
Estimating Population Size of Elusive Animals with DNA from Hunter-Collected Feces: Four Methods for Brown Bears

EVA BELLEMAIN,^{*†**} JON E. SWENSON,^{*‡} DAVID TALLMON,[†] SVEN BRUNBERG,[§]
AND PIERRE TABERLET[†]

^{*}Department of Ecology and Natural Resource Management, Agricultural University of Norway, Postbox 5003, No-1432 Ås, Norway

[†]Laboratoire d'Ecologie Alpine, Génomique des Populations et Biodiversité, CNRS UMR 5553, Université Joseph Fourier, BP 53, F-38041 Grenoble Cedex 9, France

[‡]Norwegian Institute for Nature Research, Tungasletta 2, NO-7485 Trondheim, Norway

[§]Scandinavian Brown Bear Project, Noppikoski 156, SE-794 98 Orsa, Sweden

Abstract: *Noninvasive genetic methods can be used to estimate animal abundances and offer several advantages over conventional methods. Few attempts have been made, however, to evaluate the accuracy and precision of the estimates. We compared four methods of estimating population size based on fecal sampling. Two methods used rarefaction indices and two were based on capture-mark-recapture (CMR) estimators, one combining genetic and field data. Volunteer hunters and others collected 1904 fecal samples over 2 consecutive years in a large area containing a well-studied population of brown bears (*Ursus arctos*). On our 49,000-km² study area in south-central Sweden, population size estimates ranged from 378 to 572 bears in 2001 and 273 to 433 bears in 2002, depending on the method of estimation used. The estimates from the best model in the program MARK appeared to be the most accurate, based on the minimum population size estimate from radio-marked bears in a subsection of our sampling area. In addition, MARK models included heterogeneity and temporal variation in detection probabilities, which appeared to be present in our samples. All methods, though, incorrectly suggested a biased sex ratio, probably because of sex differences in detection probabilities and low overall detection probabilities. The population size of elusive animals can be estimated reliably over large areas with noninvasive genetic methods, but we stress the importance of an adequate and well-distributed sampling effort. In cases of biased sampling, calibration with independent estimates may be necessary. We recommend that this noninvasive genetic approach, using the MARK models, be used in the future in areas where sufficient numbers of volunteers can be mobilized.*

Key Words: capture-mark-recapture, feces analysis, individual identification, population size estimates, Program MARK, rarefaction, Sweden, *Ursus arctos*

Estimación del Tamaño Poblacional de Animales Elusivos con ADN de Heces Colectadas por Cazadores: Cuatro Métodos para Osos Pardos

Resumen: *Los métodos genéticos no invasivos se pueden utilizar para estimar abundancia de animales y ofrecen varias ventajas sobre métodos convencionales. Sin embargo, se han hecho pocos intentos para evaluar la exactitud y precisión de las estimaciones. Comparamos cuatro métodos para estimar tamaño el tamaño poblacional con base en muestreo de heces. Dos métodos utilizaron índices de rarefacción y dos se basaron en estimadores de captura-marcaje-recaptura (CMR), uno combinando datos genéticos y de campo. Cazadores voluntarios y otros recolectaron 1904 muestras de heces durante 2 años consecutivos en una extensa área que contiene a una población de osos pardos (*Ursus arctos*) bien estudiada. Las estimaciones*

^{**}Address for correspondence: Laboratoire d'Ecologie Alpine, email eva.bellemain@ujf-grenoble.fr
Paper submitted December 2, 2003; revised manuscript accepted May 25, 2004.

poblacionales en nuestra área de estudio de 49,000 km² en el centro-sur de Suecia variaron entre 378 y 572 osos en 2001 y entre 273 y 433 en 2002, dependiendo del método de estimación utilizado. Las estimaciones del mejor modelo en el programa MARK parecieron ser las más precisas, con base en la estimación de tamaño poblacional mínimo de osos radio-marcados en una subsección de nuestra área de muestreo. Adicionalmente, los modelos MARK incluyeron heterogeneidad y variación temporal en las probabilidades de detección, que parecieron estar presentes en nuestras muestras. Sin embargo, todos los métodos incorrectamente sugirieron una proporción de sexos sesgada, probablemente por diferencias de sexo en las probabilidades de detección y bajas probabilidades generales de detección. El tamaño poblacional de animales elusivos se puede estimar confiablemente en áreas extensas con métodos genéticos no invasivos, pero resaltamos la importancia de un esfuerzo de muestreo adecuado y bien distribuido. En caso de muestreo sesgado, puede ser necesaria la calibración con estimaciones independientes. Recomendamos que este método genético no invasivo, utilizando modelos MARK, sea empleado en futuras áreas donde se pueda movilizar un número suficiente de voluntarios.

Palabras Clave: análisis de heces, captura-marcaje-recaptura, estimaciones de tamaño poblacional, identificación individual, MARK, rarefacción, Suecia, *Ursus arctos*

Introduction

Population size estimates are important for the proper conservation and management of species but are often difficult to obtain, especially for rare or elusive species. Noninvasive, DNA-based methods have recently been developed to estimate population sizes, with the major advantage that genetic samples (i.e., hairs or feces) are easily collected without the need to see or disturb the animal (Taberlet et al. 1999). Because each individual is characterized by a unique multilocus genotype, it is possible to determine the number of animals sampled and, through the use of statistical models, to estimate population size. But these methods are not straightforward and have several limitations, mainly because of the low quantity and quality of DNA contained in such samples. Particularly for microsatellites, two main types of scoring error, allelic dropout and false alleles, can lead to incorrect genotypes and consequently to biased estimates (Taberlet et al. 1996, 1999; Smith et al. 2000). Those problems are now well known and methods have been proposed to counter them (Schwartz et al. 1999; Mills et al. 2000; Miller et al. 2002; Creel et al. 2003; reviewed in Paetkau 2003).

Although several estimators and indices are available to estimate population size from noninvasive genetic data, capture-mark-recapture methods (CMR; Seber 1982) and rarefaction analysis (Kohn et al. 1999) are the most common. Most researchers use only one method (e.g., Banks et al. 2003; Boersen et al. 2003; Flagstad et al. 2004; Mowat & Paetkau 2003), which makes them subject to the biases of this method. Few studies have compared population size estimates based on genetic methods with those based on other field data to test the reliability and accuracy of the genetic or field methods. Kohn et al. (1999) compared genotypes from coyote (*Canis latrans*) feces with genotypes from live-trapped coyotes; however, the difference in the time scale of observations did not allow the authors to compare these estimates directly. Creel et al. (2003) used a wolf (*Canis lupus*) population of known

size (recently released) to assess the impact of genotyping error on population size estimates from a census based on genetic analysis of feces. Eggert et al. (2003) compared two genetic methods (rarefaction analysis and CMR) with a “dung-count method” based on counting fresh dung piles in the field and known defecation rates to estimate population size of elephants. Wilson et al. (2003) had to combine the use of trapping records and fecal genotypes of Eurasian badgers (*Meles meles*) to obtain a reliable estimate of group size because each method on its own yielded sampling sizes that were too small. To our knowledge, no one has yet combined field data (with the number of known individuals present) with genetic data to estimate population size at a large scale.

For this study, we examined the Scandinavian brown bear (*Ursus arctos*) population, which is well studied genetically and in the field, to compare the estimates obtained from two rarefaction indices, a closed population CMR method (White & Burnham 1999) and a Lincoln-Peterson CMR method (Seber 1982) that combines genetic and field data. We conducted our noninvasive sampling in 2 consecutive years (autumn 2001 and 2002) to document annual variations in estimates. We were also able to evaluate whether it was feasible to use hunters to voluntarily collect fecal samples over large areas (~50,000 km²).

Methods

Study Area and Studied Populations

The study area covered 48 hunting areas (49,000 km²) in two counties (Dalarna and Gävleborg) in south-central Sweden (Fig. 1). The distribution of brown bears continues north of our study area, but our area included the southern, eastern, and western edges of this population (Swenson et al. 1998b). Elevations range from about 200 m in the south to about 1000 m in the west at the

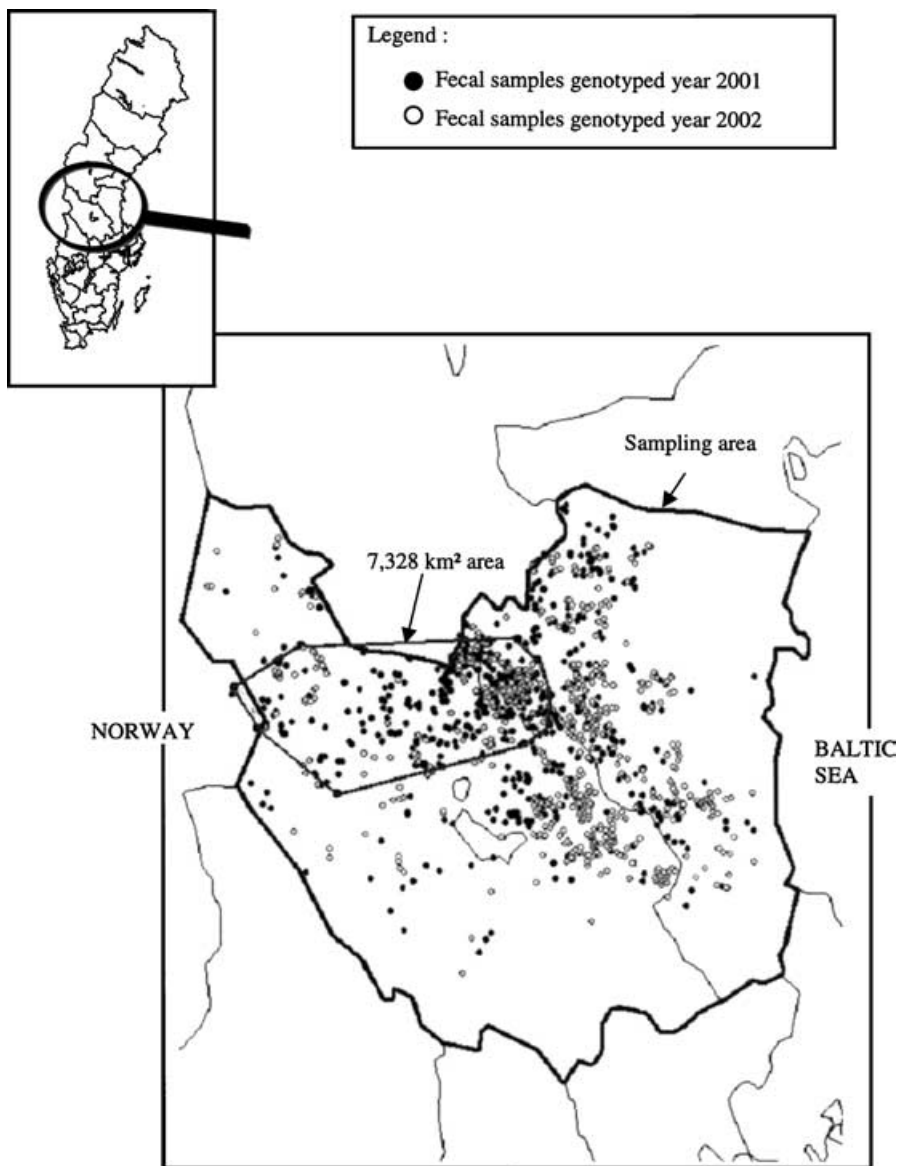


Figure 1. Map of Sweden with the sampling area for feces collection (Dalarna and Gävleborg counties) and with the area defined as the 100% minimum convex polygon formed by the positions of all the marked females in estrus during the mating season (7,328 km²).

Norwegian border, but only a small part of the area is above timberline, which is at about 750 m. Most of the area is covered with intensively managed boreal coniferous forest, dominated by Scots pine (*Pinus sylvestris* L.) and Norway spruce (*Picea abies* L.). Lakes and bogs are common. The area is more completely described in Dahle and Swenson (2003a). The Scandinavian brown bear population has been well studied in the field for more than 18 years (Swenson et al. 1998a). Bears were captured by immobilizing them with tranquilizing darts shot from helicopters. Each bear was fitted with a radio collar and a tissue sample was collected for genetic analyses (Swenson et al. 1998a; Waits et al. 2000).

Fecal Sampling

Searches for bear feces were conducted throughout the study area (Fig. 1), even where bears were rare or nonex-

istent, for 2 consecutive years. The study periods were from mid-August to late October (week 34 to week 44) in 2001 and from early August to late October (week 32 to week 44) in 2002. Fecal samples were collected opportunistically by cooperating moose (*Alces alces* L.) hunters, volunteers, and personnel from the Scandinavian Brown Bear Research Project. Before sampling, we mailed material explaining the project to the leaders of the 48 hunting areas in the study area and to officials of the Swedish Association for Hunting and Wildlife Management and requested their support. Project personnel then described the project at meetings with the hunters in most of the 48 hunting areas. The designated coordinator for each area dispensed collection bottles and sent the collected samples to the research administrator. Hunters picked up each fecal sample with a stick of wood and put 1 cm³ of it in a 20-mL bottle. A different stick and bottle were used for each sample. For each fecal sample, a sampling

date, a geographical location, the hunting team name, and coordinates (along the Swedish Grid) were recorded by the volunteers. Because of strict regulations on the distribution of alcohol, the local coordinator sent the bottles to the research administrator and the bottles were then filled with 95% alcohol to preserve the samples until DNA extraction.

There were about 30,000 moose hunters each year in these two counties. In 2001, 80% of the hunting teams agreed to collect fecal samples, and in 2002, 70% agreed. Hunters collected 90% of the samples; volunteers and project personnel collected 10%.

DNA Extractions and Typing

For every collected fecal sample, DNA extractions were performed using the Qiaamp DNA Stool Kit (Qiagen, Netherlands) and following the manufacturer's instructions. This kit was developed especially for this type of material. All extractions occurred in a room dedicated to processing hair and feces. Tubes containing samples and tubes without feces were treated identically to check for exogenous DNA contaminations.

Each DNA extract was first screened for species-diagnostic amplification with one microsatellite (G10P; Paetkau & Strobeck 1994). Polymerase chain reaction (PCR) products were separated on an agarose gel and viewed under ultraviolet (UV) light. Only the DNA samples that showed a fluorescent DNA band were kept. Next, six microsatellite primers (Mu10, Mu23, Mu50, Mu51, Mu59, G10L; Paetkau & Strobeck 1994; Taberlet et al. 1997) were amplified using PCR. Those six microsatellites were chosen for their discriminatory power (loci with the lowest probability of identity) based on the Scandinavian brown bear tissue data set (Waits et al. 2000; E.B. et al. unpublished data). The number of alleles per locus ranged from six to nine, with a mean observed heterozygosity of 0.70.

For sex identification, we used newly defined sex primers (Bellemain & Taberlet 2004). The amplification of microsatellites was carried out following the multiplex preamplification method (Piggott et al. 2004). This new method allows one to maximize the number of samples that contain the critical threshold amount of DNA for accurate genotyping. The first step involves a multiplex preamplification that simultaneously amplifies all microsatellite loci to be subsequently genotyped. Postamplification aliquots are then used as a template in locus-specific PCRs to genotype individuals. One primer of each pair was synthesized with a fluorescent dye group (6-FAM, TET, or HEX) on the 5' end to allow detection and sizing of fragments on an ABI Prism 3100 DNA sequencer (Applied Biosystems, Foster City, California).

We repeated each amplification four times. We reduced the number of replicates recommended by the multiplex tubes approach (Taberlet et al. 1996) because a prelimi-

nary study showed that reliable genotypes could be obtained after four replicates with this new multiplex preamplification method (i.e., the genotyping error rate was not much different when repeating the amplifications four times instead of eight [E.B., unpublished data]). We typed samples as heterozygous at one locus if both alleles appeared at least twice among the four replicates and as homozygous if all the replicates showed identical homozygous profiles. If neither of those cases applied, we treated the alleles as missing data. The gels were analyzed using the Genemapper software package (version 3.0, Applied Biosystems, Foster City, California). In total, the laboratory work (extractions, amplifications, and data analysis) involved 7 work months in 2001 and 6 work months in 2002, assuming 20 hours of laboratory work per week.

Reliability of the DNA Results

Genotypes from different samples were considered to represent an identical bear when all the alleles at all loci were identical. To be conservative, however, we assumed that two samples belonged to the same individual when there was only one mismatch for one allele at one locus. To verify the reliability of the genotypes and to calculate an error rate, about 5% of the successfully genotyped samples were randomly chosen to be amplified another four times (including the sex primer). The genotypes were then compared with the first typing.

To quantify the power of the microsatellite loci used, we computed the probability of identity (PI; i.e., the overall probability that two individuals drawn at random from a given population share identical genotypes at all typed loci [Paetkau & Strobeck 1994]). We also computed the PI between siblings (PIsibs; Waits et al. 2001).

We examined the tissue dataset for bears from southern Sweden (655 genotypes) only for the seven loci used for fecal genotyping (six microsatellites and one sex primer) and checked the proportion of identical genotypes or those mismatching at one allele. We also looked at the distribution of loci differences between pairwise genotypes for the feces and the tissue data sets to ensure that allelic dropout was not a problem for the former.

The reliability of the sex primer was verified by (1) comparing sexes assigned to feces samples with sexes from identically genotyped marked bears and (2) checking the consistency of sex determination within a set of identical genotypes.

Finally, for animals with multiple fecal samples, and corresponding to radio-marked individuals, we checked the percentage of fecal samples located inside the home range of the corresponding radio-marked bear. Minimum home ranges of radio-marked bears were estimated using 95% minimum convex polygon with a minimum of 16 locations per year, as described in Dahle and Swenson (2003a).

Table 1. Closed population models used to estimate male and female brown bear abundance.

| Model* | Description | 2001 | | 2002 | |
|---------------|--|---------------|------|---------------|------|
| | | $\Delta AICc$ | w | $\Delta AICc$ | w |
| <i>pb*t</i> | heterogeneity and temporal variation in detection probability | 0.00 | 1.00 | 0.00 | 1.00 |
| <i>pb*t*g</i> | heterogeneity and temporal variation in detection probability for each sex independently | 16.86 | 0.00 | 58.48 | 0.00 |
| <i>pt</i> | temporal variation in detection probability | 21.57 | 0.00 | 65.11 | 0.00 |
| <i>pt*g</i> | temporal variation in detection probability for each sex independently | 38.27 | 0.00 | 81.97 | 0.00 |
| <i>pb</i> | heterogeneity in detection probabilities | 295.39 | 0.00 | 232.99 | 0.00 |
| <i>pb*g</i> | heterogeneity in detection probability for each sex independently | 295.96 | 0.00 | 236.68 | 0.00 |
| <i>p.</i> | constant detection probability | 306.46 | 0.00 | 292.65 | 0.00 |
| <i>Pg</i> | sex-specific detection probabilities | 307.71 | 0.00 | 291.81 | 0.00 |

*Each model has different constraints on capture (p) probabilities. The $\Delta AICc$ value and weight (w) for each model each year is shown. Each model uses two parameters to estimate the abundances of males and females.

Population Size Estimation

Independently, we applied the methods described below to bear genotypes obtained in 2001 and 2002.

RAREFACTION INDICES

Following the method described in Kohn et al. (1999), we compared the multilocus genotype of each sample with all those drawn previously and calculated the population size as the asymptote of the relationship between the cumulative number of unique genotypes and the number of samples typed. This curve is defined by the equation $y = (ax)/(b + x)$, where a is the asymptote, x is the number of feces sampled, y is the number of unique genotypes, and b is the rate of decline in the value of slope. Eggert et al. (2003) derived another estimator with a similar equation: $y = a(1 - e^{-bx})$. These are referred to as the Kohn and Eggert methods, respectively. We analyzed data with the software package GIMLET (Valière 2002), with 1000 random iterations of the genotype sampling order and rarefaction curves obtained from those iterations were plotted using R software (version 1.7.1; available at <http://www.r-project.org>).

It is very difficult to obtain an analytical expression to calculate confidence intervals (CI) for rarefaction curves, so we employed the iterative approach, which is commonly used. This approach, however, gives an indication only of the sampling variance, not of the estimator variance.

CMR ESTIMATORS

The first CMR estimator was based on genetic data only. We grouped identical multilocus genotypes and compiled a "capture" and "recapture" history for each individual by dividing the data set into 11 weekly sampling periods for 2001 and 13 weekly sampling periods for 2002 (only the weeks with an adequate sample size for estimating population size were considered). If an individual's feces were captured two or more times within the same capture pe-

riod, only one capture was considered. We analyzed data as conventional CMR data with the closed-capture models of the program MARK (White & Burnham 1999). We developed and analyzed a set of a priori models, then ranked them by AICc (Akaike's information criterion corrected for small sample size) values following analysis in MARK. These ranged in complexity from simple models with constant detection probabilities to those containing heterogeneity and temporal and sex-specific variation in detection probabilities (Table 1).

The second CMR estimator combined field and genetic data. Brown bears have been marked and followed with radio transmitters in the study area for more than 18 years (Swenson et al. 1998a). The number of living bears with a functioning radio transmitter in the study area during the sampling period was known from radio telemetry data. Consequently, it was possible to identify marked bears among the fecal samples, and the following Lincoln-Peterson estimator was applied to estimate the population size (Seber 1982):

$$N = \frac{(C + 1)(M + 1) - 1}{R + 1},$$

where C is the number of unique feces genotypes, M is the total number of marked bears alive in the study area during the sampling period, and R is the number of marked bears identified among the fecal samples in the area. We used binomial and Poisson confidence intervals for this value depending on the R/C ratio (Krebs 1999).

Sex Ratio Estimate from Independent Data

We used data from radio-marked bears for the period from 1996 to 2002 (years with a high and stable bear hunting pressure) to calculate survival rates with the Kaplan-Meier method (Pollock et al. 1989) for males and females in our study area. Using these survival rates and observed reproductive rates, we determined the sex ratio in the population, assuming a 50:50 sex ratio at birth, which occurs in brown bears in the United States (Schwartz et al. 2003). We compared the sex ratio estimated from the

fecal samples with the population sex ratio to correct for bias if necessary.

Results

Microsatellite Typing

In 2001, 1066 fecal samples were collected and 728 (~70%) were successfully amplified for six to seven loci (including the sex locus). From these 728 samples, 311 unique genotypes were obtained (43% males and 57% females). Each multilocus genotype was found from 1 to 13 times, with a mean of 2.40 ± 2.20 (SE) times (2.38 ± 2.34 for females and 2.41 ± 1.99 for males).

In 2002, 838 fecal samples were collected and 665 (~80%) were successfully amplified for six to seven loci (including the sex locus). From these 665 samples, 239 unique genotypes were obtained (45% males and 55% females). Each multilocus genotype was found from 1 to 25 times, with a mean of 2.78 ± 3.73 times (2.54 ± 3.02 for females and 3.39 ± 4.75 for males). In 2002 we sampled 53% of the individuals identified in 2001. Those individuals, however, comprised 71% of the genotypes successfully extracted from feces in 2002, which was significantly different from expected if the detection probability had been the same for the individuals first detected in 2001 and for those first detected in 2002 ($\chi^2 = 24.97$; $p < 0.005$). A few samples for which the sex could not be determined were removed from the data set before further analysis.

Reliability of the DNA Results

The PI among the six amplified microsatellite loci was low ($PI = 1.38 \cdot 10^{-6}$; $PI_{sibs} = 4.52 \cdot 10^{-3}$), which allowed us to reliably identify each individual. In the tissue database, none of the individuals were genetically identical, six were genetically different for only one allele at one locus, and nine were different for the sex locus only. In addition, the proportion of genotype mismatches for one locus was

identical for the tissue and fecal data sets, suggesting that allelic dropout was not an issue.

About 5% of the samples were amplified another four times to check the reliability of the results. For 2001, 854 alleles were identical and 8 were different from the first typing. For 2002, 503 alleles were identical and 11 were different. This error rate of about 2% resulted from allelic dropout, false alleles, or contaminations (Taberlet et al. 1996).

Indeed, when comparing sex determination based on fecal samples with the sex determination of identically genotyped marked bears, we found six incompatibilities (corresponding to three individuals) among 295 genotype comparisons. When comparing the consistency of sex determination within identical genotypes, we found 16 incompatibilities, corresponding to 11 individuals in 1049 comparisons. Consequently, the error rate of this sex primer was below 2%. To be conservative, we considered that samples with identical genotypes, but different sexes, belonged to the same individual.

Eighty percent of the 77 fecal samples from 24 radio-marked bears were located within their minimum home ranges as estimated by radio telemetry. The other 20% were close (<10 km from the edge of the defined home range), except for one fecal sample that was located 20 km away.

Comparison of Population Size Estimates

Estimates of population sizes with confidence intervals from the different methods showed consistent patterns between 2001 and 2002 (Fig. 2). In both years there was considerable variation among the four methods, and the estimates were uniformly lower for all methods in 2002 than in 2001. All methods suggested a female-biased sex ratio. In one case, there was no overlap in the 95% confidence interval for the estimated number of females and males, implying a considerable bias in the sex ratio.

Kohn's method yielded the highest point estimates and the largest confidence intervals. The values generated

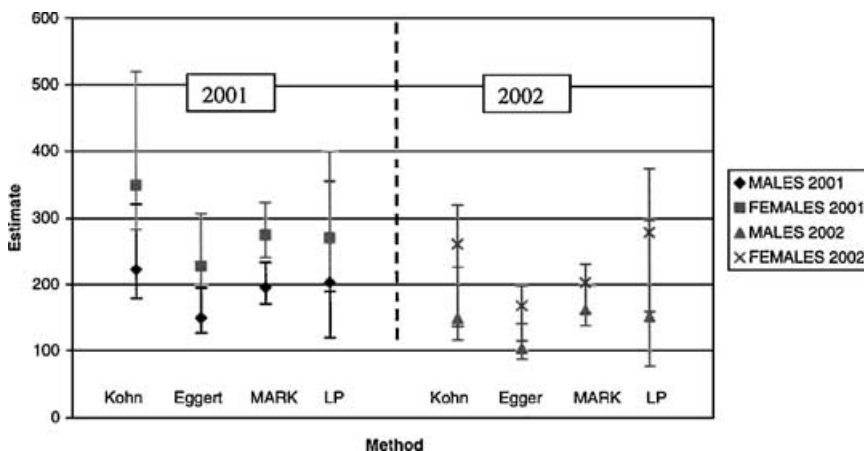


Figure 2. Brown bear population estimates by sex and year in southern Sweden based on different methods (estimate \pm 95% CI).

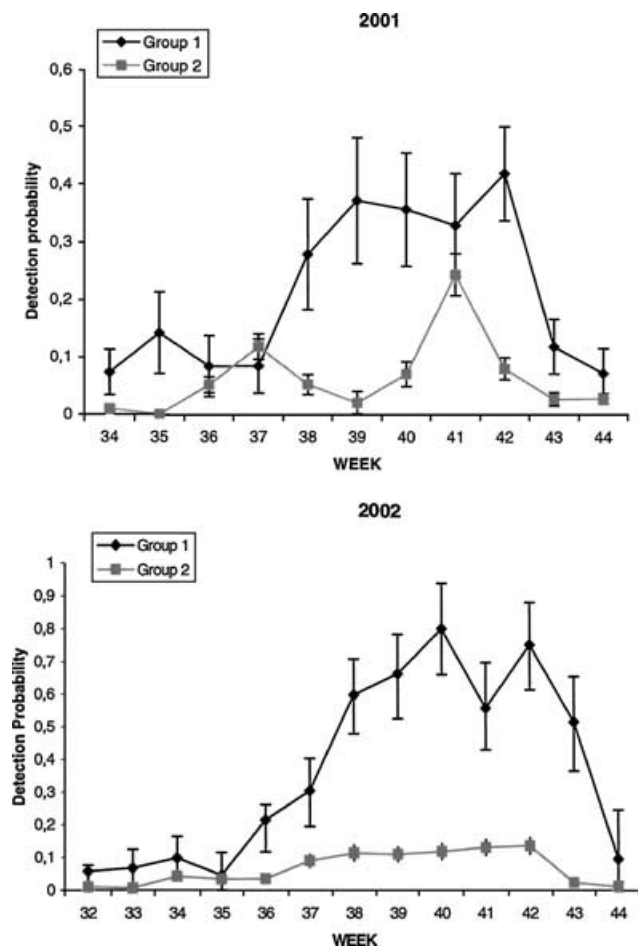


Figure 3. Detection probabilities over time for the southern population of Scandinavian brown bears. Group 1 (triangles) and Group 2 (squares) respectively represent the "lower" and "higher" detection probability groups.

with Eggert's method were consistently lower than the CMR-based estimates. This pattern was evident in both 2001 and 2002.

For both years, the best approximating model of the MARK closed-capture candidate model set included heterogeneity and temporal variation in detection probabilities, whereas the models receiving the least support were relatively simple ones that included less variability in detection probability (Table 1). This suggests that there was considerable variation among individual bears in their detection probability and that these heterogeneous detection probabilities changed over sampling intervals within each year. The latter was clearly evident (Fig. 3).

The 2001 Lincoln-Peterson estimate was calculated using the number of unique feces genotypes ($C = 319$), total number of marked bears in the study area ($M = 56$), and number of marked bears identified among the fecal samples in the area ($R = 36$). In 2002 these values were

$C = 238$, $M = 57$, and $R = 28$. The Lincoln-Peterson total population size estimates were the most similar between years (Fig. 2).

Estimate of Sex Ratio from Independent Data

Survival rates were not significantly different between males and females for either subadults (1–4 years old: survival of males = $0.842 \pm 0.051SE$, $n = 44$; females = 0.838 ± 0.037 , $n = 102$; $Z = 0.64$, $p = 0.74$) or adults (5 years old: survival of males = 0.850 ± 0.046 , $n = 70$; females = 0.885 ± 0.030 , $n = 116$; $Z = 0.053$, $p = 0.52$). The calculated sex ratio based on these survival probabilities and assuming a stable population was not significantly different from 50:50 ($\chi^2 = 0.0256$; $p = 0.89$) with 49.2% males and 50.8% females.

Discussion

Reliability of the DNA Results

We calculated a low genotyping error rate (<2%) and a low probability of identity. A similar error rate was determined for the sex PCR. However, because we allowed one genetic mismatch at one allele for one locus (including the sex locus) to be considered as two samples belonging to the same individual, there is a small probability that we underestimated the real number of different genotypes. This probability can be estimated from the tissue data set. Fifteen genotypes out of 655 were identical at all seven loci considered, except for one allele (including the sex locus). This gives an underestimation probability of about 0.4% ($15/(655 \times 654/2)$), which can be considered negligible. Furthermore, we checked that allelic dropout, a potentially serious problem with low-quantity DNA samples, was not an issue in our data set. We also verified the geographical consistency of our results. Only one fecal sample from a radio-marked bear was located relatively far (20 km) from the edge of the corresponding telemetry home range. This sample, however, was from an estrous female, and estrous females wander extensively during the mating season (Dahle & Swenson 2003b). She may also have been inside her usual home range because the home ranges based on a minimum of only 16 relocations were underestimates of the true home ranges (Dahle & Swenson 2003a). Therefore, we consider our results reliable.

We also emphasize that the quality of the results can vary greatly. The genotyping error we measured includes both intrinsic errors (PCR artifacts and allelic dropout resulting from low-quantity and low-quality DNA) and laboratory- and experimenter-dependent errors (e.g., sampling errors, tube mixing, and scoring errors). Consequently, a thorough pilot study, which will offer insight

into sources of errors and the opportunity to reduce these sources, is a requisite step in projects such as ours (Bonin et al. 2004).

The Assumptions and Reliability of Each Estimator

Noninvasive DNA-based methods are an efficient method for estimating animal population sizes, but one must be careful when interpreting the estimates obtained. Boulanger et al. (2002) pointed out that the most important challenge that needs to be addressed is designing studies to simultaneously minimize closure violations and capture probability variation while obtaining adequate sample sizes. Our rarefaction indices assumed that detection probabilities are constant over space and time and equal among all individuals. We showed that this assumption was violated (Fig. 3). The CMR methods we used are based on the following assumptions: (1) closed population, (2) uniform probability of capture, and (3) no loss of mark (i.e., the genetic tag in our case) by the animals during the sampling period. The CMR models in MARK, though, allow one to relax the second assumption.

The population closure assumption rests on there being no immigration, emigration, birth, or death. Thanks to radio telemetry data, we know that the study area included the home ranges of marked bears and encompassed the edge of the population distribution to the south, west, and east (Fig. 1; Swenson et al. 1998b). Thus, bear movements outside the sampling areas are improbable, except some to the north. Births did not occur during our study period because bears give birth in late winter. Our study occurred during the bear hunting season, when most bear mortality occurs. In our study area, 22 bears were killed in 2001 and 34 were killed in 2002. Because there was heterogeneity in detection probabilities, the effect of mortality on the MARK closed population estimator is difficult to quantify (Kendall 1999). Nevertheless, the sampled feces included feces deposited prior to the hunting season, and defecation rates are high during the period of hyperphagia, which begins in July (Roth 1980). Therefore, hunter-killed bears could also have been detected in our sampling, and fewer feces would have been available for those killed early.

The assumption of uniform detection probability implies no behavioral response, no heterogeneity among individuals, and no temporal variation in the detection probability. A noninvasively assigned genetic tag should not affect the probability of recapturing the animal (e.g., there is no behavioral response). Temporal variation seemed likely, however, given that sampling effort changed over time in the study area, with changes in the number of hunters available to collect samples. In addition, some bears were more likely to be detected than others. Results from MARK showed that there was a noticeable increase

in detection probabilities for the "high" detection probability group in week 38 in 2001 and 2002 (Fig. 3), which coincided with the start of the moose hunting season. This detection probability dropped when the first heavy snows covered the scats (week 43 in 2001; week 41 in 2002). Moreover, when comparing genotypes obtained in the 2 sampling years, the individuals we detected in 2002 also occurred most frequently in the 2001 samples (53% of the individuals from 2001 were detected in 2002, corresponding to 71% of the genotypes), which further confirmed heterogeneity among individuals.

The bias in the sex ratio, based on the fecal samples, also confirmed an unequal probability of detection for the sexes. Results in many studies show that bears exhibit heterogeneous detection probabilities as a function of sex (Mace et al. 1994; Boulanger 2002). Here, the sex ratio based on population estimators was biased in favor of females in both sampling years, although the actual population sex ratio, based on survival rates for adult and subadult radio-marked bears, was 50:50, which means that females had a higher probability of detection than males. A difference in genotyping success between genders is improbable because males and females have very much the same diet; in the study area, in autumn, 99% of feces contain berries (Opseth 1998). The most probable reason for the biased sex ratio is a difference in habitat use between males and females. Female bears on the study area preferred forested habitats to a significantly greater degree than males, and males appeared to avoid bogs less than females, although the latter difference was not statistically significant (Swenson et al. 1996). Moose hunters typically concentrate their time, and therefore their search for bear feces, in forested habitats, spending little time on bogs. Thus, the hunters would have had a higher probability of encountering feces from female bears. Only MARK allowed us to consider different models that evaluated the violations of the assumptions described above. The model selected using AICc, however, did not include sex-specific differences in addition to heterogeneity in detection probability, although the second-best model, which received little support, included sex-specific differences (Table 1). Heterogeneity in detection probability may have been the major source of variation, compared with sex-specific differences. With larger sample sizes it is possible that the more complicated sex-specific models may have received more weight.

The third assumption (animals do not lose their genetic tag) would be violated in cases in which genotyping errors occur, causing, for example, two fecal samples to be considered to be from different individuals even though they came from the same bear. Because we documented a low genotyping error rate and grouped genotypes that were different at one allele for one locus in order to avoid this type of problem, we are confident that this assumption was met.

Comparisons of Estimates for the Same Population Generated in Two Different Years

Population size estimates for brown bears were lower in 2002 than in 2001, although the confidence intervals overlapped for the 2 years and for all methods. We cannot conclude, however, that the population size decreased. A measurable decrease in population size over 1 year is unlikely, and this population of brown bears has been documented as growing rapidly (Saether et al. 1998), even though there was a lower harvest rate at that time. The differences in the estimates between 2001 and 2002 probably resulted from differences in sampling effort and from the smaller sample size in 2002. If we consider only the 7328-km² study area, within part of our sampling area (Fig. 1), defined as the 100% minimum convex polygon (MCP) formed by the positions of all the marked females in estrus during the mating season, only 154 fecal samples were genotyped in 2002, whereas 359 were genotyped in 2001. The 2002 sampling was also more concentrated in areas with a lower density of bears (southwest of the study area; Fig. 1), inducing an unequal spatial sampling. Consequently, many bears were missed in the 2002 sampling.

Comparing Estimates Obtained from the Different Methods

Because all our estimators and indices are based on noninvasive genetic data, we could not conclude independently which estimate was best. We were able, however, to compare our estimates with independent minimum population size estimates from the 7328-km² study area defined above (Fig. 1), which is rare for this type of study. Minimum population size estimates were calculated from the known minimum number of radio-marked females in estrus inside this smaller study area, and based on long-term demographic data on marked bears' survival rates and reproduction, with the equation described in Appendix 1. Because this area was located entirely within ours, we considered only the fecal samples collected in the smaller area and estimated the population size for this area so we could compare our results directly with the minimum population size estimate. There were known minima of 14 and 22 radio-marked females in estrus inside the study area at the end of the mating seasons in 2001 and 2002, respectively (Solberg et al., unpublished data). This resulted in a calculated minimum population size of 105 (95% CI: 90–124) bears in 2001 and 164 (95% CI: 141–195) bears in 2002. To avoid overestimates resulting from possible synchrony in females entering estrus (Saether et al. 1998), we used the mean value of minimum population size estimates between 2001 and 2002, which was 134 (95% CI: 115–159). We considered this value as a baseline with which we could compare our estimates, even though we are aware that this is an underestimate of the true population size (because unmarked females were not considered in the calculation).

Table 2. Brown bear population size estimates obtained from the area defined as the 100% minimum convex polygon (MCP) formed by the positions of all the marked females in estrus during the mating season (7328 km²).^a

| | 2001 (CI) | 2002 (CI) |
|-------------------|---------------|----------------|
| Kohn's method | 247 (205–359) | 236 (141–1121) |
| Eggert's method | 167 (146–220) | 140 (93–580) |
| MARK ^b | 223 (188–282) | 157 (119–227) |
| LP ^c | 219 (157–314) | 204 (136–272) |

^aThe minimum population size estimated from the known number of radio-marked estrous females was 134 (95% CI: 115–159).

^bEstimate from the best model of the program MARK (White & Burnham 1999).

^cEstimate from the Lincoln-Petersen estimator combining field and genetic data.

A comparison of our estimates with this minimum estimate (Table 2) showed that Eggert's method provided an underestimate of the true population size because the values were very close to the minimum population size. Estimates from Kohn's method, the Lincoln-Peterson estimator, and MARK appeared to be the most credible because they were above the known minimum population size. The rarefaction curves obtained for both years from the observed data (Fig. 4), however, showed that the estimate derived from Kohn's equation may have overestimated population size. Other researchers have also found that Kohn's method gave an estimate higher than what would be expected by the rarefaction curve (Wilson et al. 2003; H. Brøseth, personal communication).

Based on the above findings, we conclude that the CMR methods, namely Lincoln-Peterson and MARK closed population estimators, were most reliable. Both estimators have their limitations, however. Although the Lincoln-Peterson estimator gave the most consistent estimates between the 2 study years, it had the rarely available advantage of combined conventional and noninvasive data—which means more cost per effort—and it always had very large confidence intervals. The MARK estimator, which allows one to incorporate heterogeneity and temporal variation in detection probabilities, had reasonably small confidence intervals and relied only on noninvasively collected data. But this estimate fluctuated across years more than the Lincoln-Peterson estimator, thus appearing to be more sensitive to potential sources of bias.

Finally, to obtain a total population estimate, we considered that the MARK program gave the most credible estimate, but like the other methods, it implied a biased sex ratio that does not exist. We also considered that the 2001 estimate was more credible than the 2002 one. The combination of fecal analysis and demographic data showed that females had a higher probability of detection, so we believe that the estimated number of females was more credible than the number of males. Consequently, the total population size estimate was corrected by doubling

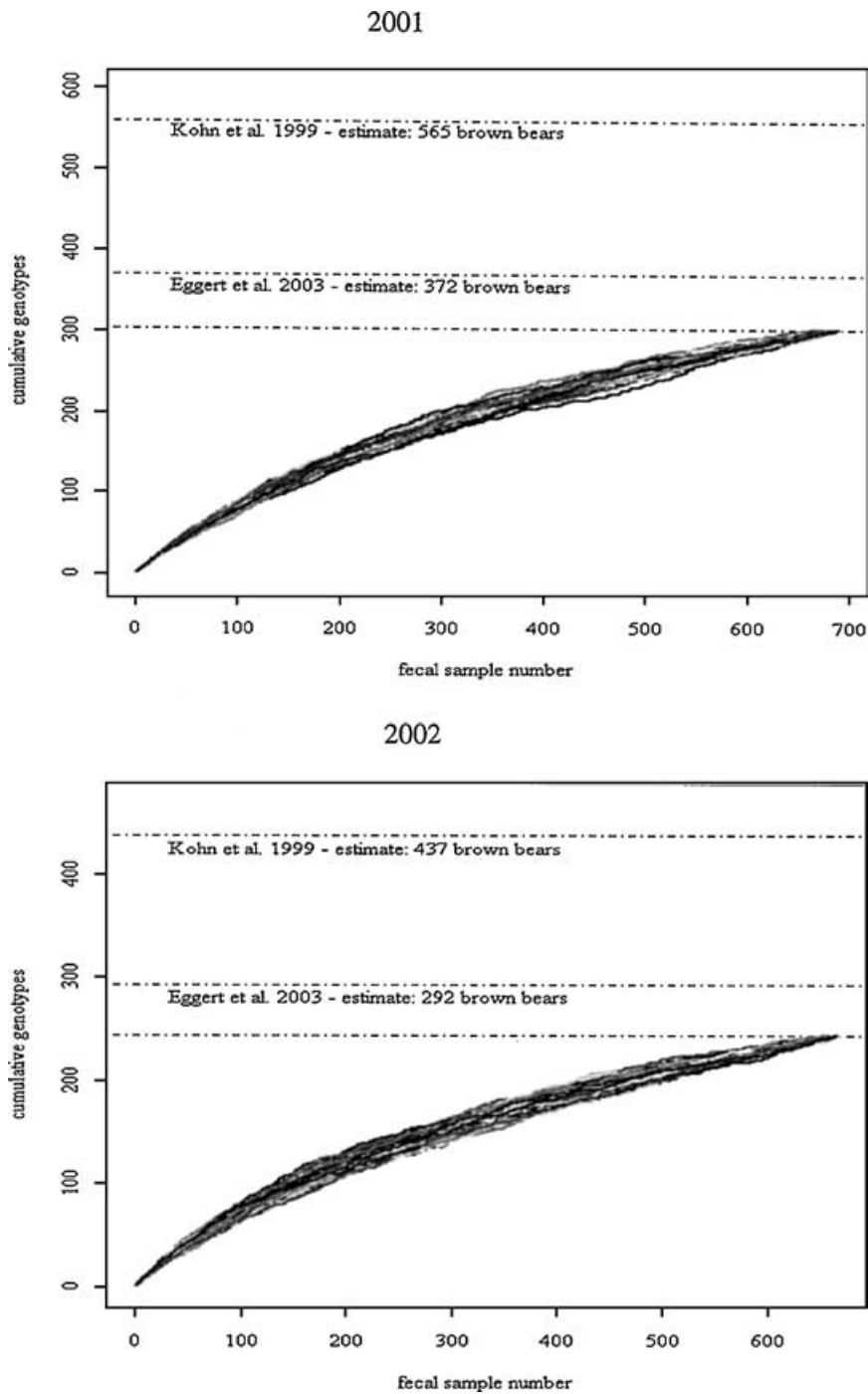


Figure 4. Rarefaction curves obtained from the observed data for the southern population of Scandinavian brown bears compared with estimates obtained from the Kohn and Eggert methods. The different curves represent the different iterations of genotype sampling order.

the estimated number of females (i.e., 550 [482–648] bears in Dalarna and Gävleborg counties).

Recommendations

Our research benefited from a large, cooperative, and well-organized group of hunters such as is found in northern Europe. Their motivation was to help obtain an accurate estimate of the number of bears, and their involve-

ment was important for the local acceptance of the results. In addition, the volunteers helped make it economically possible to sample such a large area.

We did find, however, evidence of uneven sampling effort over time (within and between years) and space (among hunting areas within the study area). Those differences probably contributed to the temporal variation and heterogeneity in detection probabilities in the preferred MARK model. In addition, extremely low detection probabilities can make estimates imprecise and less

useful for management. This suggests a trade-off between the reduced cost of obtaining a large-scale sample with hunter-collected samples and the representativeness of the samples. The easiest means to minimize heterogeneity and temporal variation in detection probabilities while increasing detection probabilities would be to try to ensure even coverage of the sampling area by adding sampling effort to areas less visited by hunters or where the hunters do not wish to participate. Demographic models that can accommodate these factors and that can account for them in population size estimation should also be used. The results from MARK also suggested a biased sex ratio and no sex difference in detection probability in the preferred model. This was inconsistent with the independent field data from the population. One should not uncritically accept that a population has a biased sex ratio from results of fecal sampling. Despite those concerns, we recommend this approach for estimating the population size of elusive animals over large areas, and where many competent and interested volunteers are available and willing to collect a large number of samples.

Acknowledgments

This study was supported by the Norwegian Directorate for Nature Management and the Swedish Association for Hunting and Wildlife Management. This study was possible because of the ongoing Scandinavian Brown Bear Research Project (SBBRP), which is also funded by the funding sources mentioned above, as well as by the Swedish Environmental Protection Agency, the Norwegian Institute for Nature Research, the World Wildlife Fund-Sweden, Orsa Besparingssskog, and several private foundations. D. Tallmon was supported by a National Science Foundation postdoctoral fellowship (INT 0202707) during preparation of this manuscript. We are very grateful to cooperative hunters, personnel from the SBBRP, and all other volunteers who enthusiastically collected so many fecal samples. We are grateful to H. Brøseth for making the plots of rarefaction curves, and many thanks to C. Poillot, E. Zeyl, B. Kovak, L. Gielly, C. Miquel, and D. Rioux for their help in the laboratory.

Literature Cited

- Banks, S. C., S. D. Hoyle, A. Horsup, P. Sunnucks, and A. C. Taylor. 2003. Demographic monitoring of an entire species (the northern hairy-nosed wombat, *Lasiorhinus krefftii*) by genetic analysis of non-invasively collected material. *Animal Conservation* **6**:101-107.
- Bellemain, E., and P. Taberlet. 2004. Improved noninvasive genotyping method: application to brown bear (*Ursos arctos*) feces. *Molecular Ecology Notes* **4**:519-522.
- Boersen, M. R., J. D. Clark, and T.L. King. 2003. Estimating black bear population density and genetic diversity at Tensas River, Louisiana using microsatellite DNA markers. *Wildlife Society Bulletin* **31**:197-207.
- Bonin, A., E. Bellemain, P. Bronken Eidesen, F. Pompanon, C. Brochmann, and P. Taberlet. 2004. How to track and deal with genotyping errors in population genetics studies. *Molecular Ecology* **13**:3261-3273.
- Boulanger, J., G. C. White, J. Woods, M. Proctor, and S. Himmer. 2002. A meta analysis of grizzly bear mark-recapture projects in British Columbia, Canada. *Ursus* **13**:137-152.
- Creel, S., G. Spong, J. Sands, J. Rotella, J. Zeigle, L. Joe, K. M. Murphy, and D. Smith. 2003. Population size estimation in Yellowstone wolves with error-prone noninvasive genotypes. *Molecular Ecology* **12**:2003-009.
- Dahle, B., and J. E. Swenson. 2003a. Home ranges in adult Scandinavian brown bears *Ursus arctos*: effect of mass, sex, reproductive status, population density and habitat type. *Journal of Zoology* **260**: 329-335.
- Dahle, B., and J. E. Swenson. 2003b. Seasonal range size in relation to reproductive strategies in brown bears *Ursus arctos*. *Journal of Animal Ecology* **72**:660-667.
- Eggert, L. S., J. A. Eggert, and D. S. Woodruff. 2003. Estimating population sizes for elusive animals: the forest elephants of Kalum National Park, Guana. *Molecular Ecology* **12**:1389-1402.
- Flagstad, O., E. Hedmark, A. Landa, H. Broseth, J. Persson, R. Andersen, P. Segerstöm, and H. Ellegren. 2004. Colonization history and noninvasive monitoring of a re-established wolverine population. *Conservation Biology* **18**:676-688.
- Kendall, W. L. 1999. Robustness of closed capture-recapture methods to violations of the capture assumptions. *Ecology* **80**:2517-2525.
- Kohn, M. H., E. C. York, D. A. Kamradt, G. Haught, R. M. Sauvajot, and R. K. Wayne. 1999. Estimating population size by genotyping feces. *Proceedings of the Royal Society of London* **266**:657-663.
- Krebs, C. J. 1999. *Ecological methodology*, second edition. Addison Wesley, Menlo Park, California.
- Mace, R. D., S. C. Minta, T. L. Manley, and K. E. Aune. 1994. Spatial and temporal interactions of male and female grizzly bears in northwest Montana. *Journal of Wildlife Management* **61**:39-52.
- Miller, S. D., P. Joyce, and L. P. Waits. 2002. Assessing allelic dropout and genotype reliability using maximum likelihood. *Genetics* **160**:357-366.
- Mills, L. S., J. J. Citta, K. Lair, M. Schwartz, and D. Tallmon. 2000. Estimating animal abundance using noninvasive DNA sampling: promise and pitfalls. *Ecological Applications* **10**:283-294.
- Mowat, G., and D. Paetkau. 2003. Estimating marten *Martes americana* population size using hair capture and genetic tagging. *Wildlife Biology* **8**:201-209.
- Opseth, O. 1998. Brown bear (*Ursus arctos*) diet and predation on moose (*Alces alces*) calves in the southern taiga zone in Sweden. *Candidatus Scientarum thesis*. Norwegian University of Science and Technology, Trondheim.
- Paetkau, D., and C. Strobeck. 1994. Microsatellite analysis of genetic variation in black bear populations. *Molecular Ecology* **3**:489-495.
- Paetkau, D. 2003. An empirical exploration of data quality in DNA-based population inventories. *Molecular Ecology* **12**:1375-1387.
- Piggott, M. P., E. Bellemain, P. Taberlet, and A. C. Taylor. 2004. A multiplex pre-amplification method that significantly improves microsatellite amplification and error rates for faecal DNA in limiting conditions. *Conservation Genetics* **5**:417-420.
- Pollock, K. H., S. C. Winterstein, C. M. Bunck, and P. D. Curtis. 1989. Survival analysis in telemetry studies: the staggered entry design. *Journal of Wildlife Management* **53**:7-15.
- Roth, H. U. 1980. Defecation rates of captive brown bears. *Proceedings of the International Conference on Bear Research and Management* **4**:249-253.
- Saether, B. E., S. E. Engen, J. E. Swenson, O. Bakke, and F. Sandergrén. 1998. Assessing the viability of Scandinavian brown bear, *Ursus arctos*, populations: the effects of uncertain parameters estimates. *Oikos* **83**:403-416.
- Schwartz, M. C., D. Tallmon, and G. Luikart. 1999. Using genetics to

- estimate the size of wild populations: many methods, much potential, uncertain utility. *Animal Conservation* **2**:321–323.
- Schwartz, C. C., S. D. Miller, and M. Haroldson. 2003. Grizzly bear. Pages 556–586 in G. Feldhamer, B. Thompson, and J. Chapman, editors. *Wild Mammals of North America*. The Johns Hopkins University Press, Baltimore, Maryland.
- Seber, G. A. F. 1982. The estimation of animal abundance and related parameters. 2nd edition. Macmillan, New York.
- Smith, K. L., et al. 2000. Cross-species amplification, noninvasive genotyping, and non-mendelian inheritance of human STRPs in savannah baboons. *American Journal of Primatology* **51**: 219–227.
- Swenson, J. E., T. M. Heggberget, P. Sandström, P. Wabakken, A. Bjärvall, A. Söderberg, R. Franzén, J. D. C. Linnell, and R. Andersen. 1996. Brunbjørnens arealbruk i forhold til menneskelig aktivitet. En utredning foretatt i forbindelse med Forsvarets planer for Regionfelt Østlandet, del 5. Norwegian Institute of Nature Reserves, Oppdragsmelding. (Use of area by brown bears in relation to human activity: a study undertaken in connection with the proposal by the Department of Defense to establish a shooting range in southeastern Norway, part 5 [in Norwegian].)
- Swenson, J. E., F. Sandegren, A. Bjärvall, and P. Wabakken. 1998a. Living with success: research needs for an expanding brown bear population. *Ursus* **10**:17–23.
- Swenson, J. E., F. Sandegren, and A. Söderberg. 1998b. Geographic expansion of an increasing brown bear population: evidence for pre-emption dispersal. *Journal of Animal Ecology* **67**:819–826.
- Taberlet, P., J.-J. Camarra, S. Griffin, O. Hanotte, L. P. Waits, C. Dubois-Paganon, T. Burke, and J. Bouvet. 1997. Noninvasive genetic tracking of the endangered Pyrenean brown bear population. *Molecular Ecology* **6**:869–876.
- 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.
- Taberlet, P., L. P. Waits, and G. Luikart. 1999. Noninvasive genetic sampling: look before you leap. *Trends in Ecology & Evolution* **14**:323–327.
- Valière, N. 2002. Gimlet: a computer program for analyzing genetic individual identification data. *Molecular Ecology Notes* **2**:377.
- Waits, L., P. Taberlet, J. E. Swenson, and F. Sandegren. 2000. Nuclear DNA microsatellite analysis of genetic diversity and gene flow in the Scandinavian brown bear (*Ursus arctos*). *Molecular Ecology* **9**:421–431.
- 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.
- Wilson, G. J., A. C. Frantz, L. C. Pope, T. J. Roper, T. A. Burke, C. L. Cheeseman, and R. J. Delahay. 2003. Estimation of badger abundance using faecal DNA typing. *Journal of Applied Ecology* **40**:658–666.
- White, G. C., and K. P. Burnham. 1999. Program MARK: Survival estimation from populations of marked individuals. *Bird Study Supplement* **46**:120–138.

Appendix 1. Minimum Population Size Estimate Based on the Number of Marked Females in Estrus.

$$N = F + M + \text{cubs} + \text{year} + \text{two}$$

$$F = \frac{FO}{PFO}$$

$$M = F$$

$$\text{cubs} = \text{REPR} * F$$

$$\text{year} = \text{cubs} * \text{Scubs}$$

two = YEAR * Syear, where N = estimated total population size, F = total number of females ≥ 3 years, FO = estimated number of females in oestrus, PFO = proportion of females in oestrus, M = total number of males ≥ 3 years, cubs = number of cubs of the year, Scubs = cub survival rate, REPR = reproductive rates among adult females, YEAR = number of yearlings, Syear = yearling survival rate, and two = number of 2-year-olds.

