears from Pocosin Lakes National Wildlife Refuge, North Carolina in 2002 and St. Johns, Florida in 2001.**

Scenario	Description
1A	1 random sample/trap/period
1B	2 random samples/trap/period
1C	3 random samples/trap/period
2A	Randomly selected 40% of samples
2B	Randomly selected 50% of samples
2C	Randomly selected 60% of samples
2D	Randomly selected 70% of samples
2E	Randomly selected 80% of samples
3	1 sample/trap from 30% of traps visited each period
Scenario 4A	All samples from period 1 + 1 random sample/trap/periods 2–8
Scenario 4B	All samples from periods 1 and 2 + 1 random sample/trap/periods 3–8
Scenario 4C	All samples from periods 1, 2, and 3 + 1 random sample/trap/periods 4–8
Scenario 4D	All samples from period 1 + 2 random samples/trap/periods 2–8
Scenario 4E	All samples from periods 1 and 2 + 2 random samples/trap/periods 3–8
Scenario 4F	All samples from periods 1, 2, and 3 + 2 random samples/trap/periods 4–8

et al. 1982). These estimates formed the basis for comparison to subsequent sub-sampled datasets (i.e., estimates using the complete datasets were taken to be the best estimates of true population size, hereafter referred to as the best estimates).

We used 4 scenarios to evaluate the effect of sub-sampling from our datasets (Table 1). In scenario 1, we randomly selected a predetermined number of samples (1, 2, or 3) from each trap for each sampling period. For scenario 2, we randomly selected a percent of the total samples (40, 50, 60, 70, or 80%) regardless of the trap or period in which we collected them. Scenario 3 simulated constrained financial resources, limiting not only the number of samples taken from each trap but the number of traps sampled. For this scenario, approximately one-third of the total traps visited by bears in each period were randomly selected, and, for each sampling period, 1 sample from each of these traps was randomly selected. This scenario was not carried out for the St. Johns dataset because <30% of the traps were visited in each period. Scenario 4 was based on the idea that new captures generally decline in later sampling periods in closed populations (White et al. 1982), and it may be more cost effective to select more samples from earlier periods for analysis and fewer from later periods. For this scenario, all samples from earlier periods (period 1; periods 1 and 2; or periods 1, 2, and 3) were sampled, along with a random number of samples/trap from subsequent periods (either 2–8, 3–8, or 4–8: Table 1). We used SAS (SAS Institute, Inc. 2001) to generate 100 random data subsets (replicates) for each scenario from each study area.

We summarized CAPTURE (White et al. 1982) output for each replicate in terms of number of unique captures ( $M_{t+1}$ ), top model selected, population estimates, coefficients of variation (CV), and 95% confidence intervals (CI). We averaged these statistics for all 100 replicates for each scenario and compared data from each scenario to data from the complete datasets and all other scenarios. We initially used the best model selected for each replicate for averages of each scenario, but to avoid variability introduced by model differences, we used the same model for all replicates and all scenarios instead (i.e., the best model selected for the complete datasets in each area,  $M_{b|(\text{REMOVAL})}$ ; Otis et al. 1978). Because behavior models have performed poorly (Otis et al. 1978), particularly with sparse data, we used a more robust heterogeneity model ( $M_{h|\text{CHAO}}$ ; Chao 1987) for all replicates as well. Accuracy and precision of each scenario was assessed using percent relative bias (PRB relative to our best estimates;  $(N - \hat{N}/N) \times 100$ ) and CI coverage (% of replicates whose 95% CI contained the best estimate from the complete datasets). Cost estimates for each scenario were based on an average 2004 price of US\$50/sample (D. Paetkau, Wildlife Genetics International, Nelson, BC, Canada; L. Waits, University of Idaho, Moscow, Idaho, USA, personal communications, 2004). We based recommendations for sub-sampling noninvasive CMR datasets on these results and total cost of various scenarios.

#### **Sample quality and adjacent barbs**

In addition to the random sub-sampling scenarios, we assessed the success rate of lower quality samples

**Table 2. Summary statistics (complete datasets) for black bear hair captured in Pocosin Lakes National Wildlife Refuge (PLNWR), North Carolina, in 2002 and St. Johns, Florida, 2001. Samples were individually identified using microsatellite DNA analysis. Data were analyzed using CAPTURE (White et al. 1982).**

	Traps hit (%)	Samples/trap	High quality samples <sup>a</sup>	Successful samples <sup>b</sup>	Individual bears captured	Redundant samples <sup>c</sup>	Captures/bear	Model selected	Capture rate ( $M_{bh}$ )	Recapture rate ( $M_{bh}$ )	Capture rate ( $M_{h[CHAO]}$ )
PLNWR	97	2.9	468	398	85	129	3.15	$M_{tbbh}^d$	0.39	0.25	0.20
St. Johns	43	3.2	179	157	25	93	2.56	$M_{bh}$	0.11	0.31	0.18

<sup>a</sup>High quality samples contain  $\geq 5$  hair follicles.

<sup>b</sup>Samples that produced complete, 6-loci genotypes in the lab.

<sup>c</sup>Samples from the same bear at the same trap in the same period.

<sup>d</sup>No estimate for this model in CAPTURE, so the next best model [ $M_{bh[REMOVAL]}$ ] was used.

(with  $< 5$  hairs) and samples from the lower wire at St. Johns by determining what proportion of these samples produced complete genotypes. We also examined samples that were directly adjacent and 2 barbs apart to determine the proportion of these samples that were from the same bear and represented redundant analyses.

## Results

### Hair collection and genetic analysis

We collected 468 hair samples from 32 traps at the Pungo Unit of PLNWR and 383 samples from 25 traps at St. Johns (Table 2). Due to the high number of samples at traps, we did not collect low quality samples ( $< 5$  hairs) at PLNWR, so for purposes of comparison, we also removed low quality samples from the St. Johns dataset for most of our sub-sampling analyses. This resulted in 179 high quality samples ( $\geq 5$  hairs) in the St. Johns dataset. For the PLNWR data, bears visited 97% of the traps and deposited an average of 2.9 (SE = 0.006) samples/trap/period. Eighty-five individual bears were identified from 398 samples that produced complete genotypes in the lab (71 samples [15.2%] failed to produce complete genotypes). One-hundred twenty-nine of these samples were redundant, coming from the same bear at the same trap in the same sampling period. Each bear was captured an average of 3.15 times (Table 2). At St. Johns, bears left hair (low and high quality samples) at 49% of the traps, yielding an average of 5.2 (SE = 0.59) hair samples from each visited trap/period. When low quality samples were removed, only 43% of traps yielded samples, for an average of 3.2 (SE = 0.46) samples/trap/period. From 157 high quality samples that produced complete genotypes at St. Johns (22 samples [12.3%] failed to produce complete genotypes), 25

individual bears were identified in the lab; 6 additional bears were identified from 113 low quality samples that produced complete genotypes at St. Johns. Ninety-three of the high quality samples that produced complete genotypes in the lab were redundant. Each bear was captured an average of 2.56 times (Table 2).

All but 1 pair of genotypes differed at  $> 2$  markers in the PLNWR dataset. Two genotypes differed at only 2 markers, but each was confirmed in 3 independent samples and thus was unlikely to contain errors (Paetkau 2003). For the St. Johns dataset, 2 pairs of genotypes differed at 2 markers and 14 pairs differed at 3, but all differences were confirmed by multiple reanalysis.

### Population estimation for complete datasets

CAPTURE model selection tests showed strong heterogeneity (test 1:  $\chi^2 = 39.34$ , 3 df,  $P < 0.0001$ ), behavioral response (test 2:  $\chi^2 = 10.08$ , 1 df,  $P = 0.0015$ ), and temporal variation (test 3:  $\chi^2 = 29.13$ , 7 df,  $P = 0.0001$ ) in capture probabilities for the complete PLNWR dataset. CAPTURE subsequently selected model  $M_{tbbh}$  (Otis et al. 1978) as the best fitting model for this dataset, which incorporates all 3 forms of capture probability variation. Because an estimator for this model does not exist in CAPTURE, we used the next best fitting model,  $M_{bh[REMOVAL]}$  (discriminant function standardized score = 0.97: Otis et al. 1978) to estimate population size. This estimator generated an estimate of 86 bears with a CV of 1.8%. Capture and recapture rates were 0.39 and 0.25, respectively (Table 2). Because a high degree of heterogeneity was detected in the data, and behavior models tend to perform poorly with sparse data, we also used the more robust model  $M_{h[CHAO]}$  (Chao 1987), which estimated a population of 128 bears (CV = 14.7%, capture rate = 0.20).

**Table 3. Population estimates and cost projections from DNA analysis of bear hair collected at Pocosin Lakes National Wildlife Refuge (PLNWR), North Carolina, and St. Johns, Florida, 2001 and 2002; estimates are given for the complete datasets (best estimates) and for the average of 100 replicates for each subsampling scenario (Table 1) using models  $M_{bh[REMOVAL]}$  and  $M_{h[CHAO]}$ .**

Location	Scenario	Captures ( $M_{t+1}$ )	Samples analyzed <sup>a</sup>	Successful samples <sup>b</sup>	Cost (US\$) <sup>c</sup>	% of total cost	$M_{bh[REMOVAL]}$				$M_{h[CHAO]}$			
							Pop. est. <sup>d</sup>	95% CI <sup>e</sup>	CI coverage <sup>f</sup>	PRB <sup>g</sup>	Pop. est. <sup>d</sup>	95% CI <sup>e</sup>	CI coverage <sup>f</sup>	PRB <sup>g</sup>
PLNWR	1A	61	162	137	\$6,850	29	64	71–100	8	-26	102	79–161	78	-20
	1B	75	276	234	\$11,700	50	76	78–91	61	-12	117	94–172	99	-9
	1C	81	353	299	\$14,950	64	82	82–90	93	-5	125	100–181	100	-2
	2A	63	187	159	\$7,950	34	65	66–81	2	-24	115	86–185	95	-10
	2B	69	234	198	\$9,900	42	71	72–84	15	-17	120	92–184	95	-6
	2C	74	281	238	\$11,900	51	76	76–88	31	-12	122	96–184	97	-5
	2D	77	328	278	\$13,900	59	78	79–88	47	-9	120	97–175	99	-6
	2E	82	374	317	\$15,850	68	82	82–90	88	-5	124	100–179	100	-3
	3	45	96	81	\$4,050	17	53	63–109	28	-38	89	62–164	66	-30
	Complete	85	468	397	\$23,400	100	86	86–93	100	0	128	104–182	100	0
St. Johns	1A	22	56	49	\$2,450	27	34	24–56	68	-19	32	25–65	99	-18
	1B	23	89	78	\$3,900	44	34	26–67	50	-19	35	27–73	100	-10
	1C	24	113	99	\$4,950	55	39	29–82	62	-7	39	29–80	100	0
	2A	19	72	63	\$3,150	35	32	22–51	82	-24	32	23–76	91	-18
	2B	21	90	79	\$3,950	44	34	24–71	72	-19	34	25–76	99	-13
	2C	22	107	94	\$4,700	52	34	24–75	80	-19	33	25–68	100	-15
	2D	23	125	110	\$5,500	61	36	26–77	76	-14	37	27–79	100	-5
	2E	24	143	125	\$6,250	70	37	27–95	83	-12	37	27–78	100	-5
	3 <sup>h</sup>	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	Complete	25	179	157	\$8,950	100	42	28–139	100	0	39	29–80	100	0

<sup>a</sup>Samples sent to lab for analysis.

<sup>b</sup>Number of samples that produced complete genotypes in the lab.

<sup>c</sup>Costs are based on an average 2004 cost/sample of US\$50. Costs of analysis of failed samples excluded.

<sup>d</sup>Pop. est. = estimated population size from CAPTURE.

<sup>e</sup>Based on the average CIs from 100 replicates.

<sup>f</sup>% of 100 replicates that contain the best estimates from complete datasets.

<sup>g</sup>Percent relative bias =  $(N - \hat{N}/N) \times 100$ .

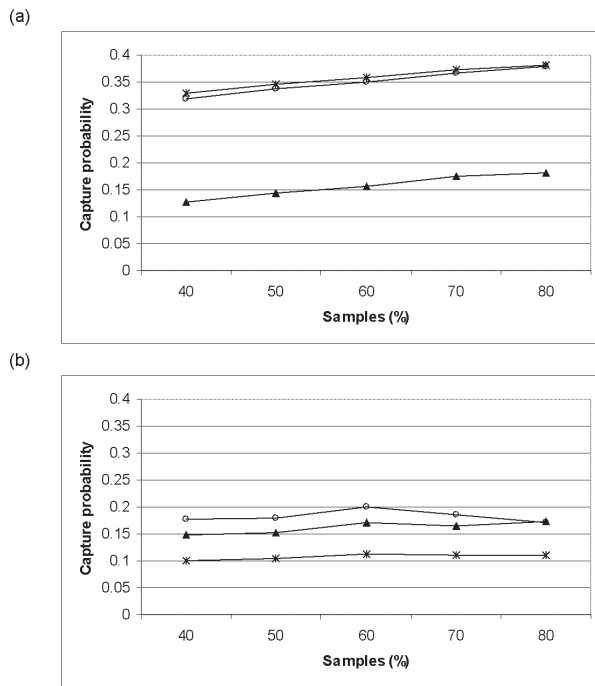
<sup>h</sup>Scenario excluded for St. Johns because 10–30% of traps were visited on average in each period.

Model selection tests (White et al. 1982) showed a high degree of heterogeneity (test 1:  $\chi^2 = 9.68$ , 2 df,  $P = 0.008$ ) and a modest behavioral response (test 2:  $\chi^2 = 4.66$ , 1 df,  $P = 0.030$ ) for the complete dataset from St. Johns (high quality samples only). The model selection routine in CAPTURE subsequently chose model  $M_{bh[REMOVAL]}$ , which estimated a population of 42 bears with a CV of 50.7%. Initial capture rate for this study was 0.11, and the recapture rate was 0.31 (Table 2). Again, due to heterogeneity detected in the data and poor performance of behavior models with sparse data, we also looked at results from the more robust model  $M_{h[CHAO]}$ , which estimated 39 bears (CV = 28.7%, capture rate = 0.18).

**Sub-sampling results**

All scenarios generated a negative bias in population estimates with sub-sampling (Table 3). Total

number of captures ( $M_{t+1}$ ) and capture probabilities increased as the number of samples increased (Fig. 1, Table 3). Accuracy (PRB) and precision (CV) also improved as the number of samples analyzed increased (Fig. 2). The estimators selected to calculate population estimates had a substantial effect on results for all scenarios. Heterogeneity models ( $M_{h[CHAO]}$ ) appeared to perform better than behavior models for most or all of the scenarios in both datasets. Population estimates generated using  $M_{bh[REMOVAL]}$  for the PLNWR data were nearly equal to the total number of captures ( $M_{t+1}$ ) and CIs were very small (Table 3), suggesting poor model performance. This was not the case when  $M_{h[CHAO]}$  was used for this dataset. Additionally, percent relative bias (PRB) for the St. Johns dataset was lower with the  $M_{h[CHAO]}$  model than the PRB generated using the  $M_{bh[REMOVAL]}$  model. CI coverage for the  $M_{h[CHAO]}$  model



**Fig. 1.** Capture probabilities for (a) data at Pocosin Lakes National Wildlife Refuge, North Carolina and (b) St. Johns, Florida using 3 estimators [ $M_{bh[REMOVAL]}$  (\*),  $M_{h[CHAO]}$  ( $\blacktriangle$ ), and best for each ( $\circ$ )]. Only data from scenario 2 (randomly selected sub-samples) were used.

also was greater than coverage under  $M_{bh[REMOVAL]}$ . Remaining results will therefore be presented with reference to results from  $M_{h[CHAO]}$  models only.

All scenarios showed a negative bias in population estimates with sub-sampling relative to our best estimates, but all mean 95% CIs contained these best estimates of 128 and 39 bears (Table 3). Total number of captures ( $M_{t+1}$ ) and capture probabilities increased as the number of samples increased (Fig. 1, Table 3). PRB and CI coverage also improved as the number of samples analyzed increased (Fig. 2, Table 3). Scenarios 1 and 2 produced similar trends in PRB and CI relative to approximate costs, with scenario 1 showing slightly less bias than scenario 2 for similar cost at St. Johns (Table 3). PRB ranged from  $-2$  (PLNWR; scenario 1c) to  $-20$  (PLNWR; scenario 1a; Table 3). For all but scenario 3, estimates were within 20% of the estimates for the complete datasets. In general, scenarios using  $\geq 60$ –70% of samples generated estimates within 10% of the estimates from the complete datasets (Table 3).

With scenario 3 (1 sample/trap from 30% of traps visited in each period), costs were reduced substantially (17% of the total cost for all samples), but PRB was the highest of any scenario. Population size was underestimated by 30%, and the CI for this scenario was 3–24% larger than CIs for other scenarios in this dataset. Because only 10–30% of all traps at St. Johns were visited in each period, scenario 3 was not carried out for this dataset.

Scenario 4 inherently imposes undesirable time variation in capture probabilities. Although models  $M_t$  (Otis et al. 1978) and  $M_{h[CHAO]}$  (Chao 1989) are robust to such time variation, they are not robust to other forms of variation that exist in the data (e.g., heterogeneity). Furthermore, it is unreasonable to introduce such variation through *post hoc* manipulation of the data (i.e., sub-sampling), so this scenario was not considered further.

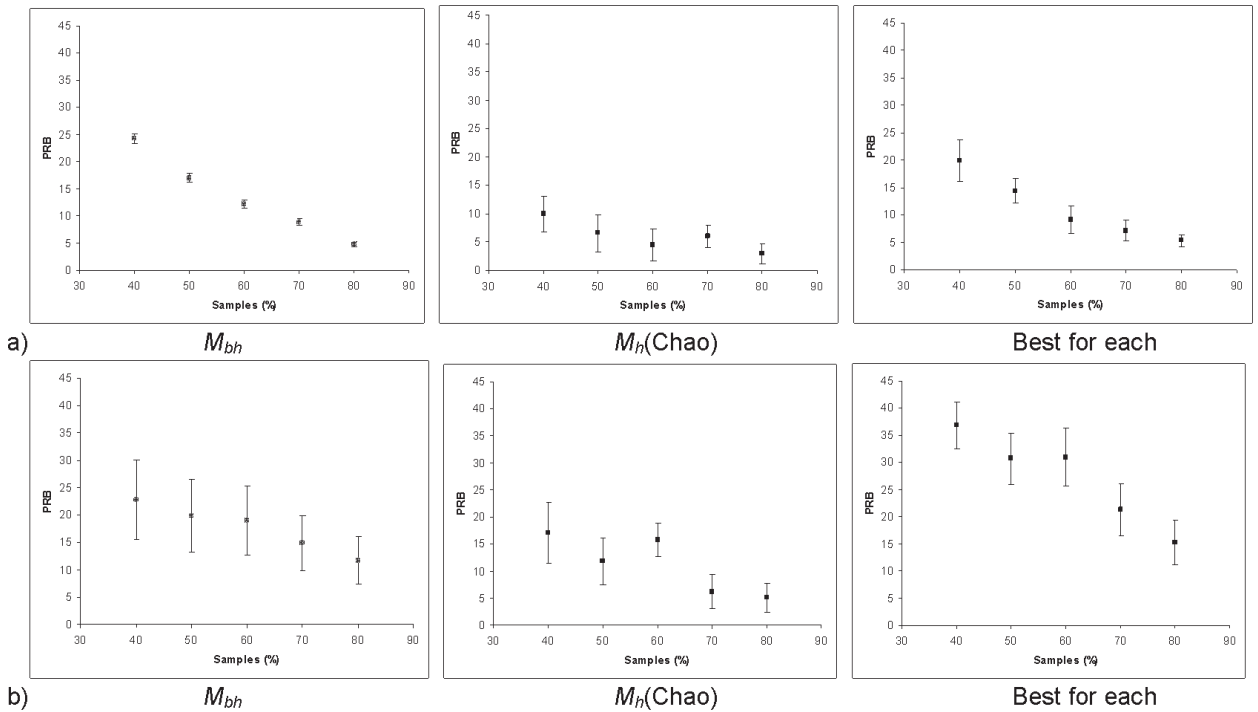
### Sample quality and adjacent barbs

Nearly 45% (91) of the 204 St. Johns samples with  $< 5$  hair follicles (low quality samples) failed to provide sufficient DNA for analysis or failed to produce complete genotypes. In contrast, 12.3% (22 of 179) of the high quality samples ( $> 5$  hair follicles) in the St. Johns dataset and 15.2% (71 of 468) of the high quality samples in the PLNWR dataset failed in the lab. Fifty of 68 (73.5%) lower wire samples had  $< 5$  hair follicles, and of those samples, 24 (48%) failed in the lab.

In the PLNWR population, 22% of samples from directly adjacent barbs were from different bears, and 55% of samples that were 2 barbs apart were from different bears. For St. Johns, 9% of directly adjacent samples and 23% of samples 2 barbs apart were from different bears.

### Discussion

Precision and accuracy of population parameter estimates is of primary concern for properly managing wildlife populations. Although noninvasive techniques show promise for providing precise and accurate estimates, our results show that sub-sampling can lead to a negative bias in population estimates. This bias can be explained in 2 ways. First, sub-sampling decreases estimates of capture and recapture rates (Fig. 1). Capture rates for CMR studies, particularly for large carnivores, are low in general (Otis et al. 1978, Foran et al. 1997, Mills et al. 2000), and sub-sampling data further reduces



**Fig. 2.** Relationship between samples (%) used and accuracy of resulting black bear population estimates. Accuracy is shown as the absolute value of the percent relative bias (PRB;  $(N - \hat{N}/N) \times 100$ ). Data is shown for (a) Pocosin Lakes National Wildlife Refuge, North Carolina, and (b) St. Johns, Florida, using 3 estimators ( $M_{bh[REMOVAL]}$ ,  $M_{h[CHAO]}$ , and best for each) from scenario 2 (randomly selected sub-samples). Error bars indicate 95% CIs on PRB for each scenario.

capture and recapture rates, producing negatively biased population estimates. This was particularly true for the St. Johns dataset since capture rates were low to begin with ( $<0.2$ ). Only 5 cases generated estimates that were within 10% of our best estimate, with most underestimating by 12–24% (Table 3).

Sub-sampling scenarios themselves also may introduce variability in capture probabilities that affect population estimates. For example, scenarios 1A and 3 may produce trap saturation bias that would cause a negative bias in estimates due to heterogeneity variation. This bias would decrease as more samples from each trap were genotyped (i.e., scenarios 1B and 1C). Scenario 4 introduces time variation that would create a negative bias as effort is reduced in later sampling periods, thereby reducing capture probabilities over time (Otis et al. 1978). Scenario 2 caused the least variation in capture probabilities since the sub-sampling is completely random, but scenario 1 may be preferable due to systematic coverage of the study area.

Model selection and behavior can bias estimates. For example, Otis et al. (1978) found that behavior

models can perform poorly, particularly with sparse data, and are less robust than other models (e.g., heterogeneity models). Their simulations showed that with small population size (i.e., 100 animals), population estimates tended to be negatively biased. Similarly, the jackknife (heterogeneity) estimator (Otis et al. 1978) underestimates population size when capture probabilities are low. Chao's (1987) heterogeneity model ( $M_{h[CHAO]}$ ) is more robust to low capture probabilities, but it can produce negative bias if recapture probabilities are high. Finally, the null model ( $M_o$ , Otis et al. 1978) was selected as the best model for an overwhelming majority of the replicates generated for the St. Johns dataset, most likely due to insufficient captures and recaptures with sub-sampling. This tends to occur with sparse data (Otis et al. 1978) and is likely the reason for the negative bias in the results for this dataset.

We found that model selection and performance had a significant influence on estimates for all scenarios. First, the  $M_{bh[REMOVAL]}$  estimator generated population estimates very close to  $M_{t+1}$  values and extremely narrow CIs for PLNWR data (Table 3).

Otis et al. (1978) showed that these models (e.g.,  $M_{bh[REMOVAL]}$ ) can perform poorly, particularly with sparse data, and are less robust to variation in capture probabilities than other models (e.g., heterogeneity models). Their simulations showed that with small population size (100 animals), population estimates tended to be negatively biased. Additionally, they found the  $M_{bh[REMOVAL]}$  estimator to have an “ill behaved” variance estimator that produces CIs that are “inappropriate for practical use” (Otis et al. 1978:42). When the more robust  $M_{h[CHAO]}$  estimator was used, estimates exhibited lower PRB, wider CIs (i.e., low precision), and larger CI coverage (Table 3). This estimator is more robust to low capture probabilities, but generally requires large sample sizes to obtain precise estimates (Boulanger 2002). Mowat and Strobeck (2000) found that under strong heterogeneity,  $M_{h[CHAO]}$  performed acceptably well, with low bias and large CI coverage, similar to our results. Wider CIs also appropriately reflect our uncertainty of estimates due to lower capture probabilities (Boulanger et al. 2004).

Several studies have shown that CAPTURE’s model selection routine performs poorly when sample sizes are small (Otis et al. 1978, White et al. 1982, Menkens and Anderson 1988), suggesting estimates generated from the top models for our datasets may not have been the most accurate. Using more robust estimates that generate lower precision (i.e.,  $M_{h[CHAO]}$ ) may be more appropriate when sample sizes and capture and recapture rates are low. Furthermore, while capture rates for CMR studies, particularly for large carnivores, tend to be low in general (Otis et al. 1978, Foran et al. 1997, Mills et al. 2000), sub-sampling data further reduces these capture and recapture rates, producing negatively biased population estimates.

Issues with study design may have increased heterogeneity and bias in our datasets. For example, violation of closure assumptions can increase heterogeneity in the data because certain bears may have a home range that overlaps only a portion of the sampling grid (‘edge’ bears), thus having a lower probability of capture than bears whose home ranges overlap the entire sampling grid (Boulanger 2002). Additionally, improper trap spacing can increase heterogeneity in CMR data by reducing trap encounter rates for certain animals (e.g., females and females with cubs who tend to have smaller home ranges). Closure violation and improper trap spacing likely were issues with our datasets, partic-

ularly at PLNWR where female home ranges were extremely small. Although  $M_{h[CHAO]}$  estimators are robust to these variations in capture rates, the tradeoff between bias and precision is evident in the wide CIs for our datasets. Increasing trap density, number of sampling periods, and moving traps between periods all can reduce heterogeneity introduced through study design issues (Boulanger et al. 2006).

Finally, genotyping error can bias parameter estimates from noninvasive genetic data. Waits and Leberg (2000) found that high rates of genotyping error leads to positive bias in population estimates, and this bias increases with sampling intensity. This is not the pattern we found with our data. Bias in our data was in most cases negative and actually decreased as sampling intensity increased. If our best estimates from the complete datasets were positively biased due to genotyping error, we would expect this bias to be constant over all levels of sub-sampling in scenario 2 (because samples were chosen randomly and thus the probability of removing an error and of removing a real bear would be fixed). This again is not the pattern we found with our data. Although genotyping error thus becomes an unlikely explanation of the bias in our data, we have no statistical assurance that this is the case. It is important to consider genotyping errors in all studies using noninvasive genetic sampling, as consequences of these errors can be severe (Taberlet et al. 1996, Mills et al. 2000, Waits et al. 2001, Miller et al. 2002, Creel et al. 2003, Paetkau 2003, McKelvey and Schwartz 2004, Roon et al. 2005).

## Recommendations

Our objective was to determine a sub-sampling procedure that balanced economic concerns with accuracy and precision of population estimates. Although no concrete solutions were evident from our analyses, certain recommendations can be made. First, maximizing capture and recapture probabilities in the field is essential for accurate and precise population parameter estimates. Pilot studies should be conducted to determine proper study design, including trap spacing, sampling duration, and sub-sampling strategies, that will maximize capture and recapture probabilities. Second, use of a robust estimator (e.g., the jackknife estimator [Otis et al. 1978] or  $M_{h[CHAO]}$  estimator for sparse data) will reduce bias in parameter estimates.

Third, we recommend the use of only high quality samples (>5 hair follicles) for analysis when budgets are constrained. Nearly half (45%) of the low quality samples (<5 hair follicles) in the St. Johns study failed to produce a useful genotype in the lab, while only 12% of high quality samples from St. Johns and 15% from PLNWR failed. The type of hair collected (guard hairs versus underfur) also can affect genotyping success. For example, 86% of PLNWR samples that relied on 1–3 guard hairs were successful in the lab, versus only 72% that relied strictly on clumps of underfur. In general, however, it is likely more preferable to analyze larger samples with more guard hairs than samples with few guard hairs (i.e., <5 hair follicles) or clumps of underfur.

Finally, our results suggest that sampling evenly over all periods is preferable to sampling more from earlier periods. With both datasets, the number of new captures declined in later periods, but sampling less from these later periods caused new captures and recaptures of animals to be missed at a rate that affected final population estimates (i.e., population size was underestimated by a relatively large margin; C. Tredick, unpublished data). Also, the costs of sampling more from earlier periods were higher in most cases than evenly sampling over all periods, with little or no improvement in population estimates. Similarly, results from scenarios 1 and 2 were similar when the same proportion of samples was analyzed, but scenario 1 may be preferable to ensure systematic spatial coverage of the study area.

Of low quality samples from the lower wire in the St. Johns dataset, 48% failed in the lab. This extra wire may only add extra time in the field and yield low quality samples, though it has proven effective in some areas (T. Eason, personal communication, 2004). Finally, though it may be prudent to avoid analyzing samples from adjacent barbs (because most samples from adjacent barbs will be from the same bear), our results showed this is not always the case (22% of samples from adjacent barbs were from different bears; 55% of samples 2 barbs apart were from different bears). This may be typical in high-density bear populations such as PLNWR, and caution should therefore be used when discarding samples from adjacent barbs.

## Acknowledgments

Funding for fieldwork and genetic analysis was provided by the US Fish and Wildlife Service, the

Florida Department of Transportation, the Florida Fish and Wildlife Conservation Commission, Virginia Polytechnic Institute and State University, and US Geological Survey. We thank D. Paetkau and his staff at Wildlife Genetics International for completing all of the genetic analysis. We also thank all graduate students, volunteers, and technicians who assisted with data collection in the field. J. Boulanger, K. McKelvey, and 2 anonymous reviewers provided useful comments which greatly improved this manuscript.

## Literature cited

- ALLEN, T.G. 1999. Black bear population size and habitat use on Alligator River National Wildlife Refuge, North Carolina. Thesis, University of Tennessee, Knoxville, Tennessee, USA.
- BOULANGER, J., G.C. WHITE, B.N. McLELLAN, J. WOODS, M. PROCTOR, AND S. HIMMER. 2002. A meta-analysis of grizzly bear DNA mark-recapture projects in British Columbia, Canada. *Ursus* 13:137–152.
- , B.N. McLELLAN, J. WOODS, M.F. PROCTOR, AND C. STROBECK. 2004. Sampling design and bias in DNA based capture-mark-recapture population and density estimates of grizzly bears. *Journal of Wildlife Management* 68:457–469.
- , M.M. PROCTOR, S. HIMMER, G. STENHOUSE, AND D. PAETKAU. 2006. An empirical test of DNA mark-recapture sampling strategies for grizzly bears. *Ursus* 17:149–158.
- CHAO, A. 1987. Estimating the population size for capture-recapture data with unequal catchability. *Biometrics* 43:783–791.
- . 1989. Estimating population size for sparse data in capture-recapture experiments. *Biometrics* 45:427–438.
- CREEL, S., G. SPONG, J.L. SANDS, J. ROTELLA, J. ZEIGLE, L. JOE, K.M. MURPHY, AND D. SMITH. 2003. Population size estimation in Yellowstone wolves with error-prone noninvasive microsatellite genotypes. *Molecular Ecology* 12:2003–2009.
- FORAN, D.R., S.C. MINTA, AND K.S. HEINMEYER. 1997. DNA-based analysis of hair to identify species and individuals for population research and monitoring. *Wildlife Society Bulletin* 25:840–847.
- GAGNEUX, P., C. BOESCH, AND D.S. WOODRUFF. 1997. Microsatellite scoring errors associated with noninvasive genotyping based on nuclear DNA amplified from shed hair. *Molecular Ecology* 6:861–868.
- HELLGREN, E.C. 1988. Ecology and physiology of a black bear population in Great Dismal Swamp and reproductive physiology in the captive female black bear. Dissertation, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, USA.

- HINESLEY, L.E. 2000. Pocosin Lakes National Wildlife Refuge: forest habitat management plan. Department of the Interior. US Fish and Wildlife Service. unpublished document. 80 p.
- MCCOWN, J.W., T.H. EASON, AND M.W. CUNNINGHAM. 2001. Black bear movements and habitat use relative to roads in Ocala National Forest. Final Report. Florida Fish and Wildlife Conservation Commission, Tallahassee, Florida, USA.
- MCKELVEY, K.S., AND M.K. SCHWARTZ. 2004. Genetic errors associated with population estimation using non-invasive molecular tagging: problems and new solutions. *Journal of Wildlife Management* 68:439–448.
- MENKENS, G.E.J., AND S.H. ANDERSON. 1988. Estimation of small-mammal population size. *Ecology* 69:1952–1959.
- MILLER, C.R., P. JOYCE, AND L.P. WAITS. 2002. Assessing allelic dropout and genotype reliability using maximum likelihood. *Genetics* 168:357–366.
- MILLS, L.S., J.J. CITTA, K.P. LAIR, M.K. SCHWARTZ, AND D.A. TALLMON. 2000. Estimating animal abundance using noninvasive DNA sampling: promise 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.
- 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., AND C. STROBECK. 1994. Microsatellite analysis of genetic variation in black bear populations. *Molecular Ecology* 3:489–495.
- , W. CALVERT, I. STIRLING, AND C. STROBECK. 1995. Microsatellite analysis of population structure in Canadian polar bears. *Molecular Ecology* 4:347–354.
- . 2003. An empirical exploration of data quality in DNA-based population inventories. *Molecular Ecology* 12:1375–1387.
- 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.
- ROON, D.A., L.P. WAITS, AND K.C. KENDALL. 2005. A simulation test of the effectiveness of several methods of error checking for non-invasive genetic data. *Animal Conservation* 8:203–215.
- SAS INSTITUTE, INC. 2001. SAS Proprietary Software Release 8.2. Cary, North Carolina, USA.
- 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.
- , AND G. LUIKART. 1999. Non-invasive genetic sampling and individual identification. *Biological Journal of the Linnean Society* 68:41–55.
- , L.P. WAITS, AND G. LUIKART. 1999. Noninvasive genetic sampling: look before you leap. *Trends in Ecology and Evolution* 14:323–327.
- WAITS, J.L., AND P.L. LEBERG. 2000. Biases associated with population estimation using molecular tagging. *Animal Conservation* 3:191–199.
- WAITS, L.P., 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., D.R. ANDERSON, K.P. BURNHAM, AND D.L. OTIS. 1982. Capture–recapture and removal methods for sampling closed populations. Report LA-8787-NERP, Los Alamos National Laboratory, Los Alamos, NM, USA.
- WOODS, J.G., D. PAETKAU, D. LEWIS, B.N. McLELLAN, M. PROCTOR, AND C. STROBECK. 1999. Genetic tagging of free-ranging black and brown bears. *Wildlife Society Bulletin* 27:616–627.

*Received: 1 March 2005*

*Accepted: 21 March 2007*

*Associate Editor: R. Harris*