DISSERTATION

STATISTICAL ASPECTS OF USING GENETIC MARKERS FOR INDIVIDUAL IDENTIFICATION IN CAPTURE-RECAPTURE STUDIES

Submitted by

Paul M. Lukacs

Department of Fishery and Wildlife Biology

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, CO

Spring 2005

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

UMI Number: 3173042

INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

UMI®

UMI Microform 3173042

Copyright 2005 by ProQuest Information and Learning Company.
All rights reserved. This microform edition is protected against unauthorized copying under Title 17, United States Code.

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

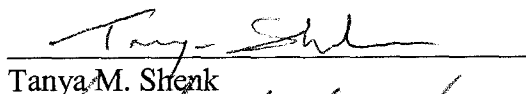
Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

COLORADO STATE UNIVERSITY

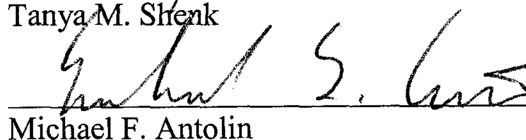
January 25, 2005

WE HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER OUR SUPERVISION BY PAUL M. LUKACS ENTITLED STATISTICAL ASPECTS OF USING GENETIC MARKERS FOR INDIVIDUAL IDENTIFICATION IN CAPTURE-RECAPTURE STUDIES BE ACCEPTED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY.

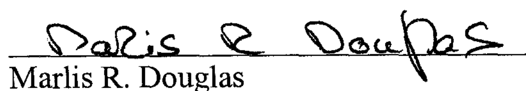
Committee on Graduate Work



Tanya M. Shenk



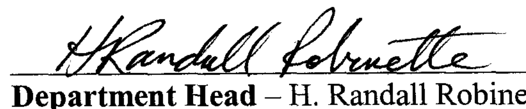
Michael F. Antolin



Marlis R. Douglas

Gary C. White

Adviser – Kenneth P. Burnham



Department Head – H. Randall Robinette

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

ABSTRACT OF DISSERTATION

STATISTICAL ASPECTS OF USING GENETIC MARKERS FOR INDIVIDUAL IDENTIFICATION IN CAPTURE-RECAPTURE STUDIES

The use of an animal's genotype as a mark for capture–recapture studies has become increasingly common in wildlife research. Frequently, animals are sampled in a non-invasive way such as collecting hair from rubs, shed feathers or feces. DNA is extracted from the samples and genotyped at a set of microsatellite loci to identify individuals. In this case, the animal's mark is self assigned. This leads to problems with misidentification of individuals. In addition, samples are often passively collected which causes a break down in the standard capture–recapture assumption of instantaneous sampling.

This dissertation focuses on developing new ways of handling the unique circumstances associated with DNA-based capture–recapture studies. I develop a method to account for genotyping error in closed population estimates of abundance. This method is used to extent the standard likelihood-based closed population capture–recapture models, the finite mixture models and the conditional likelihood models. I extend the robust design model to properly estimate survival and abundance in the face of genotyping error and allow multiple sources of data to be brought together. A new method is developed that uses additional information in dung surveys generated from

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

multiple detections of an individual within a sampling occasion to better estimate abundance.

I present suggestions for applying DNA-based capture–recapture sampling for Canada lynx (*Lynx canadensis*) in Colorado. Sampling feces in summer appears to be the most effective way to sample the lynx population. Feces are easier to collect in the field than hair and are more amenable to DNA extraction.

Finally, I present a review of the literature applicable to DNA-based capture–recapture studies. The focus of the review is to provide a starting point for researchers looking to broaden the scope of their DNA-based capture–recapture study beyond simple estimates of abundance.

Paul M. Lukacs
Department of Fishery and Wildlife Biology
Colorado State University
Fort Collins, CO 80523
Spring 2005

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

ACKNOWLEDGMENTS

There are many people to whom I am grateful for their help throughout my dissertation work. First, I would like to thank my advisor, Dr. Ken Burnham, for supporting my project and always being so excited when talking about this work with others. I also like to thank my graduate committee members, Dr. Gary White, Dr. Marlis Doulgas, Dr. Michael Antolin, and Dr. Tanya Shenk for their input on my project. In particular, Tanya provided the insight to realize that my project would have a long-term benefit for her lynx work. Gary allowed me to use his speedy dual-pro which helped in the bigger simulations. I would like to thank the Colorado Division of Wildlife and Great Outdoors Colorado for funding the research. Dr. Lori Eggert provided the African elephant dung data for chapter 3. Dr. Sara Oyler-McCance provided the Gunnison sage-grouse data for Appendix B. Brian Dreher provided black bear data that I frequently use in examples of the work in chapter 1. Finally, I would like to thank my wife for supporting me through the whole process.

Paul M. Lukacs

January 2005

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

TABLE OF CONTENTS

Chapter 1 1
Estimating population size from DNA-based closed capture–recapture data incorporating genotyping error

Chapter 2 24
Estimating survival, temporary emigration and population size in the robust design from DNA-based capture–recapture data incorporating genotyping error and multiple data sources

Chapter 3 52
Estimating population size from dung-based DNA capture–recapture data

Chapter 4 71
Sampling strategies for monitoring a reintroduced Canada lynx population in Colorado

Chapter 5 88
A review of capture–recapture methods applicable to DNA-based noninvasive sampling

Appendix A 112
Capture–recapture theory when genotypes are not unique

Appendix B 115
Suggestions for DNA-based capture–recapture on Gunnison sage-grouse

Appendix C 119
Microsatellite genotypes of lynx in Colorado

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

CHAPTER 1

ESTIMATING POPULATION SIZE FROM DNA-BASED CLOSED CAPTURE–RECAPTURE DATA INCORPORATING GENOTYPING ERROR

Abstract: Non-invasive samples of animals such as hair, feathers or scat are commonly collected for studies of wildlife populations. DNA is extracted from the samples and genotyped to identify individuals. The identifications of individuals are used as capture–recapture data to estimate population size. Some uncertainty exists in properly identifying individuals from DNA samples. I present full likelihood and conditional likelihood capture–recapture models for demographically and geographically closed populations which estimate population size in the presence of mis-identification of individuals. In addition, I extend the full likelihood capture–recapture models with mixture distributions to help account for heterogeneity in capture probability. The models presented here overcome the over-estimation problem present in standard capture–recapture estimation when errors exist in genotypes.

Key words: capture–recapture, closed population models, DNA markers, microsatellite, non-invasive sampling, population size estimation, tag-misread.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Animal identification based on DNA samples and microsatellite genotypes is widely used for capture–recapture studies (Woods et al. 1999, Boulanger et al. 2003, Eggert et al. 2003). The method shows promise in field protocols (Woods et al. 1999) and potentially minimal error rates in the DNA analysis (Paetkau 2003). Some studies show much higher error rates in individual identification (Creel et al. 2003). There will be some level of uncertainty, although in some situations the uncertainty level is small, in the identification of individuals from microsatellite genotypes.

Closed population capture–recapture analysis has received substantial attention over the past century (e.g. reviews by Otis et al. 1978, Schwarz and Seber 1999). More recently, it has been extended to conditional likelihood parameterizations that allow individual covariates to better estimate capture probability (Huggins 1989, 1991) and mixture models to estimate population size in the presence of individual heterogeneity in capture probability (Norris and Pollock 1996, Pledger 2000). The major focus of research has been developing methods to handle varying capture probability. Any methods developed in the future will also have to account for varying capture probability to obtain robust estimates of population size.

While DNA-based capture–recapture studies and standard tagging studies share several common characteristics, they differ in others. In a standard tagging study, the researcher attaches a unique tag to the animal and keeps a list of tags that have been used. In a DNA-based study, the genotype of the individual acts as the tag. Therefore, all individuals are tagged prior to the beginning of the study. Unfortunately, the researcher does not know what genotypes exist in the population and must obtain samples from the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

animals to extract DNA. In a standard tagging study, if a tag is read that does not match one known to be in the population, the researcher knows that the tag was incorrectly read and then either rereads the tag or ignores the observation. In DNA-based studies, the researcher does not have the luxury of immediately knowing which genotypes may be incorrect. Thus, a new form of sampling uncertainty is introduced. For both standard tagging and DNA-based studies, capture probability is <1.0. This necessitates a way to infer what portion of the population is not captured in order to determine the total population size. For a DNA-based study capture probability is a combination of the probability of encountering a sample (hair, scat, feather, etc.) and the probability that the sample yields a sufficient quantity and quality DNA to amplify and genotype.

Current closed population capture–recapture analysis for estimating population size assumes an animal's mark is permanent and read correctly when the animal is captured (Otis et al. 1978). The use of genotype-based identification can meet these assumptions in some situations, but the cost may be high. The cost comes in 2 pieces that clearly interact: (1) the monetary cost of analyzing the DNA and (2) the information loss when discarding samples that contain some degree of uncertainty in their identification. For example, the protocol described by Paetkau (2003) places a high emphasis on certainty of the genotype of the sample. In doing so, a large number of samples may be culled during the analysis. It may be beneficial to allow a small degree of uncertainty in the identification of a sample, perhaps 1 – 5%, if such a tradeoff would allow enough additional samples to be used in the estimation of population size to make up for the addition of a parameter to model this uncertainty.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

When genotyping error exists, it has been shown that population size estimates derived from capture–recapture assuming no error are biased (Waits and Leberg 2000, Mills et al. 2000, Creel et al. 2003). Creel et al. (2003) demonstrated that the error level they have in their wolf (*Canis lupus*) data could produce population size estimates that were biased by 5.5 times the true population size. Waits and Leberg (2000) also found large overestimation when genotyping errors were present. In addition, the authors showed underestimation in population size when multiple individuals share the same genotype. Mills et al. (2000) show in detail the underestimation effects of individuals sharing genotypes. All of these conclusions are logical when one considers that the statistical inference being made in a capture–recapture study is to the number of genotypes in the population. If errors are being made, there will be more genotypes observed than individuals in the population. If multiple individuals share the same genotype, there will be more individuals in the population than genotypes.

Currently, many studies collect far more hair, feathers, or scat samples than they can afford to have genotyped. Therefore, the limiting factor for sample size is often funding rather than a lack of sampled DNA. During the analysis, some samples are culled due to lack of confidence in genotyping results. It would be beneficial to be able to cull fewer samples and therefore increase usable sample size while taking into account the possibility of errors in genotyping.

I present a class of models for estimating the size of a demographically and geographically closed population when there exists some probability of mis-identifying individuals. The mis-identification occurs in such a way that it is unknown on a case by

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

case basis if the sample is correctly identified or not. The method extends the full likelihood models of Otis et al. (1978) and the conditional likelihood models of Huggins (1991). In addition, mixture models similar to those of Pledger (2000) can be built that estimate both heterogeneity in capture probability and mis-identification.

METHODS

The notation presented here follows Otis et al. (1978) and Pledger (2000) where applicable. The model assumes there are t sampling occasions, the population of interest is well defined, and the population is demographically and geographically closed during sampling. The following notation will be used to describe the models:

p_{ia} probability of initially observing a genotype at time i and from mixture component a, the second subscript is omitted for models not including mixtures.

c_{ia} probability of subsequently observing a genotype at time i and from mixture component a, the second subscript is omitted for models not including mixtures.

π_a probability of a genotype belonging to an animal in mixture a

α probability that a genotype is identified correctly given it is observed for the first time

f_0 the number of genotypes in the population that are never observed

N population size

$\mathbf{h} = \{h_1, h_2, ..., h_t\}$ encounter history vector $h_i = 1$ if the genotype is observed, 0 otherwise

M_{t+1} number of distinct genotypes observed

$n_{\mathbf{h}}$ a count of the number of times encounter history $\mathbf{h}$ is observed

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

A the number of mixtures in a given model, usually 1 or 2. $A = 1$ corresponds to the case of no mixtures.

I assume a set of loci, currently microsatellites, are being used that have enough loci and enough alleles per locus to ensure with high probability that all individuals within the population are unique if correctly genotyped. I further assume that a genotyping error will lead to a genotype that is not identical to any member of the population. In addition, 2 errors made at different trapping occasions are assumed to never produce identical genotypes. These assumptions were asserted as reasonable by Paetkau (2003) and are further addressed in the discussion.

Given the above assumptions, I computed the probability of each encounter history. If a genotype is first observed at time k and subsequently observed in the future, the probability of the encounter history is.

$$\Pr[\mathbf{h}] = \left[\prod_{i=1}^{k-1}(1-p_i)\right]\left[p_k\alpha\right]\left[\prod_{j=k+1}^{t} c_j^{h_j}(1-c_j)^{1-h_j}\right].$$

The expression states that the genotype was not observed from occasions 1 to $k-1$ with probability $(1-p_i)$. It is observed with probability p_k. The genotype is correctly assigned with probability α. Subsequently, it is detected or not detected at each occasion with probability c_i or $(1-c_i)$ respectively. For a genotype that is only observed at occasion k and never seen again the probability of the encounter history is:

$$\Pr[\mathbf{h}] = \left[\prod_{i=1}^{k-1}(1-p_i)\right]\left[p_k\alpha\left(\prod_{j=k+1}^{t}(1-c_j)\right) + p_k(1-\alpha)\right].$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Heuristically, the probability expression states that the genotype was not observed from occasions 1 to $k-1$ with probability $(1 - p_i)$. It is observed with probability p_k. Then it is either correctly genotyped with probability α and not seen again from occasions $k+1$ to t with probability $(1 - c_i)$, or it was incorrectly genotyped with probability $(1 - \alpha)$ and, by assumption, never seen again.

The full multinomial likelihood function can be constructed given the probability of each capture history. The likelihood is:

$$L(f_0, p_i, c_i, \alpha | n_{\mathbf{h}}, M_{t+1}) \propto \frac{(f_0 + M_{t+1})!}{f_0!} \prod_{\mathbf{h}} \Pr[\mathbf{h}]^{n_{\mathbf{h}}} .$$

I obtained parameter estimates by numerically optimizing the log–likelihood function. I used a quasi-Newton optimization function in SAS PROC IML (SAS Inc. 2002). The variance–covariance matrix can be obtained by numerically estimating the information matrix, inverting and taking its negative.

Note that N is not in this likelihood. I estimated N as a derived parameter. The closed capture–recapture models of Otis et al. (1978) were written equivalently with N or f_0 in the likelihood. The f_0 parameterization was chosen in modern software to easily enforce the constraint that abundance was greater than or equal to the total number of individuals captured such as is done in Program MARK (White and Burnham 1999). This constraint was necessary if marks were assumed to be correctly read, but it does not hold if marks can be read incorrectly. It is possible to observe more genotypes than are actually in the population. Therefore, I estimate N as:

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$\hat{N} = \hat{\alpha}\left(\hat{f}_0 + M_{t+1}\right).$$

The variance of $\hat{N}$ is estimated as:

$$\hat{\text{var}}[\hat{N}] = \hat{\alpha}^2\,\hat{\text{var}}[\hat{f}_o] + \left(\hat{f}_0 + M_{t+1}\right)^2\,\hat{\text{var}}[\hat{\alpha}] + 2\hat{\alpha}M_{t+1}\,\hat{\text{cov}}[\hat{f}_0, \hat{\alpha}] + \hat{\text{var}}[M_{t+1}]\ .$$

Having M_{t+1}^2, a potentially large positive number, in the variance was troubling, but $\hat{\text{var}}[\hat{\alpha}]$ is typically small because $\hat{\alpha}$ is often near one and the multinomial variance is therefore small. In addition, the $\hat{\text{cov}}[\hat{f}_0, \hat{\alpha}]$ is typically small and sometimes negative keeping $\hat{\text{var}}[\hat{N}]$ on the same order of magnitude as the closed capture–recapture models not incorporating recaptures.

The parameters p, c, and α can be modeled as functions of group covariates as is commonly done in generalized linear models (McCullagh and Nelder 1989) and in Program MARK (White and Burnham 1999). The α parameter should almost always be modeled with a sine link because it will be very near the boundary of 1.0 in many studies. The sine link allows for better estimation of the number of estimable parameters and of the shape of the log–likelihood function at its maximum while constraining the parameter to be within [0 – 1] than a logit function (White and Burnham 1999).

I construct a conditional multinomial likelihood function by conditioning on the probability that an animal is never captured, similar to the models of Huggins (1989, 1991). Therefore, I remove f_0 from the likelihood. The probability of an encounter history is now:

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$\Pr[\mathbf{h}_c] = \frac{\Pr[\mathbf{h}]}{1 - \left[\prod_{i=1}^{t} (1 - p_i)\right]} .$$

The likelihood function is:

$$L(p_i, c_i, \alpha | n_{\mathbf{h}}) \propto \prod_{\mathbf{h}_c} \Pr[h_c]^{n_{\mathbf{h}}} .$$

Again N is a derived parameter. It is estimated as:

$$\hat{N} = \hat{\alpha} \sum_{i=1}^{M_{t+1}} \left[\frac{1}{1 - \prod_{j=1}^{t} (1 - \hat{p}_{ij})} \right] .$$

The large sample estimated variance is:

$$\hat{\mathrm{var}}[\hat{N}] = \begin{bmatrix} \frac{\partial N}{\partial \alpha} & \frac{\partial N}{\partial p_1} & \cdots & \frac{\partial N}{\partial p_t} \end{bmatrix} \begin{bmatrix} \hat{\mathrm{var}}[\hat{\alpha}] & \hat{\mathrm{cov}}[\hat{\alpha}, \hat{p}_1] & \cdots & \hat{\mathrm{cov}}[\hat{\alpha}, \hat{p}_t] \\ \hat{\mathrm{cov}}[\hat{p}_1, \hat{\alpha}] & \hat{\mathrm{var}}[\hat{p}_1] & & \hat{\mathrm{cov}}[\hat{p}_1, \hat{p}_t] \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\mathrm{cov}}[\hat{p}_t, \hat{\alpha}] & \hat{\mathrm{cov}}[\hat{p}_t, \hat{p}_1] & \cdots & \hat{\mathrm{var}}[\hat{p}_t] \end{bmatrix} \begin{bmatrix} \frac{\partial N}{\partial \alpha} \\ \frac{\partial N}{\partial p_1} \\ \vdots \\ \frac{\partial N}{\partial p_t} \end{bmatrix} .$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The variance can be computed numerically or analytically.

The conditional likelihood models allow parameters to be modeled as functions of individual covariates. Given the observations are generally collected through noninvasive methods, many standard individual covariates that can affect capture probability will not be collected, for example length of fish captured with electro-fishing equipment. An interesting covariate for α may be a measure of the quality of the DNA sample collected for the first time the genotype is observed.

The full likelihood models can be extended to incorporate heterogeneity in capture probability with a mixture distribution (Pledger 2000). The full likelihood probability of an encounter history with mixtures for a genotype that is encountered more than 1 time is

$$\Pr[\mathbf{h}] = \sum_{a=1}^{A} \pi_a \left[\prod_{i=1}^{k-1} (1 - p_{ia}) \right] \left[p_{ka} \alpha \right] \left[\prod_{j=k+1}^{t} c_{ja}^{h_j} (1 - c_{ja})^{1-h_j} \right].$$

For a genotype that is only observed once, the probability of the encounter history is

$$\Pr[\mathbf{h}] = \sum_{a=1}^{A} \pi_a \left[\prod_{i=1}^{t-1} (1 - p_{ia}) \right] \left[p_{ta} \alpha \left(\prod_{j=t+1}^{t} (1 - c_{ja}) \right) + p_{ka} (1 - \alpha) \right],$$

where $\sum_{a=1}^{A} \pi_a = 1$. The likelihood, estimation of N and the variance of N follow the full likelihood results presented above. I used a trust region optimization in SAS PROC IML (SAS Institute, Inc. 2002) to fit the model.

I compared the genotype mis-identification models presented here with standard

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

closed capture–recapture models for a simulated closed population experiencing genotyping error. I used models representing several forms of variation in detection probability including changes in behavior due to previous encounter, changes across time, and constant capture probability (Table 1). The data were simulated in a factorial assignment with five levels of α ranging from 0.95 – 0.99 and five levels of constant capture probability ranging from 0.1 – 0.5. Five sampling occasions were used for each population. Capture probabilities ranging from 0.1 – 0.5 for each of five sampling occasions cover a large part of the range of possible levels of encounters from very few animals encountered to nearly all animals encountered. Each design point was replicated 200 times. SAS code used for simulation and estimation is available from the authors.

I did not include a constant capture probability form (model M_0 of Otis et al. [1978]) of the genotype misread model in the analysis because it was not reasonable to expect the initial capture probability to equal the recapture probability when initial captures include both correct and incorrect genotypes while recaptures were only correct genotypes. It is important to note that p is the probability of observing a genotype correctly or incorrectly, whereas c is the probability of correctly observing a genotype that has been seen at a previous occasion. The time varying capture probability model has the same logical flaw as the constant capture probability model for the genotype misread models, but it was included. One could argue that the time varying capture probability model allows enough flexibility to be reasonable if true p is nearly equal to c.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

RESULTS

In most situations, the full likelihood mis-identification models had lower bias in estimating population size than the standard capture–recapture models (Figure 1). The percent bias in estimated population size was often half as large for the mis-identification models as it was the standard capture–recapture models. Bias was worst for the genotyping error models when capture probability was 0.1. This results from a lack of recaptures due to the low capture probability. As capture probability increased, bias quickly decreased.

Confidence interval coverage of estimated population size was near the nominal 95% level for the mis-identification models with $p \leq 0.3$. Confidence interval coverage was well below the nominal 95% when mis-identification was present, but not estimated (Figure 2). Confidence interval coverage broke down for both the standard capture–recapture models and the mis-identification capture–recapture models when capture probability was >0.3 and misidentification is present.

The estimation of genotyping error performed poorly when capture probability was 0.1. This result would also hold for capture probabilities less than 0.1 given 5 sampling occasions were used. When capture probability was 0.1 an animal only has a 0.08 probability of being caught more than once. Recaptures are required to estimate the probability of correctly identifying a genotype. The low number of recaptures caused the genotyping error model containing both time variation and a behavioral response in capture probability to occasionally fail to converge to a reasonable estimate of population size when capture probability was 0.1.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

When capture probability was 0.5 and sampling occurs on 5 occasions, approximately 97% of animals were detected. Thus, nearly all individuals were expected to be captured at least once and 81% were expected to be caught more than once. Therefore, there were few encounter histories with only a single observation, so it was difficult to estimate genotyping error rate effectively. This is a minor issue because such a situation is rarely feasible in the field, and when it occurs the confidence interval width is very small, and bias is trivial (1 – 4%). The estimate is typically only off of the true value by a few animals.

DISCUSSION

I make three assumptions beyond those needed for a standard capture–recapture study to estimate genotyping error rate from capture–recapture data. The assumptions are relatively easily met in real world problems. I assume that a set of loci are used that contain enough genetic information such that each individual is uniquely identified if the genotype is correctly read. For a wide range of species these systems exist, such as bears (*Ursus* spp.; Paetkau 2003), elephants (*Loxodonta cyclotis;* Eggert et al. 2003), Canada lynx (*Lynx canadensis;* Schwartz et al. 2003), sage-grouse (*Centrocercus urophasianus;* Taylor et al. 2003), and more are being developed continuously. Resolution power can be assessed prior to beginning a study by computing the probability of identity for the marker set (Waits et al. 2001). Therefore, it is not difficult to obtain the power necessary to discriminate among individuals when no errors are present.

I assume that any error in genotyping will result in a genotype that does not match another individual in the population of interest. Far more genotypes are possible than

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

individuals that exist in many wildlife species. For example, a set of six loci each with three alleles has 46,656 possible genotypes. Therefore, the chance of an error resulting in an existing genotype rather than some other genotype is quite low. If this assumption is violated and an error results in an existing genotype, there is only a trivial effect on the estimation of population size from a closed capture–recapture model. If the error results in an animal that has never been seen before then that animal is no longer at risk of capture and the animal that was truly caught remains at risk of capture. Thus M_{t+1} and $\hat{p}$ are virtually unchanged, hence and $\hat{N}$ is also nearly unchanged. For example, consider the case of the simplest form of a capture–recapture model with time varying capture probability, the Lincoln–Petersen estimator. Three quantities are needed to estimate abundance from this model: (1) the number caught in the first sample (n_1), (2) the number caught in the second sample (n_2), and (3) the number of marked animals caught in the second sample (m_2). If 50 animals are caught at each sample and 25 of the animals caught during the second sample are recaptures, then the estimated abundance is 50 × 50 / 25 = 100. Now if one of the animals caught on the first occasion is mis-identified as an animal that is not caught on the first occasion, but the animal is caught on the second occasion, then none of the statistics change and the population estimate remains 100. Thus, a violation of this assumption is trivial.

I assumed that errors were never repeated in exactly the same way to produce an identical incorrect genotype. This assumption can be violated in 2 ways. First, an individual can be sampled twice and incorrectly genotyped in the same way twice.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Paetkau (2003) showed that occurred in about 15% of his samples that were incorrectly genotyped. For capture–recapture this is a minor issue because the individual is still correctly matched across samples even though the genotype is not correct. Second, two different individuals could both be incorrectly genotyped and coincidentally produce a matching incorrect genotype. A large number of factors would all have to happen each with low probability to generate the same genotype incorrectly twice from independent samples (Paetkau 2003). Therefore, this assumption may be the weakest of the 3 assumptions made here, but the consequences of a violation of the assumption are minor.

Estimation of the model parameters, and most importantly $\hat{N}$, is good when α is high (≥ 0.95). This is reasonable performance for the estimator. Lab protocols can easily keep error rates within that range (Paetkau 2003). Yet, even with an error rate of only 1% per sample, substantial bias in $\hat{N}$ can occur if it is not taken into account. The mis-identification models perform well when error rates exceed 5%, but the variance on estimated population size becomes large and quickly makes the results relatively uninformative about population size. Despite the large variance on population size, one would learn about the high error rate. The cutoff for an allowable level of genotyping error depends on the precision required for the study.

The models presented here rely on recaptures to estimate the probability of correctly genotyping a sample. Heuristically, this quantity is estimated by the imbalance in the number of genotypes observed only once to those observed more than once. If error is present in genotyping, there will be an excess of genotypes observed only once.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

A portion of these (α) are truly seen only once and the remainder ($1 - \alpha$) are seen once because the genotypes are mistakes. There are two ways to ensure obtaining recaptures: one way is to sample intensively to get capture probability high, and the other way is to sample on more occasions.

The capture–recapture models presented here are based on the same likelihood function as the standard capture–recapture models they generalize. Therefore, model selection criteria such as AIC_c can be used to compare models (Burnham and Anderson 2002). The researcher need not worry whether they should or should not use a model incorporating genotyping error. Estimation can be performed with both standard capture–recapture models and the genotyping error models presented here, and AIC_c will determine which model is better supported by the data. Resulting abundance estimates may be model averaged to further reduce the effects of model selection bias (Burnham and Anderson 2002).

Although these models were developed in the context of identification from microsatellite data, the concept extends to other types of analyses as well. Identification based on photographs is commonly attempted in marine mammal studies (Jefferson 2000). These models could be used for that application as well. The models presented here do not fit well with mis-identification of physical tags because the set of tags available for capture should be known. Therefore, if a tag is read that does not match a tag in the population, it is known to be incorrect.

The models presented here are most applicable to smaller populations up to several thousand individuals. In theory, the method applies to any size population.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Populations that are very large are quite expensive and time consuming to sample with DNA based methods due to the large number of samples that would need to be processed and large number of loci needed to resolve individuals.

MANAGEMENT IMPLICATIONS

Ignoring genotyping error when it is in fact present will lead to over-estimation of animal abundance. DNA-based capture–recapture is typically used on species with small population sizes that are difficult to observe. Over-estimating the size of a small population could lead to potentially detrimental conclusions for an endangered or exploited species. The methods I presented allow genotyping error rate to be directly estimated and abundance appropriately corrected. Therefore, management decisions can be made based on an accurate assessment of the population size.

ACKNOWLEDGMENTS

The Colorado Division of Wildlife, Great Outdoors Colorado, and Dr. Tanya Shenk provided support for this research. Dr. Gary White, Mr. Eric Bergman, Dr. Robert Steidl, and an anonymous referee provided comments on the manuscript.

LITERATURE CITED

Boulanger, J., G. C. White, B. N. McLellen, J. Woods, M. Proctor, and S. Himmer. 2003. A meta-analysis of grizzly bear DNA mark-recapture projects in British Columbia, Canada. Ursus 13: 137–152.

Burnham, K. P., and D. R. Anderson. 2002. Model selection and multimodel inference: a practical information-theoretic approach. Springer-Verlag, New York, New York, USA.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

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.

Eggert, L. S., J. A. Eggert, and D. S. Woodruff. 2003. Estimating population sizes for elusive animals: the forest elephants of Kakum National Park, Ghana. Molecular Ecology 12: 1389–1402.

Huggins, R. M. 1989. On the statistical analysis of capture experiments. Biometrika 76: 133–140.

_____. 1991. Some practical aspects of a conditional likelihood approach to capture experiments. Biometrics 47: 725–732.

Jefferson, T. A. 2000. Population biology of the Indo-Pacific hump-backed dolphin in Hong Kong waters. Wildlife Monographs 144.

McCullagh, P., and J. A. Nelder. 1989. Generalized linear models. Second edition. Chapman Hall, New York, New York, USA

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.

Norris, J. L., and K. H. Pollock. 1996. Nonparametric MLE under two closed capture–recapture models with heterogeneity. Biometrics 52: 639–649.

Otis, D. L., K. P. Burnham, G. C. White, and D. R. Anderson. 1978. Statistical inference from capture data on closed animal populations. Wildlife Monographs 62.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Paetkau, D. 2003. An empirical exploration of data quality in DNA-based population inventories. Molecular Ecology 12: 1375–1387.

Pledger, S. 2000. Unified maximum likelihood estimates for closed capture–recapture models using mixtures. Biometrics 56: 434–442.

SAS Institute Inc. 2002. SAS version 9. Carey, North Carolina, USA.

Schwartz, M. K., L. S. Mills, Y. Ortega, L. F. Ruggiero, and F. W. Allendorf. 2003. Landscape location affects genetic variation of Canada lynx (*Lynx canadensis*). Molecular Ecology 12: 1807–1816.

Schwarz, C. J. and G. A. F. Seber. 1999. Estimating animal abundance: review III. Statistical Science 14: 427–456.

Taylor, S. E., S. J. Oyler-McCance, T. W. Quinn. 2003. Isolation and characterization of microsatellite loci in Greater Sage-Grouse (*Centrocercus urophasianus*). Molecular Ecology Notes 3: 262–264.

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., and K. P. Burnham. 1999. Program MARK: survival estimation from populations of marked animals. Bird Study 46: S120–S139.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

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.

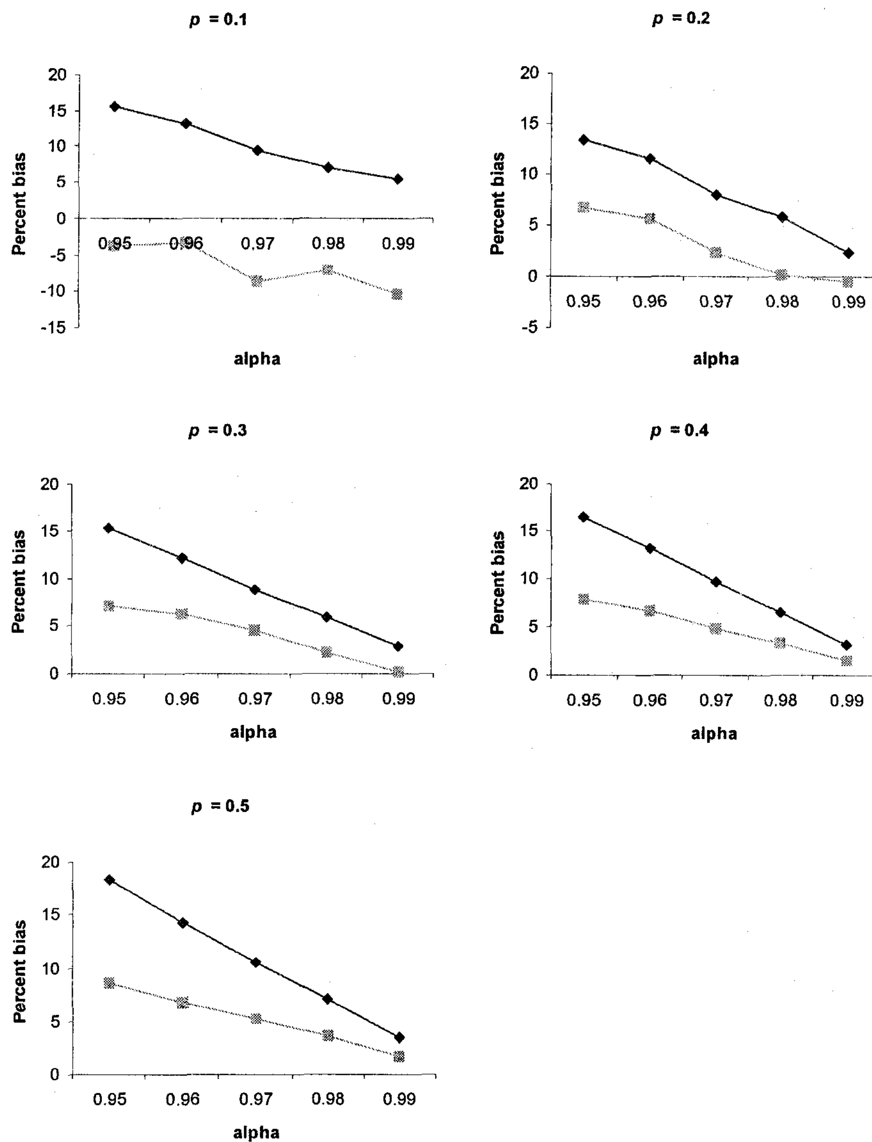
Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 1. Models representing different forms of variation in capture probability and whether they are used in capture–recapture models incorporating genotyping error. Sources of variation include changes in capture probability due to a behavioral response to first encounter, changes across time, and constant capture probability.

		Including variation attributable to genotype	
Model	Variation in capture probability	Yes	No
M_0	Constant		X
M_b	behavior	X	X
M_t	time	X	X
M_{t+b}	time + behavior	X	

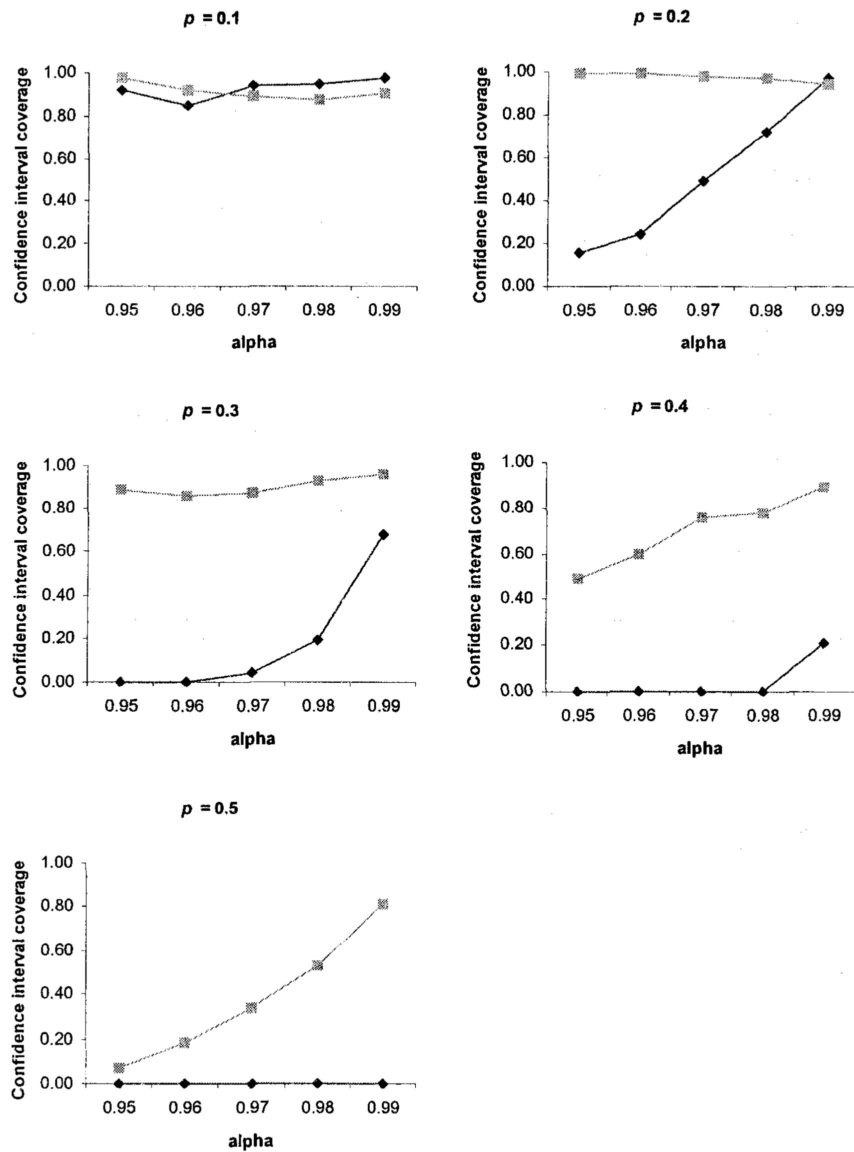
Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Figure 1. Percent bias in estimated abundance for closed capture–recapture models including an estimate of genotyping error (gray squares) and not accounting for genotyping error (black diamonds) across 5 levels of capture probability (p) and 5 levels of the probability of correctly genotyping a sample (alpha). Percent bias is averaged across models containing different forms of variation in capture probability. Each design point was replicated 200 times. The coefficient of variation for all bias estimates is <1%.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Figure 2. Confidence interval coverage of estimated abundance for closed capture–recapture models including an estimate of genotyping error (gray squares) and not accounting for genotyping error (black diamonds) across 5 levels of capture probability (p) and 5 levels of the probability of correctly genotyping a sample (alpha). Confidence interval coverage is averaged across models containing different forms of variation in capture probability. Each design point was replicated 200 times.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

CHAPTER 2

ESTIMATING SURVIVAL, TEMPORARY EMIGRATION AND POPULATION SIZE IN THE ROBUST DESIGN FROM DNA-BASED CAPTURE–RECAPTURE DATA INCORPORATING GENOTYPING ERROR AND MULTIPLE DATA SOURCES

***Abstract*:** For many species, non-invasive sampling of hair, scat or other tissue has the potential to be very useful and in some cases is already widely used to answer ecological questions. These samples are genotyped and the genotypes are used to identify individuals. There is some level of uncertainty when identifying individuals from genotyping results. I present an extension to the robust design capture–recapture model (Kendall et al. 1997) that allows for the estimation of genotyping error rate and properly estimates population size, survival, temporary emigration, and capture probability in the face of genotyping error. I further extend the model to allow estimation of all of these parameters from multiple sources of data including radio telemetry and laboratory analysis. The models presented here extend to other forms of capture–recapture which involve some level of uncertainty in individual identification such as photographic identification.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Keywords**:** abundance, capture–recapture, microsatellites, non-invasive sampling, tag misread

INTRODUCTION

Non-invasively sampled feathers, hairs and scats are being widely collected to answer ecological questions (Eggert et al. 2003, Kohn et al. 1999, Palsbøll et al. 1997 among others). DNA can be extracted from cells found in these samples. One use for these data are to obtain a microsatellite genotype of the animal which left the sample in order to identify individuals. This can be a very powerful method to study species which are expensive and dangerous to handle or are very difficult to capture. Thus far, DNA-based capture–recapture studies have been largely used for estimating abundance (Woods et al. 1999, Boulanger et al. 2003), but capture–recapture theory has been developed to estimate a wide range of ecologically interesting parameters such as survival, emigration rates and population growth rates. Therefore, it is important to broaden the scope at which this data are being applied.

Standard capture–recapture theory assumes that individual animals are correctly identified each time they are captured. For DNA-based studies, not all individuals are necessarily identified correctly. Errors may arise in genotyping that are caused by a number of sources including degradation of the sample, mutation within the PCR reaction, allelic dropout and other sources. When errors are present in genotyping, it will appear as if there are more individuals in the population than actually exist. It will also

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

appear as if survival rates are lower than they actually are because errant genotypes will never be observed again. Some studies are able to minimize error to a great degree (Paetkau 2003). Other studies show error rates that are much larger (Creel et al. 2003). Given that error in genotyping is a distinct possibility and its effect is to bias parameter estimates, it is important to be able to explicitly model the error rate in a capture–recapture model.

The central issue facing the use of genotypes as marks is that the researcher never knows the list of marks in the population. With conventional capture–recapture studies, no animals in the population are marked prior to the study. Then, each animal captured is given a unique mark which is subsequently recorded by the researcher. Therefore, the list of marks used is known. When genotypes are used as marks, every animal has a mark prior to the beginning of the study. Unfortunately, the researcher does not know what any of the marks are or even how many marks exist. This causes uncertainty in interpreting the mark of an individual because there is often no way to verify that the mark is read correctly. In a conventional capture–recapture study, if a researcher records a tag number that does not match a tag on the list, then the tag would either be reread or the capture would be ignored. Researchers involved in DNA-based capture–recapture studies do not have this luxury.

In Chapter 1, I developed a method to estimate genotyping error rate in demographically and geographically closed population capture–recapture models. The method uses three basic assumptions in order to develop a likelihood to estimate

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

genotyping error rate, population size and capture probability. First, the method assumes that a molecular marker set is used that has sufficient power to resolve individuals with a high degree of certainty (e. g. $P_{ID}<1x10^{-5}$ for a population size in the hundreds). Typically, these marker sets are microsatellites, but they can be single nucleotide polymorphisms or any other marker with the power to resolve individuals. Such marker sets exist for a large number of species and more are being developed. Second, it is assumed that a genotyping error will not result in a genotype that exists in the population. Given the typically small population sizes relative to the huge number of possible genotypes, it is unlikely for an error to result in the genotype of an individual that already exists in the population. Moreover, violation of this assumption has hardly any effect and in some cases does not change the sufficient statistics of a closed capture–recapture model and therefore will not change the estimates. Third, it is assumed that the same error will not be generated twice. A large number of rare events would have to occur for two errors to result in an identical erroneous genotypes. These assumptions are practical and can be realistically met in DNA-based capture–recapture studies. In addition to their use in capture–recapture, the assumptions have been asserted as useful in the examination of error in molecular techniques (Paetkau 2003).

I extend the results of Chapter 1 to a demographically and geographically open population model and sampling scheme often referred to as Pollock's robust design (Pollock 1982, Kendall et al. 1997). The robust design is composed of two types of sampling periods. Primary sampling periods are separated by a relatively long length of

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

time during which the population is assumed to be demographically and geographically open. Within each primary period are secondary periods which are very close together in time, so that the population can be assumed to be demographically and geographically closed over these intervals. The robust design allows the estimation of population size at each primary period, survival probability between primary periods, temporary emigration between primary periods, and capture probability during a secondary period. In addition, this extension allows for the estimation of genotyping error rate at each primary period.

I then further extend the robust design with genotyping error model to incorporate data from laboratory studies of genotyping error and radio-telemetry data for survival. The ability to combine multiple sources of data helps estimate parameters more precisely than could be done with any one data source analyzed alone. Moreover, much of this data are already being collected, therefore it would be most efficient to use the data in a single analysis.

STATISTICAL MODEL AND NOTATION

The notation largely follows Kendall et al. (1997) and Chapter 1. The robust design is a sampling scheme with t primary sampling periods. Each $i = 1, 2, \ldots, t$ of the primary periods contains l_i secondary sampling periods. The l_i secondary periods are not required to be equal in number. The population is assumed to be open between primary periods and closed between secondary periods within a primary period. An example of a design for a relatively long lived species could have $t = 5$ primary periods each separated

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

by one year and $l_1, \ldots, l_5 = 4$ secondary periods per primary period separated by one day (Figure 1).

Parameters

ϕ_i Probability of an animal surviving from primary period i to $i + 1$ given it is alive a i

γ_i' Probability of being off the sampling area at time i given the animal was off the area at $i - 1$

γ_i'' Probability of being off the sampling area at time i given the animal was on the area at $i - 1$

p_{ij} Probability of initially observing a genotype at secondary sample j of primary period i

c_{ij} Probability of subsequently observing a genotype at secondary sample j of primary period i

p_i* Probability of being observed at least once during primary period i.

$$p_i^* = 1 - \prod_{j=1}^{l_i} (1 - p_{ij})$$

α_i Probability that the first observation of a genotype in primary period i is correctly assigned

f_{i0} Number of genotypes in the population which are never observed at primary period i

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

N_i Population size at primary period i

Statistics

I Total number of secondary capture occasions. $I = \sum_{i=1}^{t} l_i$

$\mathbf{h} = \{h_1, h_2, ..., h_I\}$ Encounter history vector $h_i = 1$ if the genotype is observed, 0 otherwise

$\mathbf{h}_i$ Encounter history of the secondary samples within primary period i

$\mathbf{h}_\mathrm{p}$ Encounter history of the primary periods. $h_{\mathrm{p}i}$ equals the sum of the encounter history for the secondary occasions within primary period i.

x_{ij} An indicator variable that equals 1 if the genotype is observed at secondary occasion ij and 0 otherwise

$M_{i,t+1}$ Number of unique genotypes observed during primary period i

I use the same assumptions here as are used in Chapter 1, and discussed in the introduction, to estimate genotyping error rate within a primary period. The model is based on a multinomial likelihood and count data expressed as encounter histories (i. e., 1 if observed, 0 otherwise for each occasion). From those assumptions cell probabilities can be computed for each possible capture history and a likelihood can be built. The robust design likelihood is composed of two pieces, the likelihood for the open population primary periods, L_1, and the likelihood for the secondary periods within each primary period, L_2.

The likelihood for the secondary periods within each primary period can be developed from the cell probabilities of each possible encounter history. For genotypes

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

first observed at secondary occasion k and subsequently observed again within primary period i, the probability of the encounter history is

$$\Pr[\mathbf{h}_i] = \left[\prod_{j=1}^{k-1} (1 - p_{ij})\right]\left[p_{ik}\alpha_i\right]\left[\prod_{g=k+1}^{l_i} c_{ig}^{x_{ig}} (1 - c_{ig})^{1-x_{ig}}\right]$$

The expression states that the genotype was not observed from occasions 1 to $k - 1$ with probability $(1 - p_i)$. It is observed with probability p_k. The genotype is correctly assigned with probability α. Subsequently, it is detected or not detected at each occasion with probability c_i or $(1 - c_i)$ respectively. For a genotype which is only observed at secondary occasion k and never observed again, the probability of the encounter history is

$$\Pr[\mathbf{h}_i] = \left[\prod_{j=1}^{k-1} (1 - p_{ij})\right]\left[p_{ik}\alpha_i\left(\prod_{g=k+1}^{l_i} (1 - c_{ig})\right) + p_{ik}(1 - \alpha_i)\right].$$

Heuristically, the probability expression states that the genotype was not observed from occasions 1 to $k - 1$ with probability $(1 - p_i)$. It is observed with probability p_k. Then it is either correctly genotyped with probability α and not seen again from occasions $k + 1$ to t with probability $(1 - c_i)$, or it was incorrectly genotyped with probability $(1 - \alpha)$ and, by assumption, never seen again.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The multinomial likelihood for the secondary periods is the product of the likelihood of each secondary period within a primary period

$$L_2 = \prod_{i=1}^{t} \left[\frac{(f_{i0} + M_{i,t+1})!}{f_{i0}!} \prod_{\mathbf{h}_i} \Pr[\mathbf{h}_i]^{n_{\mathbf{h}_i}} \right].$$

The likelihood for the primary periods is also based on a multinomial distribution. For the case of $K = 4$ primary occasions, the cell probabilities for the first encounter after release are given in Appendix 1. From these cell probabilities, it is possible to construct an expression for an encounter history. Let **A** be an upper diagonal matrix of cell probabilities similar to those represented in Appendix 1 and let a_{ij} represent the elements of **A**. In addition, let b be the set of primary occasions on which the genotype is observed and subsequently seen again. Let r be the primary occasion of the last observation of the genotype. For example, for a genotype observed at times 1, 3 and 4, $b \subset \{(1,3), (3,4)\}$ and $r = 4$. For a genotype that is observed more than once, otherwise stated as the sum of the elements of $\mathbf{h}_p > 1$, the probability of the encounter history is

$$\Pr[\mathbf{h}_p] = \alpha_t \left[\prod_{(i,j)\in b} a_{ij} \right] \left(1 - \sum_{j=r}^{t} a_{rj} \right).$$

For a genotype that is only observed once (sum of $\mathbf{h}_p = 1$), the probability of the encounter history across primary periods is

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$\Pr[\mathbf{h}_p] = \alpha_t \left[1 - \sum_{j=r}^{t} a_{rj} \right] + (1 - \alpha_t) .$$

The multinomial likelihood across primary periods is

$$L_1 = \prod_{\mathbf{h}_p} \Pr[\mathbf{h}_p]^{n_{\mathbf{h}_p}} .$$

The full likelihood for the robust design incorporating genotyping error is given by the product of L_1 and L_2. Parameter estimates can be found by numerical optimization of the log-likelihood. I used a quasi-Newton optimization in SAS PROC IML (SAS Institute Inc. 2002). The variance-covariance matrix of the parameters can be computed by inverting the Hessian matrix and taking its negative. Also, profile likelihood intervals could be used to assess variation.

Up to this point, N_i has not appeared in the likelihood and it does not enter into the likelihood. N_i must be estimated as a derived parameter for this likelihood. It is estimated by primary period as

$$\hat{N}_i = \hat{\alpha}_i \left(\hat{f}_{i0} + M_{i,t+1} \right)$$

The variance of $\hat{N}_i$ is estimated as

$$\hat{\text{var}}[\hat{N}_i] = \hat{\alpha}_i^2 \, \hat{\text{var}}[\hat{f}_{i0}] + \left(\hat{f}_{i0} + M_{it+1} \right)^2 \hat{\text{var}}[\hat{\alpha}_i] + 2\hat{\alpha}_i M_{it+1} \, \hat{\text{cov}}[\hat{f}_{i0}, \hat{\alpha}_i] + \hat{\text{var}}[M_{it+1}] \ .$$

$$\hat{\text{var}}[\hat{N}_i] = \hat{\alpha}_i^2 \, \hat{\text{var}}[\hat{f}_{i0}] + \left(\hat{f}_{i0} + M_{it+1} \right)^2 \hat{\text{var}}[\hat{\alpha}_i] + 2\hat{\alpha}_i M_{it+1} \, \hat{\text{cov}}[\hat{f}_{i0}, \hat{\alpha}_i] + \hat{\text{var}}[M_{it+1}] \ .$$

The estimates of N across primary periods may covary depending on the model structure

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

of the capture probabilities and genotyping error rates.

Model parameters may be functions of group covariates as is commonly used in general linear models (McCullagh and Nelder 1989). The α parameter will typically be modeled on a sine link because it is near the boundary of 1.0 in many studies (White and Burnham 1999). If the likelihood of the secondary periods within a primary period is replaced with the conditional likelihood given in Chapter 1, parameters, except for *N*, may be modeled as functions of individual covariates as well as group covariates.

The robust design likelihood can be supplemented with additional sources of data to better estimate the probability of correctly assigning a genotype and survival probability. Laboratory results can be used to help estimate the probability of correctly assigning a genotype. Data are likely to exist from known individuals in the population. Genotypes from tissue or blood samples can be compared to those from hair or scat samples. The tissue or blood samples can be genotyped with virtually no error and therefore serve as a reference sample. The count of correctly and incorrectly matched hair or scat samples will follow a binomial distribution. Therefore, the likelihood for estimating the probability of correctly assigning a genotype is

$$L_3 = \prod_{i=1}^{t} \binom{n_i}{x_i} \alpha_i^{x_i} (1 - \alpha_i)^{n_i - x_i},$$

where n_i is the number of reference samples from primary period i, x_i is the number of correctly matched samples from primary period i and α_i is the probability of correctly

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

assigning the genotype as defined above. As with the robust design, α_i can be modeled as a function of covariates through a link function. A laboratory estimate of genotyping error adds considerable information and greatly reduces the variance in estimated α. Therefore, if it is possible to obtain some tissue samples from the same individuals as some of the hair or scat samples, this would help reduce the variance in estimated abundance. These samples may be available from individuals found dead or from hunters.

Many populations that are being sampled with noninvasive DNA capture–recapture also have a portion of the population being monitored with radio-telemetry. The telemetry data can provide information to help estimate survival in the robust design. In the absence of staggered entry between primary periods, survival across primary periods is simply a binomial with the number of animals radio-tagged and the number that survived the period. The likelihood is therefore

$$L_4 = \prod_{i=1}^{t} \binom{m_i}{y_i} \phi_i^{y_i} (1 - \phi_i)^{n_i - y_i},$$

where m_i is the number of radio tagged animals alive at primary period i and y_i is the number of animals from m_i remaining alive at i+1.

Now the full likelihood for the robust design with multiple sources of data are the product of the 4 likelihood pieces

$$L = L_1 L_2 L_3 L_4 .$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

If no laboratory data are available, L_3 equals 1 and if no telemetry data are available L_4 equals 1. Without laboratory data, there must be at least 3 secondary sampling occasions per primary occasion to estimate α, but with laboratory data available only 1 secondary sample is needed. Despite that, it is always beneficial to have multiple secondary occasions to better estimate all model parameters.

MODEL TESTING

I used simulated data to evaluate the properties of the robust design model incorporating genotyping error. The properties of the binomial likelihoods for the laboratory genotyping error data and the telemetry survival data (Pollock et al. 1989, White and Garrott 1990) are well established and perform well. Therefore, I concentrated model evaluation on only the robust design portion of the model. Capture–recapture data were simulated in a 3 factor factorial design with four levels of the probability of correctly genotyping a sample (0.96 – 0.99), five levels of capture probability (0.1 – 0.5) and two levels of survival probability (0.6, 0.9). Recapture probability was set equal to capture probability. Each of 40 design points was replicated 100 times due to long computing times. All simulations used 4 primary occasions and 5 secondary occasions within each primary occasion. Abundance was started at 1,000 individuals and no new individuals were added between primary periods. Population size, capture probability and recapture probability were estimated separately for each primary occasion and survival probability and the probability of correctly genotyping a sample were estimated as equal across primary occasions. The probability of temporary emigration was fixed to

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

zero for all simulations because it is known to be very difficult to estimate precisely and to allow clear inferences about the estimates of abundance and survival.

RESULTS

The models provided good parameter estimates in the simulated examples. Abundance at the first primary occasion was well estimated across varying levels of the probability of correctly genotyping a sample (Table 1). Abundance at the first primary period was the only abundance estimate summarized because the subsequent periods incorporate a random survival component into the true abundance. Therefore, abundance is different for each replicate after the first period. As would be expected, the estimates of abundance improved as the rate of genotyping error decreased. Survival was well estimated across the range of probabilities of correctly genotyping a sample (Table 2). Beyond the first primary period, abundance is a function of survival and therefore a fixed true value of abundance is not available to compare the estimated abundance to for future sampling periods.

Capture probability greater than 0.1 provided reliable estimates of abundance (Table 3). With a capture probability of 0.1, estimates of abundance were biased high (>12% bias, SE = 1.2), were imprecise and failed to converge in 29 of 900 replicates. As capture probability decreases below 0.1, abundance estimates will further degrade unless the number of secondary occasions is increased. Abundance was best estimated when capture probability was in the range of 0.2 – 0.3. Bias in estimated abundance increased slightly at higher capture probabilities. Survival was generally well estimated across the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

entire range of capture probabilities (Table 4). One anomaly occurred where mean estimated survival was over-estimated when capture probability was 0.2 and true survival was 0.9.

DISCUSSION

Noninvasive DNA sampling techniques open many opportunities for capture–recapture studies where they were not previously feasible. This will allow ecologists to answer important questions regarding species which are difficult to detect or handle. One necessary step is to expand DNA-based capture–recapture studies beyond estimates of abundance at a single point in time. The robust design allows ecologists to explore more interesting ecological questions.

The robust design is a natural extension from abundance estimation to explorations of survival and emigration rates. The expanded robust design model presented here allows abundance, survival and emigration to be estimated properly in the face of genotyping error. For studies currently estimating abundance annually in a closed population capture–recapture framework, the robust design offers survival and emigration rates with little or no additional effort. Moreover, these parameters can be modeled as functions of covariates in a general linear models framework to answer a wider array of ecological questions.

The simulation results presented suggest the estimator performs well across varied capture probability and genotyping error levels. It is not surprising that the estimator performed poorly when a capture probability of 0.1 is used because in those cases an

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

animal has only a 0.08 probability of being caught 2 or more times within a primary occasion containing 5 secondary sampling occasions. Therefore, relatively few individuals are recaptured. Recaptures are essential in order to estimate genotyping error rate because the error rate is estimated from the over abundance of genotypes only detected once relative to those detected more than once. Without recaptures or an alternate data source, there is no way to estimate genotyping error. If capture probability is low and there is no direct way to increase it, estimates of abundance and survival can be improved by increasing the number of secondary sampling occasions. This increases p^* that in turn increases the total number of unique animals caught and improves the parameter estimates.

The results in Chapter 1 demonstrate that abundance would be overestimated if genotyping error is not taken into account. These results follow for the estimated abundance in the robust design simulations used here because the robust design model used for estimation had p estimated separately for each primary period and abundance was only considered in the first primary period.

A useful feature of this robust design model is that it shares the same likelihood as a robust design model not including genotyping error. Therefore, model comparisons can be made with information–theoretic criteria such as AIC_c (Burnham and Anderson 2002). This allows one to examine the information loss in assuming no genotyping error when some may be present. If genotyping error is relatively small, it may be advantageous in terms of precision to assume zero error. Conversely, when genotyping error rate is larger

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

the added bias would suggest assuming zero error is a poor choice. AIC_c can quantitatively assess such tradeoffs. Given the recent debate in the literature over the effect of genotyping error on capture–recapture studies (McKelvey and Schwartz 2004a, Paetkau 2004, McKelvey and Schwartz 2004b) and evidence supporting both no error (Paetkau 2003) and considerable error (Creel et al. 2003), it is important to have a data driven way to resolve the issue.

Combining data sources is an important way to gain a better understanding of an ecological system. Here I present a model that incorporates noninvasive genetic data, laboratory testing of the accuracy of the genotyping methods and radio-telemetry data. For species such as bears (*Ursus* spp.) all of these data are routinely available. Moreover, combining telemetry data with DNA-based data can provide a more representative view of the population than with either sampling method alone. Animals that are radio-collared are often those that are more easily captured, for example they live near roads or are involved in human-wildlife conflicts. In addition, radios may affect the survival or behavior of the animal for birds or smaller mammals. Therefore, the radio-telemetered animals may not be representative of the population as a whole. When radio-telemetry data are combined with DNA-based capture–recapture data, one can address questions of differences in characteristics of the population segments.

The standard robust design has been extended to handle other situations. Kendall and Nichols (2002) developed a modification to the robust design to account for unobservable states. This modification brings together the robust design and the multi-

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

state models (Schwartz et al. 1993, Lebreton and Pradel 2002). Schwartz and Stobo (1997) extended the robust design to be geographically open within a primary period. It may be possible to combine the results presented here with these two models, but complications are likely to arise. In both cases there are far more possible outcomes for an unobserved animal, therefore teasing a single observation due to genotyping error apart from all of the other possible causes of a genotype only being observed once will be difficult.

Without accounting for genotyping error when using DNA-based capture–recapture methods, the potential exists to over-estimate abundance and under-estimate survival. DNA-based sampling is often used for rare and exploited species. Therefore, accurate assessment of demographic parameters is key for management of the species. The robust design model presented here allows for appropriate estimates of abundance and survival in the face of genotyping error.

ACKNOWLEDGMENTS

The Colorado Division of Wildlife, Great Outdoors Colorado and Dr. Tanya Shenk provided support for this research.

LITERATURE CITED

Boulanger, J., G. C. White, B. N. McLellen, J. Woods, M. Proctor, and S. Himmer. 2003. A meta-analysis of grizzly bear DNA mark-recapture projects in British Columbia, Canada. Ursus **13**: 137–152.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Burnham, K. P., and D. R. Anderson. 2002. Model selection and multimodel inference, second edition. Springer-Verlag, New York, NY, USA.

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.

Eggert, L. S., J. A. Eggert, and D. S. Woodruff. 2003. Estimating population sizes for elusive animals: the forest elephants of Kakum National Park, Ghana. Molecular Ecology 12: 1389–1402.

Kendall, W. L., J. D. Nichols, and J. E. Hines. 1997. Estimating temporary emigration using capture–recapture data with Pollock's robust design. Ecology **78**: 563–578.

Kendall, W. L., and J. D. Nichols. 2002. Estimating state-transition probabilities for unobservable states using capture-recapture/resighting data. Ecology 83: 3276–3284.

Kohn, M. H., E. C. York, D. A. Kamradt, G. Haught, R. M. Sauvajot, R. Wayne. 1999. Estimating population size by genotyping faeces. Proceedings of the Royal Society of London, Series B 266: 657–663.

Lebreton, J.-D. and R. Pradel. 2002. Multistate recapture models: modelling incomplete individual histories. Journal of Applied Statistics 29: 353–369.

McKelvey, K. S., and M. K. Schwartz. 2004a. Genetic errors associated with population estimation using non-invasive molecular tagging: problems and new solutions. Journal of Wildlife Management **68**: 439–448.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

McKelvey, K. S., and M. K. Schwartz. 2004b. Providing reliable and accurate genetic capture–mark–recapture estimates in a cost-effective way. Journal of Wildlife Management **68**: 453–456.

Paetkau, D. 2003. An empirical exploration of data quality in DNA-based population inventories. Molecular Ecology **12**: 1375–1387.

Paetkau, D. 2004. The optimal number of markers in genetic capture–mark–recapture studies. Journal of Wildlife Management **68**: 449–452.

Palsbøll, P. J., J. Allen, M. Bérubé, P. J. Clapham, T. Feddersen, P. S. Hammond, R. R. Hudson, H. Jørgensen, S. Katona, A. H. Larsen, F. Larsen, J. Lien, D. K. Mattila, J Sigurjónsson, R. Sears, T. Smith, R. Sponer, P Stevick and N. Øien. 1997. Genetic tagging of humpback whales. Nature **388**: 767–769.

Pollock, K. H. 1982. A capture–recapture design robust to unequal probability of capture. Journal of Wildlife Management **46**: 752–757.

Pollock, K. H. 1989. Survival analysis in telemetry studies – the staggered entry design. Journal of Wildlife Management **53**: 7–15.

SAS Institute Inc. 2002. SAS version 9.0. Cary, North Carolina, USA.

Schwarz, C. J., J. F. Schweigert, and A. N. Arnason. 1993. Estimating migration rates using tag-recovery data. Biometrics 49: 177–193.

Schwarz, C. J. and W. T. Stobo. 1997. Estimating temporary migration using the robust design. Biometrics 53: 178–194.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

White, G. C. and R. A. Garrott. 1990. Analysis of wildlife radio-tracking data. Academic Press, San Diego, California, USA.

White, G. C. and K. P . Burnham. 1999. Program MARK: survival estimation from populations of marked animals. Bird Study 46: S120–S139.

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.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

APPENDIX 1

Cell probabilities for the primary sessions of a robust design can be constructed through the use of matrix algebra. To do so, a few matrices must be defined.

Matrices

$$\mathbf{f}_{t+1} = \begin{bmatrix} \gamma''_{t+1} & (1-\gamma''_{t+1})(1-p^*_{t+1}) \end{bmatrix}$$

$$\mathbf{G}_i = \begin{bmatrix} \gamma'_i & (1-\gamma'_i)(1-p^*_i) \\ \gamma''_i & (1-\gamma''_i)(1-p^*_i) \end{bmatrix}$$

$$\mathbf{d}_i = \begin{bmatrix} (1-\gamma'_i) \\ (1-\gamma''_i) \end{bmatrix}$$

Given these matrices and the other model parameters, cell probabilities can be constructed as is shown in Table A1.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table A1. Multinomial cell probabilities for primary periods of a 4 period robust design. The values represent the elements of matrix **A**.

Primary release occasion	Primary recapture occasion 2	3	4
1	$\phi_1(1-\gamma_2'')p_2^*$	$\phi_1\mathbf{f}_2\ \phi_2\mathbf{d}_3\ p_3^*$	$\phi_1\mathbf{f}_2\ \phi_2\mathbf{G}_3\ \phi_3\mathbf{d}_4\ p_4^*$
2		$\phi_2(1-\gamma_3'')p_3^*$	$\phi_2\mathbf{f}_3\ \phi_3\mathbf{d}_4\ p_4^*$
3			$\phi_3(1-\gamma_4'')p_4^*$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 1. Average abundance estimates and standard errors for the first primary occasion of a robust design model including genotyping error at 4 levels. The probability of correctly genotyping a sample is denoted α and abundance is N and SE is standard error of the mean of $\hat{N}$. Mean estimated N and mean standard error of $\hat{N}$ is averaged over 5 levels of capture probability from 0.1 – 0.5. Mean N is based on 1,000 simulation replicates per level of α.

α	mean N	SE	mean SE(N)
0.96	1098	18.1	31.79
0.97	1073	15.6	31.42
0.98	1054	16.4	27.75
0.99	1037	17.7	25.41

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 2. Average estimated survival between primary occasions for a robust design model including genotyping error at 5 levels. The probability of correctly genotyping a sample is denoted α and survival is ϕ. Means and standard errors (SE) are computed over 500 simulation replicates at each level and averaged over varying levels of capture probability.

	mean estimated ϕ			
α	true $\phi = 0.6$	SE	true $\phi = 0.9$	SE
0.96	0.595	0.001	0.911	0.002
0.97	0.596	0.001	0.915	0.002
0.98	0.591	0.001	0.899	0.002
0.99	0.601	0.001	0.909	0.002

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 3. Average abundance estimates by capture probability for the first primary occasion of a robust design model including genotyping error at 5 levels. The capture probability is denoted p and abundance is N and SE is standard error of the mean of $\hat{N}$. Mean $\hat{N}$ is averaged over 5 levels of the probability of correctly genotyping a sample from 0.95 – 0.99. Mean N is based on 1,000 simulation replicates per level of p.

p	mean N	SE	mean SE(N)
0.1	1120	12.17	71.11
0.2	1041	3.11	35.14
0.3	1050	1.50	19.25
0.4	1059	1.19	11.53
0.5	1067	1.19	6.72

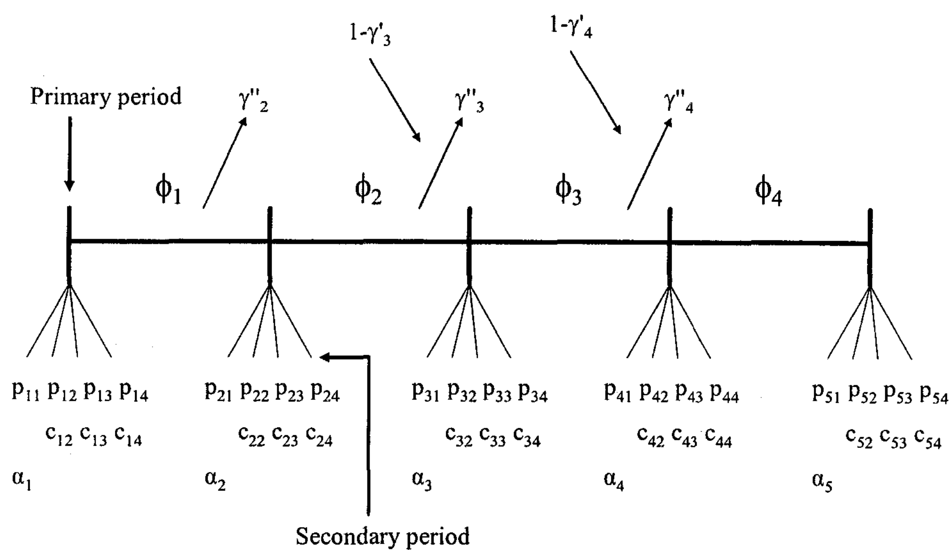
Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 4. Average survival estimates by capture probability for a robust design model including genotyping error at 5 levels. The capture probability is denoted p and survival is ϕ and SE is standard error of the mean of N. Mean ϕ is averaged over 5 levels of the probability of correctly genotyping a sample from 0.95 – 0.99. Mean ϕ is based on 1,000 simulation replicates per level of p.

	mean estimated ϕ			
p	true $\phi = 0.6$	SE	true $\phi = 0.9$	SE
0.1	0.620	0.001	0.927	0.002
0.2	0.609	0.001	0.960	0.002
0.3	0.596	0.001	0.905	0.001
0.4	0.583	0.001	0.890	0.001
0.5	0.574	0.001	0.876	0.001

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

FIGURE 1. An example robust design with $t = 5$ primary periods and $l_1, \ldots, l_5 = 4$ secondary periods.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

CHAPTER 3

ESTIMATING POPULATION SIZE FROM DUNG-BASED DNA CAPTURE–RECAPTURE DATA

ABSTRACT

Non-invasive samples of animal dung are being collected for use with DNA-based capture–recapture analysis to estimate the abundance of the species. A typical sampling strategy involves repeated samples of transects or grids to search for dung. Searches performed on different transects or on different occasions are considered the capture sessions for capture–recapture analysis. Often multiple samples of dung from the same individual are found within an occasion. Current discrete-time capture–recapture models do not account for this additional data. I present a generalization of the geographically and demographically closed population capture–recapture models that use the numbers of dung found for each individual during each sampling occasion to help estimate heterogeneity in capture probability. I adopt a Bayesian approach using Markov chain Monte Carlo to facilitate computing. I demonstrate that in the face of heterogeneity caused by unequal amounts of dung available per individual, the estimator developed here out performs standard capture–recapture abundance estimation. I present an example using dung-based capture–recapture data from African elephants. Heterogeneity in capture probability causes under-estimates of abundance. The method developed here

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

allows individual heterogeneity in capture probability to be estimated with additional data available in dung-based DNA capture–recapture studies.

Key-words: capture–recapture, closed population models, DNA markers, microsatellites, non-invasive sampling, population size

INTRODUCTION

Non-invasive samples of animal dung are currently being collected for use in estimating animal abundance with capture–recapture methods (Kohn et al. 1999; Creel et al. 2003; Eggert et al. 2003). The dung samples are genotyped for a set of microsatellite loci. Individual genotype samples are matched across sampling occasions. Matching samples are considered to be recaptures of the individual. Current methods assume genotypes are accurately obtained and the set of microsatellite loci are sufficient to uniquely resolve individuals. In some cases these assumptions seem reasonable (Eggert et al. 2003; Paetkau 2003) while in other cases genotyping error rate has been shown to be quite high (Creel et al. 2003).

Standard discrete-time capture–recapture theory assumes an individual animal is detected only once per sampling occasion (Otis et al. 1978). This is a reasonable assumption when marked animals are physically captured or directly sighted when encountered. For dung sampling, observations of the animal are indirect. Possibly an animal leaves multiple dung piles within the sampled area within the time interval used for sampling. A reasonable assumption is that not all animals deposit the same number of

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

dung piles in a given unit of time. Therefore, in surveys using dung sampling additional data are available in the count of the number of dung piles identified for each individual sampled on a given occasion.

When multiple dung piles are potentially available for each individual and individuals deposit varying numbers of dung piles, heterogeneity is introduced into the probability of detecting an individual animal. Some methods, such as finite mixture models (Pledger 2000), frequency of count methods (Chao and Lee 1992) and the jackknife estimator (Burnham and Overton 1979), account for some unspecified heterogeneity, but no methods specifically target this dung sampling-DNA form of heterogeneity in capture probability. When present and not taken into account, heterogeneity in capture probability results in under-estimates of population size because individuals with high capture probability are sampled disproportionately relative to individuals with low capture probability. Therefore, the average capture probability for the population is over-estimated and population size is under-estimated.

Recently, *N*-mixture models using Poisson or negative binomial mixing distributions have been developed to estimate animal abundance (Royle and Nichols 2003; Royle 2004). These models attempt to estimate the mean abundance of animals per plot based on counts in each plot being a realization of a Poisson process with a common mean. Royle and Nichols (2003) use the Poisson and negative binomial mixtures to account for heterogeneity in the probability of detecting presence or absence of a species due to varying abundance on each plot. This idea can be used to model the mean number

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

of dung piles available per individual and then that model is imbedded into a capture–recapture model to estimate animal abundance.

I present an extension to standard capture–recapture theory that uses the additional data of the number of dung samples per occasion matching each genotype to account for the heterogeneity in capture–probability caused by varying numbers of dung piles produced per individual. The model is fit using Bayesian Markov chain Monte Carlo (MCMC) methods. I then contrast the method to standard capture–recapture methods. I also present an analysis of African elephant (*Loxodonta cyclotis*) dung DNA-based capture–recapture data as an example of the method.

METHODS

Statistical Notation

The notation presented here largely follows Otis et al. (1978) and Royle and Nichols (2003) where applicable. The model assumes there are t sampling occasions, the population of interest is well defined and it is demographically and geographically closed during the sampling period.

Parameters

r_i probability of detecting an individual dung pile on occasion i.

λ mean abundance of dung per individual animal

a shape parameter for the negative binomial

f_0 the number of genotypes not detected during the study

N population size, more precisely the number of genotypes in the population

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

μ_j mean of the prior distribution of parameter j. Parameter j may be the logistic transformation of r_i, the log transformation of λ, or the log transformation of f_0.

σ_j standard deviation of the prior distribution of parameter j. Parameter j may be the logistic transformation of r_i, the log transformation of λ, or the log transformation of f_0.

Statistics

$\mathbf{h} = \{h_1, h_2, \ldots, h_t\}$ encounter history vector where h_i is the count of the number of dung piles with matching genotypes sampled on occasion i. For example, $\mathbf{h} = \{3,2,0,4\}$ for an individual genotype detected 3 times on occasion 1, twice on occasion 2, not detected on occasion 3, and detected 4 times on occasion 4.

M_{t+1} number of distinct genotypes observed. $M_{t+1} = N - f_0$

Statistical Model and its Assumptions

I assume that genotyping occurs without error. I further assume that genetic markers with sufficient power to resolve individuals are used (Paetkau 2003). I assume that the number of dung piles deposited by each individual is a realization of Poisson process with mean λ constant across the duration of the study or a negative binomial process with mean λ and the variance is $\lambda + \frac{\lambda^2}{a}$. In order for an individual animal to be detected, there must be at least 1 dung pile available for sampling, therefore the zero-truncated Poisson and negative binomial distributions are used. Detection probability per dung pile is assumed to be equal to and independent of other dung piles. The likelihood

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

developed from these assumptions is

$$L(f_0, \lambda, r_i | \mathbf{h}) \propto \frac{(f_0 + M_{t+1})!}{f_0!}$$

$$\cdot \prod_{i=1}^{M_{t+1}} \left[\sum_{j=\max(h_{i.})}^{\infty} \left(\prod_{k=1}^{t} (1-(1-r_k)^j)^{I_{h_{ik}>0}} (1-r_k)^{jI_{h_{ik}=0}} \Pr[j \text{ piles} | \text{individual } i] \right) \right]$$

$$\cdot \prod_{i=0}^{f_0} \left[\sum_{j=1}^{\infty} \left(\prod_{k=1}^{t} (1-r_k)^j \Pr[j \text{ piles} | \text{individual } i] \right) \right]$$

where I is an indicator function returning a 1 if the argument is true and 0 otherwise. The model describes the probability of observing h_i dung piles at occasion k given the maximum number of dung piles observed for a genotype and the mean deposition rate. The probability of observing each count is weighted by the probability that a specific number of dung piles are available to be sampled. Then the probability of failing to observe any dung piles from a given genotype is computed and weighted across possible counts of available piles.

Specifically, the mixture model allows the probability of every possible number of dung piles per animal to be computed. The first summation begins at the largest number of dung piles found for each individual on any occasion, for example h. Then the probability of detecting h dung piles is computed and multiplied by the probability of h dung piles existing. Next, one dung pile is added, now $h + 1$, and the probabilities are recalculated and added to the previous probability. This is repeated up to an arbitrary maximum number of dung piles (theoretically infinite) for every individual detected. The

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

second summation computes the probability of an individual never being detected and multiplies that by f_0, the number of individuals not detected.

I considered two possible distributions for the number of dung piles available. First, the Poisson may be reasonable when the variance in the number of dung piles per individual is small. The zero-truncated Poisson is given as:

$$\Pr[j \text{ piles}] = \frac{e^{-\lambda}\lambda^{j}}{\left(1 - e^{-\lambda}\right)}$$

where $\left(1 - e^{-\lambda}\right)$ in the denominator provide for zero truncation. In many situations, the variation in counts from natural systems exceeds that of a Poisson distribution. In such cases a negative binomial distribution often fits the data well. The zero-truncated negative binomial is:

$$\Pr[j \text{ piles}] = \frac{\Gamma(j+a)}{j!\,\Gamma(a)}\left(\frac{(1/a)\lambda}{1+(1/a)\lambda}\right)^{j}\left(\frac{1}{1+(1/a)\lambda}\right)^{a}\left(1-\left(\frac{1}{1+(1/a)\lambda}\right)^{a}\right)^{-1}.$$

The mean of the untruncated negative binomial is λ and the variance is $\lambda + \frac{\lambda^2}{a}$. Thus, the negative binomial has more flexibility to handle over dispersion. The negative binomial is zero-truncated as well to handle the assumption that each individual must leave at least one dung pile in order to be sampled.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Model parameters are transformed on link functions to have properties more desirable for estimation, similar to what is done in standard capture–recapture software such as Program MARK (White and Burnham 1999). Capture probability, r_i, is transformed to a logit scale while f_0 and λ are transformed on a log scale. These transformations give the parameter estimators roughly normal sampling distributions and force the back transformed parameters to be bounded on their appropriate scales. For this reason, normal prior distributions are used with parameters $\mu = 0$ and $\sigma = 1{,}000$ for all transformed model parameters. These prior distributions are essentially flat and uninformative across the likely range of model parameters. In practice, the summation across the Poisson mixture must be stopped at a finite number. Typically, summing to a value of roughly 10 times the largest number of dung observations per individual is sufficient to get to a region of the Poisson distribution where the probability of occurrence is nearly zero.

Model Testing

The models presented above were compared to standard capture–recapture estimation using simulated data. First, data were generated from a Poisson distribution with a mean number of dung piles per individual available at each occasion of 7. The population consisted of 1,000 individuals. Each dung pile had a capture probability of 0.1 across all sampling occasions and sampling occurred on 5 occasions. The counts of the number of dung piles were used for the models presented in this paper, hereafter referred to as MDP (Multiple Detections with Poisson). Counts were collapsed to zeros and ones for the standard capture–recapture analysis. Two standard capture–recapture

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

models were used for comparison to MDP; 1) model M_0 which has a constant capture probability (Otis et al. 1978) and a two component finite mixture model with constant capture probability within components M_{h2} (Pledger 2000). M_0 represents a model that completely ignores individual heterogeneity in capture probability while M_{h2} accounts for heterogeneity in capture probability without assigning a reason for it. The simulation was repeated 100 times. For exploration of the posterior of MDP, the Poisson summation was stopped at 50. When the mean number of dung piles is 7, summing to any value above 20 results in the same description of the posterior distribution.

Evidence exists that animal dung piles exhibit extra-Poisson variation (White and Eberhardt 1980). Therefore, I repeated the simulation above except data were generated from a negative binomial distribution with a mean of 7 and variance of 14 ($a = 7$). This allowed the effectiveness of the Poisson model assumption to be explored when the true variance in the number of dung piles was twice that of the Poisson variance. It also allows the negative binomial model to be tested, hereafter called MDB (Multiple Detections with negative binomial).

I use a Metropolis-Hastings random walk Markov chain Monte Carlo (MCMC) algorithm to explore the posterior distribution (Metropolis et al. 1953; Hastings 1970). Five independent chains are run with a burn-in period of 2,000 iterations and a sampling period of an additional 5,000 iterations for each of the parameters. Convergence to the posterior distribution was very quick, therefore I used fewer burn-in iterations than sampling iterations and only 5 chains which are both fewer than recommended by Gelman et al. (2004). This allowed us to place more effort into simulation replicates and

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

less into individual analyses without changing the results. The variance ratio method is used to assess convergence of the Markov chain (Gelman et al. 2004).

Comparison between the two techniques is somewhat difficult because standard capture–recapture models are likelihood-based while MDP and MDB are Bayesian. In addition, the only common parameter between the models is population size (N) which is derived from the sum of f_0 and M_{t+1}. Detection probability is not directly comparable between the standard capture–recapture models and MDP or MDB. In the standard capture–recapture models, detection probability (p) is the probability of detecting at least one dung pile. For model MDP, detection probability (r) is the probability of detecting a single dung pile. For comparison, I examine the average MLE for N from M_0 and the average posterior mean from MDP and MPB. I also examine confidence interval coverage for M_0 and M_{h2} and credible interval coverage for MDP and MDB.

Example

Elephant dung was collected and genotyped to estimate the population size in Parc National de la Marahoué, Côte d'Ivoire. This data was previously analyzed by Eggert (2004) using the jackknife estimator and an accumulation curve. Dung samples were collected within seven sections of the park. Each of the seven sections was sampled once to generate a seven-sample capture-recapture data set (MRC7). In addition, areas of high concentrations of elephants were sampled a second time. The first seven samples were collapsed into a single sample and joined with the second sampling to create a second data set with two capture occasions (MRC2). DNA was extracted from the samples and

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

amplified at six microsatellite loci and a sexing locus. Individuals were matched by genotype. Genotypes which differed at one or two loci were rechecked for accuracy.

Here I analyze the elephant data with model MDP and MDB to attempt to account for heterogeneity in the data caused by varying numbers of dung piles available per elephant. MDB was only fit to the MRC7 data because the two occasions of MRC2 would not be enough to support a negative binomial. The results are compared to the jackknife, M_0 and M_{h2} estimator results. Five independent Markov chains were run for 5,000 burn-in iterations and 5,000 additional iterations for each parameter. I again use only five chains to remain consistent with the simulations.

RESULTS

Model Testing

MDP outperformed standard capture–recapture models when the dung data were generated following a Poisson distribution. Model M_0 performed worst in terms of bias. The mean maximum likelihood estimator (MLE) of N for M_0 was 966 (SE = 2.40). Models M_{h2} and MDP performed nearly equal to each other in terms of bias. The mean MLE of N for M_{h2} was 994 (SE = 6.15) and the average posterior mean for MDP was 992, (SE = 3.95). The 95% confidence intervals on N for M_0 failed to cover the true value of 1,000 in every simulation replicate. M_{h2} performed better than M_0 in terms of confidence interval overage, but still not well. The achieved coverage was 77% (SE = 4.2). The coverage of the 95% credible intervals on N was 93% (SE = 2.6) for MDP.

MDB performed best overall when the dung data were generated from a negative binomial distribution with a variance twice as large as the assumed Poisson distribution in

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

model MDP. Mean abundance for MDB was 972 (SE = 3.9). The mode of abundance for MDB was 1003. This is the only case in the simulations where the mode and mean were substantially different. Mean abundance for M_{h2} was 1004 (SE = 54.7). MDP underestimated abundance, mean N = 939 (SE = 9.8). M_0 greatly underestimated abundance, mean N = 880 (SE = 5.8). While M_{h2} was the least biased on average, the large variance in the estimates of abundance made it indistinguishable from MDP or MDB. In addition, M_{h2} failed to converge to an MLE on 13 of 100 simulation replicates. MDB performed best in terms of credible interval coverage with 93% (SE = 2.6) of intervals covering the true value. The confidence interval coverage of M_{h2} was 70% (SE = 4.5). MDP had a credible interval coverage of 26% (SE = 4.4). M_0 again failed to cover the true value in all simulation replicates.

Example

Model MDP yielded similar results for the abundance of elephants as did the jackknife estimator, posterior mean N = 152 (SD = 11.02) for MRC7 (Figure 1). MDP produced a higher estimate than either M_0 or M_{h2}. Capture probability per dung pile was 0.017 (SD = 0.006) for the MRC7 data and 0.103 (SD = 0.023) for the MRC2 data. The estimated mean number of dung piles available to be sampled (λ) was 7.75 (SD = 2.24) for MRC7 and 9.55 (SD = 1.68) for the MRC2 data. The variance ratio statistic suggests the Markov chains converged. For the MRC7 data, $\sqrt{R}$ = 1.006 for parameter r, $\sqrt{R}$ = 1.003 for N, and $\sqrt{R}$ = 1.004 for λ. For the MRC2 data, $\sqrt{R}$ = 1.005 for parameter r, $\sqrt{R}$ = 1.004 for N, and $\sqrt{R}$ = 1.004 for λ. Values of $\sqrt{R}$ close to 1.0 suggest

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

convergence has occurred and values less than approximately 1.2 are acceptable (Gelman et al. 2004). The mean and count parameters of MDB converged to a Poisson distribution suggesting there was no overdispersion in the MRC7 data.

DISCUSSION

When the number of dung piles varies among individuals, detection probability will vary among individuals because animals that left more dung piles will be more likely to be detected. Standard capture–recapture estimates of abundance will be biased low when individual heterogeneity in capture probability is not taken into account (Otis et al. 1978). Many methods have been developed to help deal with heterogeneity in capture probability, for example jackknifing (Burnham and Overton 1979), frequency of capture methods (Chao and Lee 1992), using covariates (Huggins 1989), and finite mixtures (Pledger 2000). While all of these methods are useful for estimating abundance in the face of heterogeneity in capture probabilities, none of them specifically targets heterogeneity caused by varying numbers of dung piles available.

Here I developed a method that explicitly models the varying numbers of dung piles available for each individual. By using Poisson and negative binomial mixtures on the number of dung piles available for each individual I am able to parsimoniously estimate the mean amount of dung per individual. This adds only one or two parameters to a standard capture–recapture model. Moreover, by placing the analysis in a Bayesian framework the estimation is simple and straight forward. Attempts at using maximum likelihood estimation for this method proved unreliable. I chose to use a zero-truncated distribution and assume that all individuals left at least one dung pile. Without this

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

assumption, it could be problematic to attempt to estimate the abundance of animals with zero probability of being sampled. Moreover, it is reasonable to assume that all animals are depositing dung.

The simulation results demonstrate that in the face of varying numbers of dung piles available, estimation with capture–recapture methods not accounting for heterogeneity did not perform very well. The estimates are biased low and provide a falsely short confidence interval. When finite mixtures are used to approximate individual heterogeneity in capture probability, estimation is greatly improved. Yet, even though bias is eliminated, confidence interval coverage remains below the desired level. This result suggests that the estimator is not adequately accounting for the uncertainty in estimated population size when heterogeneity is caused by varying dung abundance. Conversely, this method is nearly unbiased and provides a realistic assessment of the uncertainty of the estimate of population size.

Evidence exists in the literature that the number of dung piles shows greater variance than a Poisson distribution allows and fit a negative binomial distribution (White and Eberhardt 1980). When I simulated data under a negative binomial distribution with a variance twice as large as the Poisson variance, only MDB performed well. This result suggests that multiple models should be run to examine distribution of the counts of dung piles and how that will affect the abundance estimate. M_{h2} did well in terms of bias, but was imprecise and unstable. Even when the point estimate was very close to the true value, the confidence interval was too wide to be informative; for example from M_{t+1} to

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

>4000. MDP underestimated abundance by about 6%, but still estimated the mean number of dung piles available appropriately.

I assumed samples are genotyped without error. In some studies, this is a valid assumption, while in others it is less so. Methods have been developed to minimize genotyping error (Paetkau 2003) and to test for genotyping error (McKelvey and Schwartz 2004). The method developed here is applicable for studies with a high degree of certainty in genotypes. Data from studies with lower certainty or an unknown level of genotyping error would be more appropriately analyzed with a method that incorporates genotyping error into the abundance estimate (Chapter 1).

Model MDP appeared to perform well for the elephant data. As expected the estimated abundance was greater than that for model M_0 which does not account for heterogeneity, but the two are similar. The results from model M_{h2} suggest there is surprisingly little heterogeneity in the data. This leads to the reason why models MDP and M_0 give similar results. The lack of heterogeneity was further supported by MDB collapsing to a Poisson distribution. If more heterogeneity had been present in the data then the difference between the two estimates would be larger.

The method presented here could have applications beyond dung sampling. When using hair snags to collect DNA samples, some individuals may leave samples a more than one hair snag. In these situations often only a subset of the samples are genotyped due to logistical or financial constraints. Therefore, the λ parameter would estimate the mean number of samples available per individual if all samples were to be genotyped.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

ACKNOWLEDGMENTS

The Colorado Division of Wildlife, Great Outdoors Colorado and Dr. Tanya Shenk provided support for model development. The African Elephant Conservation Fund, grant #98210-1-G968 administered by the U. S. Fish and Wildlife Service provided funding for Lori Eggert. Dr. Gary C. White and Dr. Michael Phillips provided comments on the manuscript.

REFERENCES

Burnham, K. P. and Overton, W. S. (1979) Robust estimation of population size when capture probabilities vary among individuals. *Ecology*, **60**, 927–936.

Chao, A., and S. M. Lee. (1992) Estimating the number of classes via sample coverage. *Journal of the American Statistical Association*, **87**, 210–217.

Creel, S., Spong, G., Sands, J. L., Rotella, J., Zeigle, J., Joe, L., Murphy, K. M., and Smith, D. (2003) Population size estimation in Yellowstone wolves with error-prone noninvasive microsatellite genotypes. *Molecular Ecology*, **12**, 2003–2009.

Eggert, L. S., Eggert, J. A., and Woodruff, D. S. (2003) Estimating population sizes for elusive animals: the forest elephants of Kakum National Park, Ghana. *Molecular Ecology*, **12**, 1389–1402.

Eggert, L. S. (2004) The elephants of Parc National de la Marahoué, Côte d'Ivoire: genetic estimate of population size and demography. Smithsonian Institution Report.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004) *Bayesian data analysis second edition*. Chapman and Hall / CRC, New York.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Hastings, W. K. (1970) Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, **57**, 97–109.

Huggins, R. M. (1989) On the statistical analysis of capture experiments. *Biometrika*, **76**, 133–140.

Kohn, M. H., York, E. C., Kamradt, D. A., Haught, G., Sauvajot, R. M., Wayne, R. (1999) Estimating population size by genotyping faeces. *Proceedings of the Royal Society of London, Series B*, **266**, 657–663.

McKelvey, K. S. and Schwartz, M. K. (2004) Genetic errors associated with population estimation using non-invasive molecular tagging: problems and new solutions. *Journal of Wildlife Management*, **68**, 439–448.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953) Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, **21**, 1087–1092.

Otis, D. L., Burnham, K. P., White, G. C., and Anderson, D. R. (1978) Statistical inference from capture data on closed animal populations. *Wildlife Monographs*, **62**.

Paetkau, D. (2003) An empirical exploration of data quality in DNA-based population inventories. *Molecular Ecology*, **12**, 1375–1387.

Pledger, S. (2000) Unified maximum likelihood estimates for closed capture–recapture models using mixtures. *Biometrics*, **56**, 434–442.

Royle, J. A. and Nichols, J. D. (2003) Estimating abundance from repeated presence–absence data or point counts. *Ecology*, **84**, 777–790.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

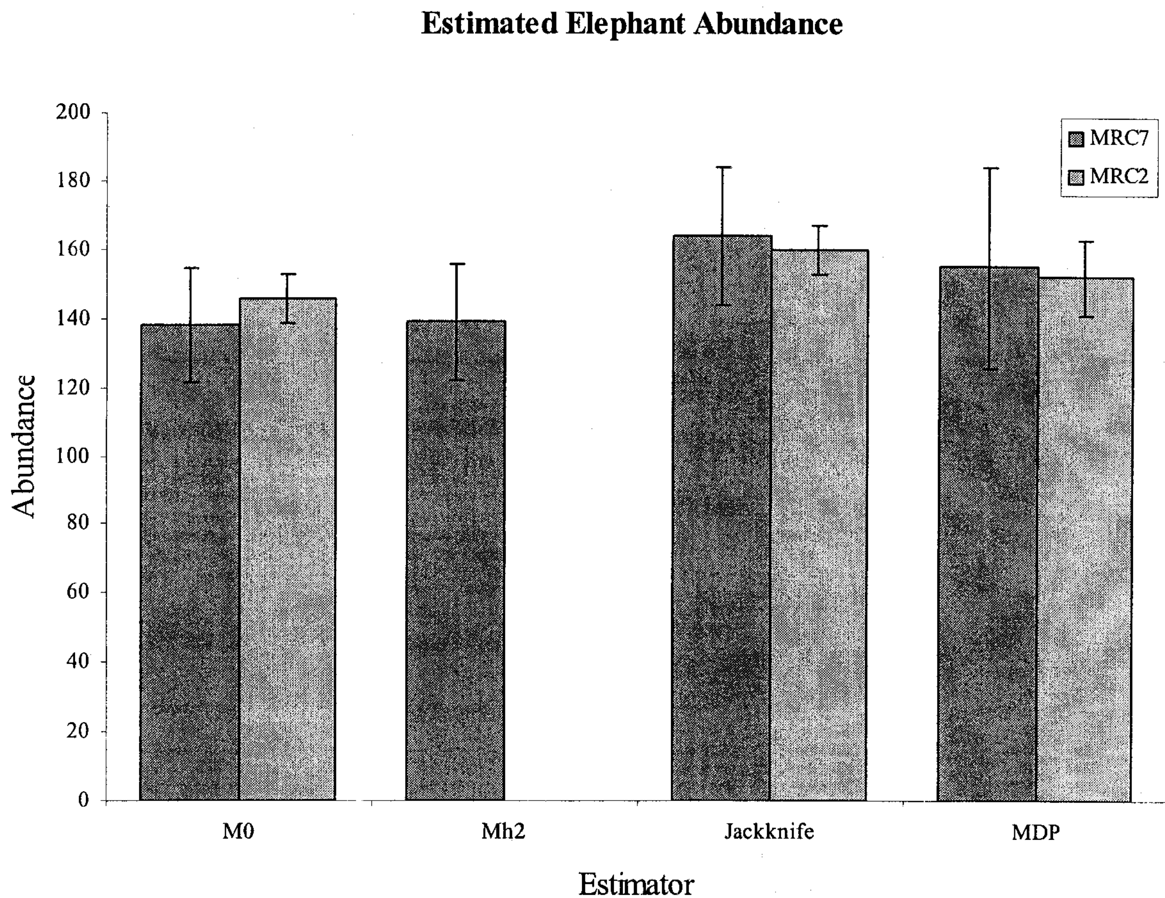
Royle, J. A. (2004) *N*-mixture models for estimating population size from spatially replicated counts. *Biometrics*, **60**, 108–115.

White, G. C. and Burnham, K. P. (1999) Program MARK: survival estimation from populations of marked animals. *Bird Study*, **46**, S120–S139.

White, G. C. and Eberhardt, L. E. (1980) Statistical analysis of deer and elk pellet-group data. *Journal of Wildlife Management*, **44**,121–131.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Figure 1. Estimates of the population size of African elephants in Parc National de la Marahoué, Côte d'Ivoire using 4 different capture-recapture estimators. Model M_{h2} is not estimable for the MRC2 data because it has only 2 occasions. Error bars represent one standard error for models M_0, M_{h2} and the jackknife. The error bars represent one standard deviation of the posterior distribution of N for model MDP.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

CHAPTER 4

SAMPLING STRATEGIES FOR MONITORING A REINTRODUCED CANADA LYNX POPULATION IN COLORADO

Abstract: Canada lynx (*Lynx canadensis*) were reintroduced in Colorado beginning in 1999 with the intent to develop a self-sustaining lynx population in the state. Currently, lynx are monitored using a labor intensive combination of satellite telemetry, radio telemetry, and snow tracking. As the population continues to grow and stabilize, the Colorado Division of Wildlife hopes to use a less intensive and invasive method to monitor the lynx population. I present a protocol for monitoring lynx in Colorado, along with an approach for how to determine sample size and intensity for such a monitoring program. The monitoring is based on finding lynx scat in summer with the aid of trained dogs and using a DNA-based robust design capture–recapture analysis to estimate abundance and survival.

Key words: capture–recapture, Canada lynx, microsatellite, monitoring, non-invasive sampling, reintroduction

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

INTRODUCTION

The Colorado Division of Wildlife (CDOW) reintroduced Canada lynx (*Lynx canadensis*) into southwestern Colorado beginning with 41 lynx released in 1999. Releases of additional lynx have continued through 2004. The goal of the reintroduction is to build a self-sustaining lynx population at the southern extent of the species range. The lynx is listed as threatened under the Endangered Species Act (U. S. Fish and Wildlife Service 2000). As a threatened species, demographic information on the Colorado population demonstrating population viability is needed to meet the recovery goals for the southern portion of the species range (U. S. Fish and Wildlife Service 2000).

The reintroduced population of lynx must be monitored to determine the success of the effort and long-term persistence of the population. CDOW outlined two primary goals for the lynx monitoring effort (Shenk 2001). First, CDOW biologists would like to know the abundance of lynx in the state. Initially, abundance was known by tracking the survival of radio-collared released animals. Now, determining abundance is more difficult because radio collars have failed on some lynx and reproduction is known to have occurred in 2003 and 2004 (Shenk 2004). Therefore, the total population size is no longer known. As time progresses the population size will be less and less reflective of the number of animals released. Second, CDOW would like to monitor the survival rate of the lynx in Colorado. Radio-telemetry is currently used for this purpose, but trapping

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

lynx to fit collars and subsequent tracking is expensive and invasive. Therefore, an alternative method for estimating survival would be beneficial.

Monitoring lynx is ongoing within the historic lynx range across public land in the United States as part of the National Lynx Survey (McKelvey 2002). The goal of the program is to determine the present range of Canada lynx within the contiguous United States. The protocol uses scent stations with hair rubs to collect samples of lynx hair with a follow-up snow tracking survey in areas where lynx hair was detected (McKelvey 2002). McDaniel et al. (2000) demonstrated that scent stations are effective for luring lynx to hair rubs. The hair is positively identified as lynx hair through genetic analysis. This program is successful and provides a starting point for lynx monitoring in Colorado.

It is beneficial for a lynx monitoring program to use non-invasive sampling and not require any direct contact with the animals. Great effort and expense are required to capture lynx. Canada lynx will always be relatively rare in Colorado. Their behavior and habitat make them difficult to capture. Thus, the effort required to trap lynx is considerable. Second, the portion of the population sampled in a radio-telemetry study may not be representative of the lynx population as a whole. The trauma of capture may adversely affect the lynx potentially decreasing survival or altering behavior. The lynx captured are more likely to be individuals that are in relatively easy areas to access such as those that spend time close to roads. Third, non-invasive sampling is preferred from an ethical standpoint (Goodall and Bekoff 2002). Non-invasive sampling allows us to learn about lynx without having to handle or disturb the animals.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Non-invasive DNA-based capture–recapture studies are becoming common. Abundance has been successfully estimated for bears (*Ursus* spp.; Woods et al. 1999), African elephants (Eggert et al. 2003, Chapter 3), and humpbacked whales (Palsbøll et al. 1997) among other taxa. In some studies, genotyping error has been problematic leading to an unrealistically high number of genotypes observed (Creel et al. 2003). Protocols have been developed for genetic analyses to minimize errors in genotyping (Paetkau 2003). In addition, explicit tests for genotyping error have been developed (McKelvey and Schwartz 2004) and methods for estimating and correcting for genotyping error rate in a capture–recapture context have also been developed (Chapter 1, Chapter 2). The microsatellite library that has been developed for the lynx is variable enough to provide individual identification (Carmichael et al. 2000, Schwartz et al 2004). Thus, enough development and testing has been performed to consider DNA-based capture–recapture as a viable long-term monitoring option for lynx in Colorado.

Here I outline the steps taken to determine the amount of effort and sample size requirements to monitor lynx population size and survival rates as based on sampling hair or scat. These steps are presented to 1) provide a guideline for the design of a monitoring program in Colorado and 2) to provide a template of the steps to take to design future DNA-based capture–recapture monitoring programs.

METHODS

Initially, I was interested in determining amplification rates and genotyping error rates for plucked lynx hair. Here I define amplification rate as the proportion of samples

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

that produce a consistent genotype across analysis runs at the desired number of loci. A genotyping error is the occurrence of 2 samples known to be from the same individual producing different genotypes at one or more loci. Genotyping error rate is the proportion of paired samples which produce different genotypes at one or more loci. Plucked hair and blood samples were collected from lynx while being held in pens prior to release. First, I submitted 100 hair samples with 3 – 10 hair follicles per sample for DNA extraction and amplification at six dinucleotide microsatellite loci (Lc106, Lc109, Lc110, Lc111, Lc118 and Lc120; Carmichael et al. 2000). Hair samples were collected directly from lynx held in captivity prior to release in Colorado in 1999 and 2000. Samples were individually placed in paper envelopes and refrigerated. Of the 100 hair samples, 61 unique lynx were represented and 39 samples were replicates of a subset of those lynx. This allocation of samples was chosen to simultaneously minimize the variance in estimating genotyping error rate and allele frequency given a total of 100 samples were to be analyzed. Genetic analyses were performed in 2003. Later, I submitted 33 paired hair and blood samples for DNA extraction and amplification at the sample microsatellite loci. Samples were collected from lynx brought to Colorado in 2003 and analyzed in the same year. I estimated amplification rate and genotyping error rate from these analyses.

I computed the bias and precision of estimated survival rate and population size given several assumptions about amplification rate, genotyping error rate, detection probability and number of sampling occasions. I assumed a starting lynx population of 200 individuals because that is a reasonable guess at the size of the Colorado lynx

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

population in the near future. To address bias and precision of estimated population size, I simulated closed population capture–recapture data with detection probability ranging from 0.1 – 0.5, for 5 and 7 sampling occasions, and a genotyping error rate of 0.03 per sample in a factorial arrangement. Data were analyzed under three closed population capture–recapture models from Otis et al. (1978) and three models from Chapter 1. The Otis et al. (1978) models represented conditions where capture probability was constant (M_0), or varied across sampling occasions (M_t),or varied due to a behavioral response of the animal (M_b). The models from Chapter 1 represented the same conditions, but added a parameter to estimate and adjust for genotyping error rate. I computed expected bias and standard error from each combination of factor levels. In addition, sample size required for varying levels of precision was computed given the likely amplification rates per sample of hair samples and fecal samples. Amplification rates for hair were based on the results of my hair samples and a 90% amplification rate per sample for fecal samples was based on information from unpublished data (M. K. Schwartz pers. comm.). While I did not have fecal samples of or own to test, I felt it was important to consider given recent successes with DNA extraction from lynx feces (M. K. Schwartz pers. comm.).

I used designs suggested from the closed population exploration above to determine the expected bias and precision for estimated survival rate. I simulated data under a robust design (Kendall et al. 1997) with per secondary occasion capture probability ranging from 0.2 – 0.4, and annual survival rate ranging from 0.6 – 0.9. Four primary occasions were simulated yielding three estimates of survival. A factorial

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

arrangement was used for the simulation. Data were analyzed using a robust design model extended to incorporate genotyping error (Chapter 2). I computed expected bias and standard error of estimated survival.

RESULTS

Amplification rates for the hair samples were low (Table 1). Across both sets of samples only 21% (SE = 4%) of the hair samples amplified at four or more loci. Only 8% (SE = 3%) of the hair samples amplified at all 6 loci. For samples which produced a consistent, amplified product at four or more loci, genotyping error rate per sample was low. Only a single pair of hair samples from one lynx did not match at all of the amplified loci resulting in a genotyping error rate of 4% (SE = 4%).

The closed population capture–recapture simulations suggested lynx population size could be estimated from non-invasive DNA-based sampling. Bias/SE for estimated abundance was minimized with per occasion capture probability in the range of 0.2 – 0.3 (Figure 1). Non-invasive DNA-based sampling failed to be effective when capture probability fell to 0.1 or less. Increasing per occasion capture probability above 0.3 provided little gain for the additional effort especially when 7 occasions are used.

Given the best estimates are produced with a capture probability in the range of 0.2 – 0.3, I estimated the number of samples required to achieve a coefficient of variation on estimated abundance of ≤ 0.2 from hair samples and from scat samples (Figure 2). Using hair as the DNA source requires 5 – 10 times more samples to be collected than using fecal samples because most of the hair samples fail to yield DNA.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Robust design simulations suggested estimating survival requires less effort than estimating abundance. For a genotyping error rate per sample of 3%, the coefficient of variation on estimated survival was always less than 0.1 for a design with 4 primary occasions and 5 secondary occasions per primary occasion. In addition, estimated survival was essentially unbiased. Therefore, estimating abundance was the driving factor for determining the sample size needed for monitoring, while survival was obtained as a by-product.

DISCUSSION

A non-invasive DNA-based capture–recapture sampling scheme holds promise for monitoring lynx in Colorado. This sampling scheme allows both abundance and annual survival to be estimated. Both parameters are needed to fulfill the goals set out by CDOW and to meet recovery goals of the U. S. Fish and Wildlife Service.

Sampling lynx scat appears to be a more efficient method than sampling hair. Scat can be found by following lynx tracks in snow until scat is encountered or with the aid of trained dogs in the summer. To obtain hair samples, a lynx must be lured to a scent rub and must choose to rub the scent pad and leave hair in the process. Moreover, DNA amplification rate is much higher for scat than it is for hair in lynx (M. K. Schwartz pers. comm.). This allows the desired estimates of abundance and survival to be obtained with far fewer samples collected in the field. Furthermore, McDaniel et al. (2000) tested the efficacy of hair snares in Yukon, Canada. They placed 78 transects of 5 hair snag scent stations and obtained 58 hair samples from lynx working in an area likely to have a higher

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

lynx density than Colorado. Monitoring lynx abundance and survival in Colorado would require 1,500 – 2,000 hair samples, a rather unreasonable number.

The robust design offers a useful analysis framework for estimating both abundance and survival. Estimating abundance requires a substantial effort each time it is attempted. If sampling for abundance periodically, the robust design allows survival to be estimated with a high degree of precision without any additional cost. This does require that the same microsatellite loci are used at each sampling session in order to identify animals across years.

Here I chose to use a standard capture–recapture model for estimation rather than the models developed in Chapter 3 which are specifically designed for use with scat data. I have three reasons for this decision. First, lynx do not leave large quantities of scat as do herbivores. Therefore, the probability of encountering multiple scats from the same individual within the same sampling occasion is low. Thus, little additional data are available from multiple detections from which to estimate heterogeneity in capture probability. Second, a robust design version of the scat abundance model does not exist yet. I would like to keep the entire analysis in the same framework, therefore I chose the standard capture–recapture models which have been extended to a robust design.

Scat samples are useful because they offer information beyond what is targeted from this study. Only a small portion of a scat is needed for the DNA analysis. The remainder of the scat could be used to explore other questions such as lynx diet. While hair samples can also be used in various types of studies, the limited quantity of hair

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

collected per sample greatly limits the number of things that can be done with a single sample. In the future, it may be possible to age lynx into either a pre-breeding age or post-breeding age category through the use of hormones in the scat. Age would be a very useful covariate in a capture-recapture study. Thus, there is more impetus to use scat rather than hair.

I recommend a robust design monitoring program for lynx in Colorado. Under this design, each summer scat would be collected by using trained dogs to find scat. The annual surveys would constitute the primary sampling occasions of the robust design. The summer sampling would be broken up into five secondary sampling occasions. The goal is to collect at least 300 scats per year.

While it is a common method and relatively easy to find lynx scat in winter by following tracks, I prefer a summer survey. During winter, much of the lynx range is inaccessible due to snow. Therefore, the sampling is heavily biased toward lynx that spend time near roads. During the summer, the entire range can be accessed. Trained dogs have been successfully employed to find bear (*Ursus* spp., Wasser et al. 2004) scat and could likely be trained to find lynx scat as well.

When designing a sampling strategy, a balance must be struck between sampling the entire area of interest and sampling areas with denser lynx abundance in order to collect enough samples. It would be useful to stratify the sampling into two strata. One stratum would be drainages and the second would be highlands. Lynx use drainages heavily. Therefore, more effort would be placed in this stratum to maximize the number

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

of scats detected. Highlands would be sampled less intensively. Less sampling in the highlands avoids placing too many resources into areas of low lynx abundance, but still allows the lynx in that habitat to have a non-zero sampling probability and inference to be made to the area.

Analysis will be performed using a suite of robust design models. A standard robust design model (Kendall et al. 1997) and a robust design with a genotyping error parameter (Chapter 2) will both be considered. Model selection and model averaging will be based on information–theoretic methods (Burnham and Anderson 2002). Both of these model types allow design flexibility to model differences in detection, survival, and abundance across time, groups and other covariates.

I feel monitoring lynx abundance and survival is feasible in Colorado. Non-invasive sampling of lynx scat using trained dogs offers a viable way to collect the necessary field data. DNA-based capture–recapture theory can be applied to obtain reliable estimates of demographic parameters.

ACKNOWLEDGMENTS

The Colorado Division of Wildlife and Great Outdoors Colorado provided support for this work. M. K. Schwartz and K. Pilgrim performed the genetic analyses.

LITERATURE CITED

Burnham, K. P. and D. R. Anderson. 2002. Model selection and multimodel inference: a practical information-theoretic approach. Springer-Verlag, New York, NY, USA.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Carmichael, L. E., W. Clark, and C. Strobeck. 2000. Development and characterization of microsatellite loci from lynx (*Lynx canadensis*), and their use in other felids. Molecular Ecology 9: 2197–2198.

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.

Eggert, L. S., J. A. Eggert, and D. S. Woodruff. 2003. Estimating population sizes for elusive animals: the forest elephants of Kakum National Park, Ghana. Molecular Ecology 12: 1389–1402.

Goodall, J. and M. Bekoff. 2002. The ten trusts. Harper Collins Publishers, New York, NY, USA.

Kendall, W. L., J. D. Nichols, and J. E. Hines. 1997. Estimating temporary emigration using capture–recapture data with Pollock's robust design. Ecology 78: 563–578.

McKelvey, K. S. 2002. Statement before the Committee on Resources, U. S. House of Representatives concerning National Lynx Survey.

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.

McDaniel, G. W., K. S. McKelvey, J. R. Squires, and L. F. Ruggiero. 2000. Efficacy of lures and hair snares to detect lynx. Wildlife Society Bulletin 28: 119–123.

Paetkau, D. 2003. An empirical exploration of data quality in DNA-based population inventories. Molecular Ecology 12: 1375–1387.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Palsbøll, P. J., J. Allen, M. Bérubé, P. J. Clapham, T. Feddersen, P. S. Hammond, R. R. Hudson, H. Jørgensen, S. Katona, A. H. Larsen, F. Larsen, J. Lien, D. K. Mattila, J Sigurjónsson, R. Sears, T. Smith, R. Sponer, P Stevick and N. Øien. 1997. Genetic tagging of humpback whales. Nature 388: 767–769.

Schwartz, M. K., K. L. Pilgrim, K. S. McKelvey, E. L. Lindquist, J. J. Claar, S. Looch, and L. F. Ruggiero. 2004. Hybridization between Canada lynx and Bobcats: genetic results and management implications. Conservation Genetics 5: 349–355.

Shenk, T. M. 2001. Post-release monitoring of lynx reintroduced to southwestern Colorado. Colorado Division of Wildlife, Federal Aid Reports WP-0670-T-1. Fort Collins, CO.

Shenk, T. M. 2004. Post-release monitoring of lynx reintroduced to southwestern Colorado. Colorado Division of Wildlife, Federal Aid Reports WP-0670-T-1. Fort Collins, CO.

U. S. Fish and Wildlife Service. 2000. Endangered and threatened wildlife and plants: final rule to list the contiguous United States distinct population segment of the Canada lynx as a threatened species. Federal Register 65, Number 58.

Wasser, S. K., B. Davenport, E. R. Ramage, K. E. Hunt, M. Parker, C. Clarke, and G. Stenhouse. 2004. Scat detection dogs in wildlife research and management: application to grizzly and black bears in the Yellowhead Ecosystem, Alberta, Canada. Canadian Journal of Zoology 82: 475–492.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

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.

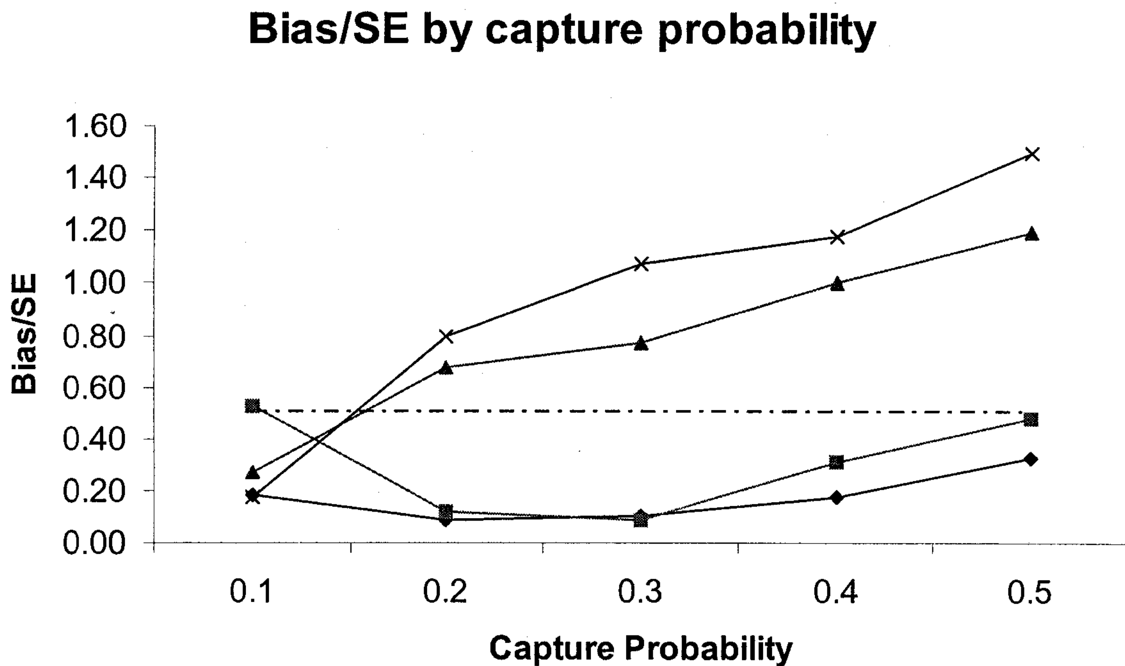
Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 1. Amplification rates and error rates per sample for 2 sets of plucked hair samples from Canada lynx released in Colorado. The 100 samples were submitted in the first round and 33 in the second. Genotyping error was computed as the ratio of the number incorrect genotypes over the number of samples that amplified at four or more loci.

	Amplification rate			
Loci Amplified	Round 1	SE	Round 2	SE
≥4	0.20	0.04	0.21	0.07
≥5	0.15	0.04	0.15	0.06
6	0.10	0.03	0.03	0.03
Error rate	0.05	0.05	0.00	0.00

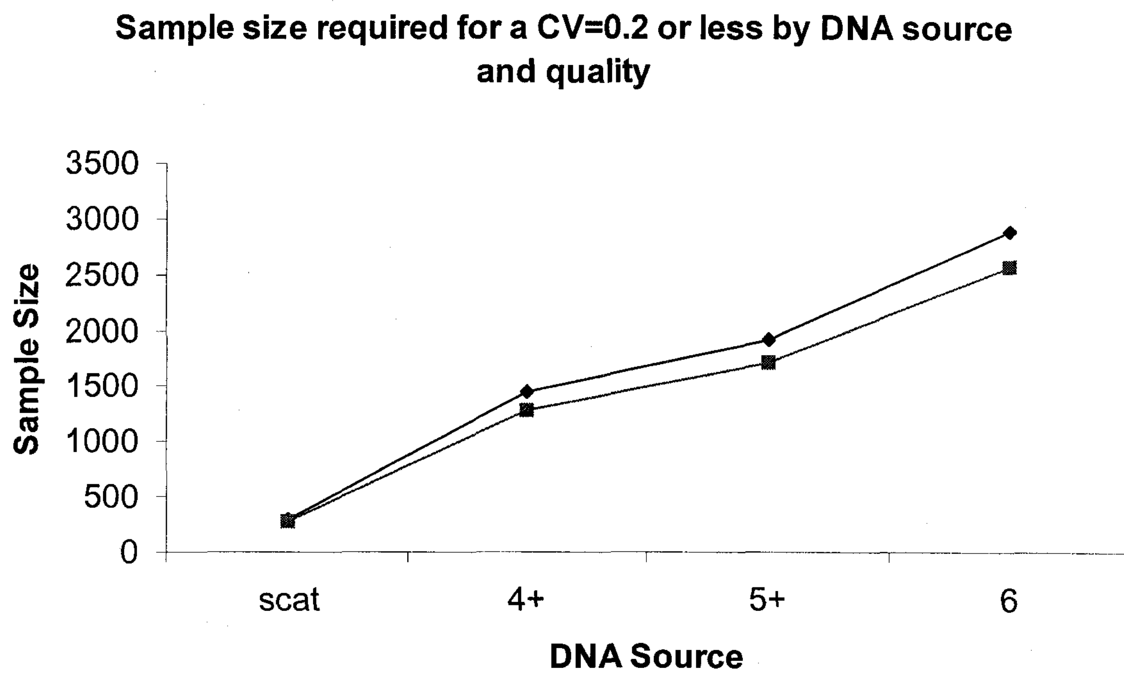
Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Figure 1. Absolute bias divided by standard error of estimated abundance by per occasion capture probability for a simulated population of 200 lynx designed to mimic the Colorado lynx population using a time varying model of capture probability. The diamond represents sampling 5 occasions and estimating genotyping error. The square symbol represents sampling 7 occasions and estimating genotyping error. The triangle symbol represents sampling 5 occasions, but not estimating genotyping error. The x represents sampling 7 occasions and not estimating genotyping error. Bias/SE exceeding 0.5 (dashed line) represents situations where confidence interval coverage is appreciably less than the intended level.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Figure 2. Sample size required to achieve a coefficient of variation on estimated abundance of ≤ 0.2 when the DNA source is scat or hair and per occasion capture probability is 0.2 (diamond) and 0.25 (square). For the hair samples, I considered examining samples that amplified four or more loci, five or more loci and all six loci. True population size is assumed to be 200 animals.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

CHAPTER 5

A REVIEW OF CAPTURE–RECAPTURE METHODS APPLICABLE TO DNA-BASED NONINVASIVE SAMPLING

Abstract: The use of noninvasive DNA sampling to identify individual animals for capture–recapture studies has become widespread in the past decade. Strong emphasis has been placed on the field protocols and genetic analyses with fruitful results. Little attention has been paid to the capture–recapture application for this specific type of data beyond stating the effects of assumption violations. Here I review the broad class of capture–recapture methods that are available for use with DNA-based capture–recapture data, noting the vast array of biologically interesting parameters that can be estimated from such data. I also present an overview of recent developments in capture–recapture theory specifically designed for DNA-based data.

Key words: abundance; capture–recapture; survival; microsatellites; sampling.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

INTRODUCTION

In recent years, the use of noninvasive sampling of DNA for capture–recapture studies has become widespread. The methods have been applied to a diverse array of taxa such as bears (*Ursus* spp., Woods et al. 1999), African elephants (*Loxodonta cyclotis,* Eggert et al. 2003), coyotes (*Canis latrans*, Kohn et al. 1999), humpback whales (*Megaptera novaeangliae*, Palsbøll et al. 1997), and painted turtles (*Chrysemys picta* Pearse et al. 2001). Large advances have been made in field protocols (Woods et al. 1999) and genetic analyses (Paetkau 2003).

Most DNA-based capture–recapture studies follow the same basic principles. First, samples containing DNA are collected at several points in time. Often the samples are noninvasive. Hair, feathers or feces are collected in a way that does not require physical contact of the animal. DNA is extracted from the samples and amplified at several microsatellite loci. Matching genotypes are considered to arise from the same individual and classified as recaptures. The data are then analyzed in a capture–recapture framework.

Little attention has been placed on the capture–recapture analysis of DNA-based data. A few studies have examined the effects of assumption violations on estimators of population abundance (Mills et al. 2000, Waits and Leberg 2000). Analyses focused on simple estimators of abundance. Only recently have capture–recapture models been specifically developed for use with the special set of circumstances that arise with DNA-based sampling (Chapters 1, 2 and 3).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Here I present a review of the literature on DNA-based capture–recapture studies. Thorough review papers dealing the genetics and field protocols already have been written (Mills et al 2000, Taberlet et al. 1999, Taberlet and Luikart 1999), therefore I choose to point out the advances that occurred since then and more heavily concentrate on data analysis issues. First, I provide an overview on the state of the genetics work in individual identification and where it is likely to head. Second, I focus on the unique features of DNA-based capture–recapture that separate it from standard tagging studies. Third, I present an overview of the broad array of analysis options available within capture–recapture theory that go far beyond simple estimates of abundance. Within this overview, I present recent developments in capture–recapture theory specifically designed for use with DNA-based data.

GENETICS FOR INDIVIDUAL IDENTIFICATION

Large progress has been made in the genetic techniques for identifying individual animals. The work done in ecology builds off a strong foundation of research in human forensics. The two fields together offer a lot of knowledge to the researcher pursuing a DNA-based capture-recapture study.

In the ecological literature, the potential for problems to arise in genotyping microsatellite loci has been demonstrated repeatedly. Creel et al. (2003) show high error rates in genotyping wolf (*Canis lupus*) fecal DNA. Kohn et al. (1999) also demonstrated errors occurring in genotypes from field collected fecal samples of coyotes. Similar results occur in laboratory test situations of plucked hair samples (Goosens et al. 1998).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The problems faced in genetic analysis of noninvasive samples fall into three categories: 1) amplification failure, 2) allelic dropout and 3) *in vitro* mutation. Amplification failure is failure of the polymerase chain reaction (PCR) to produce copies of the intended piece of the genome. The failure may be caused by a lack of the target DNA or by PCR inhibitors present in the sample. From a capture–recapture standpoint, amplification failure is an easy problem to deal with because it simply results in no data from a sample analogous to never having obtained the sample. In capture–recapture studies, it is typically assumed that capture probability is less than 1. Therefore, amplification failure can be absorbed into capture probability. Allelic dropout occurs when one allele at a locus fails to amplify or is not present in the pipetted DNA sample. The result causes the sample to appear to be homozygous at the locus when it may actually be heterozygous. This can be problematic for capture–recapture studies if it results in an incorrect genotype. It is possible, although rare, for an allele to mutate early in the PCR process and create a spurious allele. This could be problematic if an incorrect genotype results, but the rarity of the occurrence minimizes the risk.

Efforts have been made to limit genotyping error as much as possible. Paetkau (2003) shows that with careful laboratory protocol including culling poor quality samples, a reliable set of microsatellite loci and experience, genotyping error rates can be greatly decreased and in some cases virtually eliminated. Extensive culling of lesser quality samples may be a double edged sword for capture–recapture studies. On one hand, the final genotypes produced are more reliable. On the other hand, it has been demonstrated in human forensics that individual humans leave varying amounts of DNA. Some

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

individuals termed “shedders” tend to leave much more DNA than non-shedders (Alessandrini et al. 2002, Lowe et al. 2002). Likely, animals also vary by individual in their cell shedding rate. Therefore, additional heterogeneity in capture probability is likely added as low quality samples are removed.

Explicit tests have been developed to determine whether error is present in a set of genotypes and the relative amount of error (McKelvey and Schwartz 2004a). First, the examining bimodality (EB) test looks for an over-abundance of genotypes observed only once (McKelvey and Schwartz 2004a). When errors arise in the genotyping process, they typically lead to unique observed genotypes. Second, the difference in capture history (DCH) test examines the rate at which new individuals are recognized by looking at more loci (McKelvey and Schwartz 2004a). If the rate of adding more individuals exceeds that expected by increasing resolution among individuals alone, then genotyping error is likely. These tests combined provide the researcher with useful information about the quality of the genotypes. The tests both assume equal capture probability among individuals. The effectiveness of the tests when capture probability is heterogeneous has not been determined.

An additional way to increase amplification rate and reduce error is multiplex pre-amplification (Piggot et al. 2004). For this approach, an initial large volume PCR run is performed on the original sample with primers for the entire PCR panel. Then, a portion of the result of the first run is used in a second PCR with each primer individually. Piggot et al. (2004) demonstrate that the multiplex pre-amplification method greatly

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

increases amplification rate for samples nearing the lower bound of template DNA available.

In human forensics, tetranucleotide microsatellites are used because they are less likely to produce errant genotypes (Eckert et al. 2002). Development of tetranucleotide markers for most wildlife species may be impractical because the increased cost and effort of locating polymorphic loci. Future develops in techniques for finding microsatellites and developing microsatellite primers will likely make using tetranucleotides more practical.

Probability of identity (P_{ID}) is commonly computed in ecological studies as evidence of the resolution power of a set of microsatellites (Woods et al. 1999, Waits et al. 2001). P_{ID} is defined as the probability that two individuals selected at random from a population share the same genotype. This is a useful quantity for determining the resolution of the set of microsatellite loci prior to undertaking a capture–recapture study. A researcher strives to use a set of microsatellites that have a low P_{ID} without using too many loci which could become cost prohibitive. Therefore, a tradeoff exists between minimizing P_{ID} and available funding (Creel et al. 2003).

P_{ID} appears to be frequently misused as evidence that matching genotypes arise from the same individual after samples have been collected. A more useful quantity is the weight of evidence as to whether two matching samples arose from the same individual given you have them in hand. The weight of evidence is the likelihood ratio of the conditional probability of the samples matching given the hypotheses that 1) they are the same individual and 2) they are from different individuals (Evett and Weir 1998).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Moreover, the conditional probability statements allow for precise definition of the hypotheses. Without having at least two hypotheses for how the genotypes arose, there is no way to judge the relative merit of a single hypothesis (Evett and Weir 1998:29). The conditional probabilities are computed for each genotype, therefore variation in match probabilities among genotypes can be explored. Computing match probabilities by genotype is also important because the probability of a genotype being repeated in the population is increased when the genotype is detected in the population (Weir 2001). Finally, conditional match probabilities can be computed for incomplete genotypes and therefore allow the use of more data.

ISSUES FACING DNA-BASED CAPTURE–RECAPTURE

The unique issues confronted in a DNA-based capture–recapture study can be split into two general categories. First, the list of marks in the population is unknown. Second, the concept of a sampling occasion is poorly defined. This sounds daunting for a capture–recapture study, but both of these issues can be overcome when properly addressed.

In a standard capture–recapture study, the identifying marks placed on animals captured are known. In a DNA-based capture–recapture study, every animal in the population is self-marked prior to the study beginning because every individual has a genotype. Unfortunately, the researcher does not know what genotypes (marks) are in the population. Not knowing the list of marks in the population leads to two problems: 1) it is difficult to differentiate between correctly and incorrectly read marks and 2) marks may not be unique.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

When it is difficult to determine if a mark is read correctly, identification errors are more likely to be introduced into the data. In standard capture–recapture studies, when a mark is read that does not match a mark from the known list, the researcher ignores or corrects the observation. Thus few errors occur. In rare situations where one mark is misread as a different existing mark, there is little impact on the resulting parameter estimates because these is little change in the sufficient statistics. In DNA-based capture–recapture, no list of marks exists and therefore, incorrect genotypes are added as new individuals. Even small amounts of this error can have substantial impacts on parameter estimates (Wait and Leberg 2000, Creel et al. 2003). In some cases errors can largely be eliminated through strict protocols in the DNA analysis (Paetkau 2003), while in other cases errors persist (Creel et al. 2003).

Marks may not be unique when the list of marks in the population is self-assigned as it is in DNA-based identification. When animals share a common mark, they appear to be only one animal. This has been dubbed the 'shadow effect' (Mills et al. 2000). The shadow effect results in under-estimation of abundance and over-estimation of survival. To a large extent, this problem can be overcome by using a set of microsatellite loci with high power to resolve individuals. Unfortunately, using a larger set of loci results in an increased probability of genotyping error.

In most DNA-based capture–recapture studies, animals are passively detected through observations of hair, scat or feathers. This introduces some uncertainty as to when the animal left the sample. It also allows multiple samples to be observed within a

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

single sampling occasion. Both of these issues result in the concept of a sampling occasion to be blurred from the traditional definition of a sampling occasion.

When the time of deposition of a sample is unknown, the sampled population may be poorly defined. Typically, researchers set criteria for the age of a sample. In other cases, the design determines when a sample was deposited, for example hair left on a snag must have been left after the snag was set. This helps better define the population. Even so, the estimated population remains the number of animals that have used the sampled area within the time since the cutoff of sample decay. Therefore, the number of individuals that left samples may exceed the number of individuals currently in the population. This results in a type of closure violation.

In addition to uncertain timing of when samples are left, multiple samples from the same individual may be encountered within an occasion. This situation is largely avoided with active sampling because trapped animals generally cannot be trapped again until released by the researcher and visually detected animals are ignored if seen again before the next sampling occasion. Conversely, with DNA-based methods, multiple samples from the same individual are often collected and the identity is not known until a considerable monetary and time investment is put into the sample. Through this large investment, an additional form of data are available in these sampling schemes that has not been previously dealt with in capture–recapture analysis.

The concept of a capture occasion can be unclear in studies that repeat sampling across space rather than time. For example, Eggert et al. (2003) sample African elephant dung in different sections of there study area for each occasion. This sampling scheme

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

could add additional heterogeneity in capture probability. When samples are taken across space rather than time, some animals may be unavailable for sampling in some locations.

Some basic designs have been established that help to better define an occasion. Woods et al. (1999) describe a hair snag protocol for bears. They suggest placing the hair snags in the field and checking at regular intervals, much as would be done with a standard trap except the interval tends to be longer. Other options include sampling a set of transects or quadrats for dung. The transects themselves could be considered the occasions or could be sampled multiple times. Using the transects as occasions leads to a lack of uniqueness in the "time" ordering of the encounter history. Another design used is to sample continuously and set the number of occasions as the maximum number of times a single genotype is observed (Øystein et al. 2004). This method is not desirable because it greatly limits analysis options.

OPTIONS AVAILABLE IN CAPTURE–RECAPTURE ANALYSIS

Capture–recapture theory has been a field of intense research for the past century. Advances and extensions of capture–recapture models allow a vast array of biologically interesting parameters to be estimated. To cover this in detail would require a book, therefore I restrict the discussion to the types of questions that can be answered with DNA-based capture–recapture data. Currently, only a small piece of capture–recapture theory has been exploited in DNA-based studies.

Animal abundance in a closed population at one point in time is the target parameter of most DNA-based capture–recapture studies. A vast theory exists on how to estimate population size under different sets of assumptions (Otis et al. 1978, Huggins

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

1989, Pledger 2000, Borchers et al. 2002, Williams et al. 2002). Much of the current literature applying capture–recapture analysis to DNA-based data restricted analysis to simple cases of the models in Otis et al. (1978) and the jackknife estimator (Burnham and Overton 1979). Advances in closed population capture–recapture allow heterogeneity in capture probability to be modeled in more flexible frameworks. The conditional likelihood models of Huggins (1989, 1991) allow capture probability to be modeled as a function of individual covariates. Finite mixture models approximate individual heterogeneity in capture probability with group differences in capture probability (Norris and Pollock 1996, Pledger 2000). Continuous mixture models allow individual heterogeneity to be modeled on a continuous scale (Dorazio and Royle 2003).

Two extensions to the closed population capture–recapture models have been developed to specifically address issues faced with DNA-based data. First, an extension to the models of Otis et al. (1978), Huggins (1989) and Pledger (2000) has been developed that includes a parameter to estimate genotyping error (Chapter 1). The method uses the disproportionate number of genotypes only observed once relative to genotypes seen more than once to estimate genotyping error. This model allows abundance to be properly estimated in the presence of genotyping error. The method provides a major advance because it eliminates the debate over whether genotyping error is problematic in DNA-based estimates of abundance. The debate has been raging in recent literature (McKelvey and Schwartz 2004a, Paetkau 2004, McKelvey and Schwartz 2004b) without clear resolution. The models from Chapter 1 allow standard capture–recapture models to be compared directly to models expanded to estimate

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

genotyping error in an information-theoretic framework. Therefore, the data guide the decision as to whether genotyping error is important or not. The results can then be model averaged for a further reduction in bias (Burnham and Anderson 2002).

A second method was developed to take advantage of the additional data available when multiple dung samples are collected from the same individual within an occasion. In Chapter 3, I developed an estimator which uses this data to help account for the individual heterogeneity in capture probability caused by varying numbers of dung piles available. They assume numbers of dung piles deposited by animals follow either a Poisson or negative binomial distribution. Counts of dung piles per individual come from a combination of the numbers per individual and the probability of detecting a dung pile. Based on this, I fit mixture models to estimate the total unobserved abundance of dung, detection probability per dung pile, and abundance of animals. They demonstrate the effectiveness of the method through simulation and an example using African elephant dung data.

Accumulation curve analysis has been used to estimate abundance from noninvasive DNA data (Kohn et al. 1999, Eggert et al. 2003). Capture–recapture analysis has several advantages over accumulation curves. First, accumulation curves do not account for the sampling design used to obtain the data (Cam et al. 2002). Accumulation curves are merely designed to approximate the appearance of the data. Capture–recapture models directly estimate detection probability. Second, accumulation curves do not efficiently use the data collected. Accumulation curves only use the first detection of an individual, whereas capture–recapture methods can use all detections. Finally,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

accumulation curves cannot account for variation in detection probability. Detection probability is known to vary widely in many situations, therefore methods need to be able to account for these differences in order to appropriately estimate abundance. Therefore, accumulation curves are not recommended for estimating abundance.

Estimating abundance from DNA-based capture–recapture methods is likely to be most useful for relatively small populations, up to a few thousand individuals. Beyond this size, a very large number of samples would have to be collected and analyzed making the study cost prohibitive. For large populations, other methods such as double sampling (Thompson 2002), mark-resight sampling (White 1996), or distance sampling (Buckland et al. 2001) are likely to be more efficient.

Abundance is just one parameter that can be estimated from capture–recapture data. Models have been developed to estimate survival, emigration rates, movement or transition rates, fecundity and population growth. As more surveys continue across time, the desire to estimate more population parameters will increase. The Jolly-Seber (Jolly 1965, Seber 1965) model provides the basis from which much of open capture–recapture theory is built upon. Much of the work in this field is presented in proceedings from the EURING technical conferences (Morgan and Thomson 2002, Baillie and North 1999, North and Nichols 1995).

The Cormack-Jolly-Seber model (CJS, Cormack 1964, Jolly 1965, Seber 1965) allows estimates of apparent survival, the probability of surviving and remaining on the study area, of individuals conditioned on the individual being captured and marked at least once. The CJS model is useful for DNA-based capture–recapture when genotyping

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

error is not present and the population is only sampled once per year or other relevant time interval. The estimates of survival from a CJS model are robust to individual heterogeneity in capture probability.

The robust design is an obvious way to bridge the gap from single year abundance estimates to estimates of survival and emigration rates (Pollock 1982, Kendall et al. 1997, Schwarz and Stobo 1997). The robust design uses two types of sampling periods. Primary sampling periods are separated by long intervals across which survival and emigration rates are estimated. Secondary periods occur at the transition between primary periods. The secondary periods are short and used for better estimating capture probability. The secondary periods may assume a closed population (Kendall et al 1997) or an open population (Schwarz and Stobo 1997, Kendall and Bjorkland 2001). The robust design allows abundance to be estimated at each primary period. Survival and emigration rates can be estimated between primary periods.

The robust design has been extended to handle situations where genotyping error is present. In Chapter 2, I developed a robust design model that estimates genotyping error rate at each primary sampling occasion. By doing so, the model properly estimates abundance and survival. Without accounting for genotyping error, if present, abundance would be biased high and survival would be biased low. The robust design is required to extend the CJS model to estimate genotyping error because the information needed to estimate genotyping error is in repeated detections of individuals within a primary occasion.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Another particularly useful advance in capture–recapture theory is that of multi-state and multi-strata models (Schwarz et al. 1993, Lebreton and Pradel 2002). The multi-state or strata models allow one to estimate transition rates among conditions or locations. For example, one might wish to estimate the rate at which individual animals move among subpopulations. The multi-strata models have been expanded to fit into a robust design sampling scheme as well (Kendall and Nichols 2002).

An application of capture–recapture theory of particular interest to ecologists is the direct estimation of population growth rate. Typically, this would be done by computing the dominant eigenvalue of a matrix projection model such as a Lefkovitch or Leslie matrix (Caswell 2001). Such a method requires age and/or state specific estimates of survival and fertility. With the method developed by Pradel (1996), population growth rate can be estimated from capture–recapture data. An exciting step beyond estimating population growth rate is to estimate an individual animals contribution to population growth. This opens capture–recapture theory up to answer exciting evolutionary biology questions. Link et al. (2002) present a Bayesian approach combining capture–recapture theory and matrix modeling to estimate the latent individual fitness of an animal.

A recent extension of capture–recapture theory is the class of site occupancy models (MacKenzie et al. 2002). These models change the sampling unit from the individual animal to the plot. The data consist of presence or absence of an animal at each plot across sampling occasions. The models allow for imperfect detection of animals. The model estimates the proportion of plots occupied during the study. These data are particularly useful for large scale studies or surveys of rare species. MacKenzie

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

et al. (2003) extended the method to a robust design framework which allows colonization to be estimated in addition to occupancy rate. MacKenzie et al. (2004) further extended the model to handle more than one species and investigate patterns of co-occurrence.

A major advance in capture–recapture theory for application to answering ecological questions is placing the models in a general linear models (GLM) framework (Lebreton et al. 1992). The GLM framework allows parameters to be modeled as functions of covariates. The linear model is then tied to the parameters with a link function. This allows biological hypotheses to be examined and complicated models to be built in parsimonious ways. For example, a 20 year time trend on survival could be modeled with only 2 parameters in a GLM framework, rather than 20 parameters if each annual survival is estimated separately.

Multiple sources of data can be easily combined in capture–recapture analysis. For example, a segment of a population being sampled with DNA-based methods may also be radio collared. Combining the radio telemetry data with a robust design DNA-based survey could provide more information about survival and abundance than either survey could provide alone. Multiple sources of data also arise from laboratory and field data in DNA-based studies. Lab estimates of genotyping error can help better estimate abundance in the wild (Chapter 2). In addition, harvest data can be used as a source of dead recovery data and known DNA source.

Laboratory estimates of genotyping error can be particularly useful in reducing sampling variance in DNA-based capture-recapture estimates. Using a joint likelihood

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

with a binomial estimator of error rate and the capture-recapture likelihood as is done in Chapter 2 produces standard errors on estimated abundance that are 20% smaller than without the additional data for a 5-occasion case with a capture probability of 0.2 and 75 lab trials. Little additional reduction occurs from adding more lab samples because the variance from the detection process constitutes the vast majority of the remaining variation. Estimates of the standard error of genotyping error rate are reduced by 60% when 75 lab samples are tested.

Software exists for a wide range of capture–recapture analyses. Nearly all model types included in this review are available in Program MARK (White and Burnham 1999). Program M-SURGE is available for analysis of multi-state problems (Choquet et al. 2003). Program PRESENCE performs site occupancy estimation (MacKenzie et al. 2002). For unique problems, models can be coded in SAS (SAS Institute 2004), R (R Project 2004) or similar programable statistics software for greater flexibility.

CONCLUSION

The stage is set for a fruitful expansion of DNA-based capture–recapture studies to explore deeper ecological questions. Effective field protocols have been designed for a variety of organisms. Technology and techniques exist for reliable genotyping from samples that contain low numbers of copies of the organism's DNA. Capture–recapture theory exists to estimate an assortment of biologically important parameters and to cope with the situations faced with DNA-based sampling. The merging of the three areas is occurring rapidly.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

While abundance is an important starting point for many investigations, especially for rare and endangered species, it should not be the end point of a study. Now it is time to move to broader ecological questions.

ACKNOWLEDGMENTS

The Colorado Division of Wildlife, Great Outdoors Colorado and Dr. Tanya Shenk provided support for this review.

LITERATURE CITED

Baillie, S. R., and P. M. North. 1999. Large-scale studies of marked birds: proceedings of the EURING97 conference. Bird Study 46 supplement.

Borchers, D. L., S. T. Buckland, and W. Zucchini. 2002. Estimating animal abundance: closed populations. Springer-Verlag, New York, NY, USA.

Buckland, S. T., D. R. Anderson, K. P. Burnham, J. L. Laake, D. L. Borchers, and L. Thomas. 2001. Introduction to distance sampling. Oxford University Press, Oxford, U.K.

Burnham, K. P., and D. R. Anderson. 2002. Model selection and multimodel inference: an information theoretic approach. Springer–Verlag, New York, NY, USA.

Burnham, K. P., and W. S. Overton. 1979. Robust estimation of population size when capture probabilities vary among individuals. Ecology 60: 927–936.

Cam, E., J. D. Nichols, J. R. Sauer and J. E. Hines. 2002. On the estimation of species richness based on the accumulation of previously unrecorded species. Ecography 25: 102–108.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Caswell, H. 2001. Matrix population models second edition. Sinauer Associates, Inc. Sunderland, MA, USA.

Choquet R., A.-M. Reboulet, R. Pradel, O. Gimenez, J.-D. Lebreton. 2003. M-SURGE - New software for multistate recapture models.

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.

Cormack, R. M. 1964. Estimates of survival from the sighting of marked animals. Biometrika 51: 429–438.

Dorazio, R. M. and J. A. Royle. 2003. Mixture models for estimating the size of a closed population when capture probabilities vary among individuals. Biometrics 59: 351–364.

Eckert, K. A., A Mowery, and S. E. Hile. 2002. Misalignment-mediated DNA polymerase β mutations: comparison of microsatellite and frame-shift error rates using a foward mutation assay. Biochemistry 41: 10490–10498.

Eggert, L. S., J. A. Eggert, and D. S. Woodruff. 2003. Estimating population sizes for elusive animals: the forest elephants of Kakum National Park, Ghana. Molecular Ecology 12: 1389–1402.

Evett, I. W. and B. S. Weir. 1998. Interpreting DNA evidence: statistical genetics for forensic scientists. Sinauer Associates Inc., Sunderland, MA, USA.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Goosens, B., L. P. Waits, and P. Taberlet. 1998. Plucked hair samples as a source of DNA: reliability of dinucleotide microsatellite genotyping. Molecular Ecology 7: 1237–1241.

Huggins, R. M. 1989. On the statistical analysis of capture experiments. Biometrika 76: 133–140.

Huggins, R. M. 1991. Some practical aspects of a conditional likelihood approach to capture experiments. Biometrics 47: 725–732.

Jolly, G. M. 1965. Explicit estimates from capture–recapture data with both death and immigration-stochastic model. Biometrika 52: 225–247.

Kendall, W. L., J. D. Nichols, and J. E. Hines. 1997. Estimating temporary emigration using capture–recapture data with Pollock's robust design. Ecology 78: 563–578.

Kendall, W. L. and R. Bjorkland. 2001. Using open robust design models to estimate temporary emigration from capture–recapture data. Biometrics 57: 1113–1122.

Kendall, W. L. and J. D. Nichols. 2002. Estimating state–transition probabilities for unobservable states using capture–recapture/resighting data. Ecology 83: 3276–3284.

Kohn, M. H., E. C. York, D. A. Kamradt, G. Haught, R. M. Sauvajot, R. Wayne. 1999. Estimating population size by genotyping faeces. Proceedings of the Royal Society of London, Series B 266: 657–663.

Lebreton, J.-D., K. P. Burnham, J. Clobert, and D. R. Anderson. 1992. Modeling survival and testing biological hypotheses using marked animals: a unified approach with case studies. Ecological Mongraphs 62: 67–118.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Lebreton, J.-D. and R. Pradel. 2002. Multistate recapture models: modelling incomplete individual histories. Journal of Applied Statistics 29: 353–369.

Link, W. A., E. G. Cooch, and E. Cam. 2002. Model-based estimation of individual fitness. Journal of Applied Statistics 29: 207–224.

Lowe, A., C. Murray, J. Whitaker, G. Tully, and P. Gill. 2002. The propensity of individuals to deposit DNA and secondary transfer of low level DNA from individuals to inert surfaces. Forensic Science International 129: 25–34.

MacKenzie, D. I., J. D. Nichols, G. B. Lachman, S. Droege, J. A. Royle, and C. A. Lantimm. 2002. Estimating site occupancy rates when detection probabilities are less than one. Ecology 83: 2248–2255.

MacKenzie, D. R., J. D. Nichols, J. E. Hines, M. G. Knutson, and A. B. Franklin. 2003. Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. Ecology 84: 2200–2207.

MacKenzie, D. I., L. L. Bailey, and J. D. Nichols. 2004. Investigating species co-occurrence patterns when species are detected imperfectly. Journal of Animal Ecology 73: 546–555.

McKelvey, K. S., and M. K. Schwartz. 2004a. Genetic errors associated with population estimation using non-invasive molecular tagging: problems and new solutions. Journal of Wildlife Management 68: 439–448.

McKelvey, K. S., and M. K. Schwartz. 2004b. Providing reliable and accurate genetic capture–mark–recapture estimates in a cost-effective way. Journal of Wildlife Management 68: 453–456.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Mills, L. S., J. J. Citta, K. P. Lair, M. K. Schwartz, D. A. Tallmon. 2000. Estimating animal abundance using noninvasive DNA sampling: promise and pitfalls. Ecological Applications 10: 238–294.

Morgan, B. J. T., and D. L. Thomson. 2002. Statistical analysis of data from marked bird populations. Journal of Applied Statistics 29 (1-4).

Norris, J. L. and K. H. Pollock. 1996. Nonparametric MLE under two closed capture–recapture models with heterogeneity. Biometrics 52: 639–649.

North, P. M. and J. D. Nichols. 1995. Statistics and ornithology. Journal of Applied Statistics 22 (5-6).

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.

Øystein, F., E. Hedmark, A. Landa, J. Brøseth, J. Persson, R. Andersen, P. Segerström, and H. Ellegren. 2004. Colonization history and noninvasive monitoring of a reestablished wolverine population. Conservation Biology 18: 676–688.

Paetkau, D. 2003. An empirical exploration of data quality in DNA-based population inventories. Molecular Ecology 12: 1375–1387.

Paetkau, D. 2004. The optimal number of markers in genetic capture–mark–recapture studies. Journal of Wildlife Management 68: 449–452.

Palsbøll, P. J., J. Allen, M. Bérubé, P. J. Clapham, T. Feddersen, P. S. Hammond, R. R. Hudson, H. Jørgensen, S. Katona, A. H. Larsen, F. Larsen, J. Lien, D. K. Mattila, J Sigurjónsson, R. Sears, T. Smith, R. Sponer, P Stevick and N. Øien. 1997. Genetic tagging of humpback whales. Nature 388: 767–769.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Pearse, D. E., C. M. Eckerman, J. Janzen, and J. C. Avise. 2001. A genetic analogue of 'mark-recapture' methods for estimating population size: an approach based on molecular parentage assessments. Molecular Ecology 10: 2711–2718.

Piggot, M. P., E. Belleman, 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.

Pledger, S. 2000. Unified maximum likelihood estimates for closed capture–recapture models using mixtures. Biometrics 56: 434–442.

Pollock, K. H. 1982. A capture–recapture design robust to unequal probability of capture. Journal of Wildlife Management 46: 752–757.

Pradel, R. 1996. Utilization of capture–mark–recapture for the study of recruitment and population growth rate. Biometrics 52: 703–709.

R Project. 2004. R version 2.0.1. www.r-project.org.

SAS Institute, Inc. 2004. SAS version 9.1. Cary, North Carolina, USA.

Schwarz, C. J., J. F. Schweigert, and A. N. Arnason. 1993. Estimating migration rates using tag-recovery data. Biometrics 49: 177–193.

Schwarz, C. J. and W. T. Stobo. 1997. Estimating temporary migration using the robust design. Biometrics 53: 178–194.

Seber, G. A. F. 1965. A note on the multiple-recapture census. Biometrika 52: 249–259.

Taberlet, P., and G. Luikart. 1999. Non-invasive genetic sampling and individual identification. Biological Journal of the Linnean Society 68: 41–55.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Taberlet, P., L. P. Waits, and G. Luikart. 1999. Noninvasive genetic sampling: look before you leap. Trends in Ecology and Evolution 14: 323–327.

Thompson, S. K. 2002. Sampling: second edition. John Wiley and Sons, Inc. New York, NY.

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.

Weir, B. S. 2001. Forensics. *in* Handbook of statistical genetics. D. J. Balding, M. Bishop, and C. Cannings, editors. John Wiley and Sons, Inc., New York, NY, USA.

White, G. C. 1996. NOREMARK: population estimation from mark-resighting surveys. Wildlife Society Bulletin 24: 50–52.

White, G. C. and K. P . Burnham. 1999. Program MARK: survival estimation from populations of marked animals. Bird Study 46: S120–S139.

Williams, B. K., J. D. Nichols, and M. J. Conroy. 2002. Analysis and management of animal populations. Academic Press. New York, NY, 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.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

APPENDIX A

CAPTURE–RECAPTURE THEORY WHEN GENOTYPES ARE NOT UNIQUE

In DNA-based capture–recapture studies, the an individual's genotype is used as the mark. Individuals may not have unique genotypes. When individuals share a common genotype, they appear to be one individual. This phenomena is dubbed the shadow effect (Mills et al. 2000). This occurs because only a subset of the genome is examined when the individual is genotyped or in rare cases because of identical twins. Capture–recapture theory assumes that individuals are given unique marks. Without unique marks, it is difficult to determine the encounter history of each individual.

I developed a method incorporate multiple animals sharing the same genotype into capture–recapture theory. I used a Cormack–Jolly–Seber (CJS, Cormack 1964, Jolly 1965, Seber 1965) model to demonstrate the method. First, the match probability of each genotype is computed (Evett and Weir 1998). The match probability is the probability that two identical genotypes arose from different individuals. Second, given the encounters of a genotype in the study, all possible combinations of encounter histories are developed and the probability of each set of encounter histories is computed. Third, for each genotype, all possible combinations of encounter histories are compiled to make a data set for the population. Finally, a CJS analysis is run on each data set and the results

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

are weighted by the likelihood of each data set being the correct representation of the population.

While this method is easy to derive and code, physical computation is extremely impractical. I developed an Access database to manage all possible encounter histories and their probabilities of occurrence. The database also makes an input file for a SAS analysis. I developed SAS code to perform the analysis. Yet, for any problem of significance, the run time is tremendous. Frequently, for problems with roughly 10 individuals sharing a few genotypes and being sampling over 5 occasions, the number of possible combinations of encounter histories is enough to cause Access to run out of memory. Therefore, I abandoned this approach.

When dealing with individuals that share a common genotype, the problem should be addressed as a technological issue and not a statistical issue. Either the marker set should be changed or extended to allow better resolution among individuals. Even though examining more loci increases the overall rate of error in the genotype, it is a far easier problem to deal with than individuals sharing genotypes.

LITERATURE CITED

Cormack, R. M. 1964. Estimates of survival from the sighting of marked animals. Biometrika 51: 429–438.

Evett, I. W. and B. S. Weir. 1998. Interpreting DNA evidence: statistical genetics for forensic scientists. Sinauer Associates Inc., Sunderland, MA, USA.

Jolly, G. M. 1965. Explicit estimates from capture–recapture data with both death and immigration-stochastic model. Biometrika 52: 225–247.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Mills, L. S., J. J. Citta, K. P. Lair, M. K. Schwartz, D. A. Tallmon. 2000. Estimating animal abundance using noninvasive DNA sampling: promise and pitfalls. Ecological Applications 10: 238–294.

Seber, G. A. F. 1965. A note on the multiple-recapture census. Biometrika 52: 249–259.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

APPENDIX B

SUGGESTIONS FOR DNA-BASED CAPTURE–RECAPTURE ON GUNNISON SAGE-GROUSE

The Gunnison sage-grouse (*Centrocercus minimus,* Young et al. 2000) is a species of special concern in Colorado. It is also a candidate species for listing under the Endangered Species Act. Declines in range expanse, abundance and habitat availability led researchers to investigate the demography of the Gunnison sage-grouse.

DNA-based capture–recapture has been proposed as a possible way to sample grouse populations. Either feathers or feces could be collected in the field. DNA could then be extracted and amplified at a panel of microsatellite loci and a sexing locus. Considerable work has been completed on Gunnison sage-grouse population genetics and microsatellite primers have been developed (Oyler-McCance et al. 1999, Taylor et al. 2003).

We examined microsatellite data from Gunnison sage-grouse to evaluate the potential for implementation of a DNA-based capture–recapture study (S. J. Oyler-McCance unpublished data). The panel consisted of 8 dinucleotide microsatellite loci (Table 1) and one sexing locus. A total of 152 male and 112 female grouse from 8 populations were genotyped from blood samples.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Of the 264 grouse genotyped, there were 261 unique genotypes. The probability of selecting two matching genotypes from a population with the allele frequencies estimated from this population given all individuals are randomly mating, P(ID), is 0.000085. If the entire population were full sibs, P(ID-sibs)=0.015 (Waits et al. 2001). Thus, the number of matching genotypes suggests the population is highly related. The expected number of matching genotypes given the P(ID-sibs) is 4 and 3 were observed. This is also reflected in the biology of the Gunnison sage-grouse. Several of the populations sampled for this examination were from small isolated populations. At the current level of resolution, capture–recapture models would perform fairly well. Abundance would only be slightly under-estimated and survival slightly over-estimated. More microsatellite markers are being developed which will help provide a higher level of resolution (S. J. Oyler-McCance pers. comm.).

Recently, methods have been developed to extract DNA from sage-grouse feces (S. J. Oyler-McCance pers. comm.). Feces would likely be the most efficient source of DNA in terms of field collection. If abundance is the parameter of interest, the multiple detection model of Lukacs et al. (In prep) would be the most appropriate. The effort involved for a range-wide estimate of grouse abundance would likely be too great for a DNA-based study because the total number of grouse would lead to too many samples to be genotyped. This is especially true if feces from the same individual are encountered multiple times within a sampling occasion. DNA-based could be useful for investigations limited to smaller areas.

LITERATURE CITED

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Lukacs, P. M., K. P. Burnham, and L. S. Eggert. In prep. Estimating population size from dung-based DNA capture–recapture data. Journal of Applied Ecology.

Oyler-McCance, S. J., N. W. Kahn, K. P. Burnham, C. E. Braun, and T. W. Quinn. 1999. A population genetic comparison of large- and small-bodied sage grouse in Colorado using microsatellite and mitochondrial DNA markers. Molecular Ecology 8: 1457–1465

Taylor, S. E., S. J. Oyler-McCance, T. W. Quinn. 2003. Isolation and characterization of microsatellite loci in Greater Sage-Grouse (*Centrocercus urophasianus*). Molecular Ecology Notes 3: 262–264.

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.

Young, J. R., C. E. Braun, S. J. Oyler-McCance, J. W. Hupp, and T. W. Quinn. 2000. A new species of sage-grouse (Phasianidae: Centrocercus) from southwestern Colorado. Wilson Bulletin 112: 445–453.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 1. Dinucleotide microsatellite loci and number of alleles per locus for Gunnison sage-grouse in Colorado.

Locus	Number of Alleles
L1	2
S5	3
S9	16
S11	7
L3	3
L8	2
ADL230	6
L4	15

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

APPENDIX C

MICROSATELLITE GENOTYPES OF LYNX IN COLORADO

TABLE C1. Microsatellite genotypes from plucked hair samples of lynx released in 1999 and 2000. Each sample was genotyped 3 times after DNA extraction. Loci are listed in columns, alleles are given as the length (in base pairs) of the microsatellite. An 'x' indicates no alleles amplified at that locus.

Sample	Run	Lynx ID	Lc111	Lc111	Lc109	Lc109	Lc110	Lc110	Lc106	Lc106	Lc118	Lc118	Lc120	Lc120
1	1	AK99F15	152	152	172	180	94	94	104	104	145	145	196	196
1	2	AK99F15	150	152	172	180	94	94	104	104	145	145	196	196
1	3	AK99F15	150	152	172	180	94	94	96	104	145	145	196	196
2	1	AK99F17	150	150	172	180	98	100	98	104	139	143	200	200
2	2	AK99F17	150	150	172	180	98	100	98	104	139	143	200	200
2	3	AK99F17	150	150	172	180	98	100	98	104	139	143	200	200
3	1	AK99M6	x	x	x	x	98	100	98	98	143	143	x	x
3	2	AK99M6	150	150	x	x	98	98	x	x	143	143	x	x
3	3	AK99M6	152	154	x	x	106	106	102	102	x	x	x	x
4	1	AK99M9	152	156	172	178	94	98	104	104	143	147	196	204
4	2	AK99M9	152	152	172	178	94	98	104	104	143	147	196	204
4	3	AK99M9	152	156	172	178	94	98	104	106	143	147	196	204
5	1	AK99M9	152	156	172	178	94	98	102	104	x	x	196	204
5	2	AK99M9	152	156	x	x	94	98	102	104	144	144	196	204

5	**3**	AK99M9	152	156	172	178	94	98	102	104	x	x	196	204
6	**1**	AK00F4	x	x	176	176	98	100	100	104	x	x	200	200
6	**2**	AK00F4	x	x	172	172	98	100	104	104	x	x	x	x
6	**3**	AK00F4	150	150	x	x	98	100	x	x	x	x	x	x
7	**1**	BC99M2	154	154	x	x	102	104	98	102	x	x	x	x
7	**2**	BC99M2	x	x	x	x	94	102	98	98	x	x	x	x
7	**3**	BC99M2	x	x	178	178	98	104	102	102	x	x	x	x
8	**1**	AK00F5	x	x	x	x	x	x	102	102	x	x	x	x
8	**2**	AK00F5	x	x	x	x	x	x	x	x	x	x	x	x
8	**3**	AK00F5	x	x	180	180	x	x	x	x	x	x	x	x
9	**1**	BC99F8	x	x	x	x	98	98	100	100	143	143	x	x
9	**2**	BC99F8	x	x	x	x	98	98	94	102	147	147	x	x
9	**3**	BC99F8	x	x	x	x	102	102	96	104	137	145	x	x
10	**1**	BC99F9	152	154	174	176	100	104	100	104	145	145	200	200
10	**2**	BC99F9	152	154	174	176	100	104	100	104	145	145	200	200
10	**3**	BC99F9	152	154	174	176	100	104	100	104	145	145	200	200
11	**1**	BC00M12	x	x	x	x	98	98	102	102	x	x	x	x
11	**2**	BC00M12	x	x	x	x	x	x	x	x	144	144	x	x
11	**3**	BC00M12	x	x	x	x	x	x	104	104	x	x	x	x
12	**1**	BC00M12	x	x	x	x	100	100	102	102	x	x	198	198
12	**2**	BC00M12	x	x	x	x	x	x	98	98	x	x	x	x
12	**3**	BC00M12	x	x	x	x	x	x	x	x	x	x	x	x
13	**1**	BC00M16	x	x	x	x	x	x	98	104	x	x	196	196
13	**2**	BC00M16	x	x	x	x	x	x	x	x	x	x	x	x
13	**3**	BC00M16	x	x	x	x	x	x	102	102	x	x	x	x
14	**1**	BC00F6	x	x	x	x	98	98	98	102	x	x	204	204
14	**2**	BC00F6	x	x	x	x	102	102	96	102	x	x	x	x
14	**3**	BC00F6	x	x	x	x	x	x	x	x	x	x	x	x
15	**1**	BC00F18	x	x	x	x	98	98	102	102	x	x	x	x
15	**2**	BC00F18	x	x	x	x	x	x	102	102	x	x	x	x
15	**3**	BC00F18	x	x	x	x	x	x	102	102	x	x	x	x
16	**1**	BC00F14	x	x	x	x	96	96	x	x	x	x	x	x

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

16	**2**	BC00F14	152	152	x	x	x	x	100	102	x	x	x	x
16	**3**	BC00F14	x	x	178	178	x	x	100	102	x	x	x	x
17	**1**	BC00F18	x	x	180	180	96	96	100	100	143	143	196	196
17	**2**	BC00F18	x	x	x	x	92	96	100	100	x	x	x	x
17	**3**	BC00F18	x	x	172	172	x	x	100	106	x	x	x	x
18	**1**	BC00F19	x	x	x	x	96	98	102	102	x	x	x	x
18	**2**	BC00F19	x	x	x	x	96	98	102	102	x	x	x	x
18	**3**	BC00F19	x	x	x	x	x	x	x	x	x	x	x	x
19	**1**	YK99F1	x	x	x	x	102	104	100	104	x	x	200	200
19	**2**	YK99F1	150	154	x	x	x	x	102	102	137	141	x	x
19	**3**	YK99F1	150	150	x	x	x	x	96	100	145	145	x	x
20	**1**	YK99F4	150	152	172	178	94	98	98	104	143	145	200	200
20	**2**	YK99F4	150	152	172	178	94	98	98	104	143	145	200	200
20	**3**	YK99F4	150	152	172	178	94	98	98	104	143	145	200	200
21	**1**	YK99F5	150	150	178	184	98	98	98	104	147	147	204	204
21	**2**	YK99F5	150	150	178	184	94	98	98	106	143	147	204	204
21	**3**	YK99F5	150	150	178	184	98	98	98	104	x	x	x	x
22	**1**	YK99M5	152	158	174	174	94	100	100	104	x	x	196	204
22	**2**	YK99M5	152	158	174	174	94	100	104	104	x	x	196	204
22	**3**	YK99M5	152	158	174	174	94	94	100	102	x	x	196	204
23	**1**	YK99M3	x	x	x	x	x	x	96	100	x	x	196	196
23	**2**	YK99M3	x	x	x	x	x	x	x	x	x	x	x	x
23	**3**	YK99M3	x	x	x	x	x	x	x	x	x	x	x	x
24	**1**	YK00M1	x	x	x	x	x	x	x	x	x	x	x	x
24	**2**	YK00M1	x	x	x	x	x	x	x	x	x	x	x	x
24	**3**	YK00M1	x	x	176	176	x	x	x	x	x	x	x	x
25	**1**	YK00M2	150	150	x	x	94	94	100	104	x	x	x	x
25	**2**	YK00M2	x	x	x	x	x	x	102	104	x	x	x	x
25	**3**	YK00M2	x	x	x	x	x	x	x	x	x	x	x	x
26	**1**	YK00M3	x	x	176	178	x	x	98	102	145	145	x	x
26	**2**	YK00M3	x	x	x	x	x	x	98	104	145	145	x	x
26	**3**	YK00M3	x	x	x	x	x	x	x	x	x	x	x	x

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

27	**1**	YK00M7	154	154	x	x	x	x	100	102	x	x	x	x
27	**2**	YK00M7	x	x	x	x	x	x	100	102	x	x	x	x
27	**3**	YK00M7	x	x	x	x	x	x	x	x	x	x	x	x
28	**1**	YK00F1	x	x	172	172	98	98	102	102	x	x	x	x
28	**2**	YK00F1	x	x	x	x	102	102	x	x	x	x	x	x
28	**3**	YK00F1	x	x	x	x	x	x	100	100	x	x	x	x
29	**1**	YK00F6	x	x	x	x	94	94	100	106	147	147	x	x
29	**2**	YK00F6	x	x	x	x	96	98	x	x	x	x	x	x
29	**3**	YK00F6	x	x	x	x	x	x	102	104	x	x	x	x
30	**1**	YK00F9	x	x	x	x	96	102	100	100	x	x	x	x
30	**2**	YK00F9	x	x	x	x	96	96	100	100	x	x	x	x
30	**3**	YK00F9	x	x	x	x	x	x	100	102	x	x	x	x
31	**1**	YK00F11	150	150	x	x	106	106	x	x	x	x	x	x
31	**2**	YK00F11	x	x	170	170	x	x	96	102	x	x	x	x
31	**3**	YK00F11	x	x	x	x	x	x	x	x	x	x	x	x
32	**1**	YK00F13	152	152	x	x	96	96	x	x	x	x	x	x
32	**2**	YK00F13	x	x	x	x	x	x	x	x	x	x	x	x
32	**3**	YK00F13	x	x	x	x	x	x	x	x	x	x	x	x
33	**1**	YK00F16	x	x	x	x	x	x	x	x	x	x	x	x
33	**2**	YK00F16	x	x	x	x	x	x	94	100	x	x	x	x
33	**3**	YK00F16	x	x	x	x	x	x	x	x	x	x	x	x
34	**1**	YK00F19	152	152	178	178	x	x	x	x	x	x	x	x
34	**2**	YK00F19	152	152	x	x	x	x	100	102	x	x	x	x
34	**3**	YK00F19	x	x	x	x	x	x	100	106	x	x	x	x
35	**1**	YK00F20	x	x	x	x	96	98	102	102	x	x	200	200
35	**2**	YK00F20	x	x	x	x	x	x	102	106	x	x	200	200
35	**3**	YK00F20	x	x	178	178	x	x	98	98	x	x	200	200
36	**1**	AK99F4	150	152	178	180	94	100	104	104	x	x	200	200
36	**2**	AK99F4	150	152	178	180	x	x	100	104	x	x	200	200
36	**3**	AK99F4	150	152	x	x	x	x	104	104	x	x	200	200
37	**1**	AK99M1	150	150	178	180	102	104	104	106	x	x	200	200
37	**2**	AK99M1	150	150	178	180	102	104	104	106	x	x	200	200

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

37	**3**	AK99M1	150	150	178	180	102	104	104	106	x	x	200	200
38	**1**	AK99M26	x	x	x	x	98	104	x	x	x	x	196	200
38	**2**	AK99M26	150	150	174	174	x	x	100	104	x	x	200	200
38	**3**	AK99M26	154	154	170	176	x	x	104	104	x	x	200	200
39	**1**	AK00F5	x	x	x	x	x	x	98	100	x	x	x	x
39	**2**	AK00F5	x	x	170	172	x	x	x	x	x	x	x	x
39	**3**	AK00F5	x	x	174	174	x	x	100	100	x	x	x	x
40	**1**	AK00F6	x	x	172	172	100	100	96	104	x	x	x	x
40	**2**	AK00F6	x	x	x	x	104	104	102	102	x	x	x	x
40	**3**	AK00F6	x	x	x	x	x	x	x	x	x	x	x	x
41	**1**	AK00M4	x	x	174	178	104	104	94	102	x	x	x	x
41	**2**	AK00M4	x	x	x	x	x	x	98	98	x	x	x	x
41	**3**	AK00M4	x	x	172	172	x	x	104	104	x	x	x	x
42	**1**	BC99M5	150	152	174	176	94	94	98	104	141	145	196	200
42	**2**	BC99M5	150	152	174	176	94	94	98	104	141	145	196	200
42	**3**	BC99M5	150	152	174	176	94	94	98	104	141	145	196	200
43	**1**	BC00M13	x	x	170	170	98	98	98	98	x	x	x	x
43	**2**	BC00M13	x	x	170	172	x	x	98	108	x	x	x	x
43	**3**	BC00M13	x	x	x	x	x	x	x	x	x	x	x	x
44	**1**	BC00M12	x	x	x	x	x	x	x	x	x	x	x	x
44	**2**	BC00M12	x	x	176	176	x	x	102	102	x	x	x	x
44	**3**	BC00M12	x	x	170	170	x	x	102	104	x	x	x	x
45	**1**	BC00F10	x	x	172	172	94	96	102	102	x	x	x	x
45	**2**	BC00F10	x	x	176	176	x	x	102	102	x	x	x	x
45	**3**	BC00F10	x	x	x	x	x	x	x	x	x	x	x	x
46	**1**	YK00M1	152	152	178	178	x	x	100	104	145	145	x	x
46	**2**	YK00M1	x	x	x	x	x	x	100	106	145	145	x	x
46	**3**	YK00M1	x	x	x	x	x	x	100	102	x	x	x	x
47	**1**	YK99M6	154	154	172	174	94	98	100	106	145	149	200	200
47	**2**	YK99M6	146	154	172	174	94	98	102	104	149	149	200	200
47	**3**	YK99M6	150	154	172	174	94	98	100	102	145	149	200	200
48	**1**	YK00M4	148	150	x	x	94	94	x	x	x	x	x	x

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

48	**2**	YK00M4	x	x	x	x	x	x	x	x	x	x	x	x
48	**3**	YK00M4	x	x	x	x	x	x	100	102	x	x	x	x
49	**1**	YK00F9	152	152	x	x	x	x	100	104	x	x	x	x
49	**2**	YK00F9	x	x	x	x	x	x	100	104	x	x	x	x
49	**3**	YK00F9	x	x	x	x	x	x	100	104	x	x	x	x
50	**1**	AK00F2	x	x	x	x	94	94	x	x	x	x	x	x
50	**2**	AK00F2	x	x	x	x	x	x	102	104	x	x	x	x
50	**3**	AK00F2	x	x	x	x	x	x	104	104	x	x	x	x
51	**1**	AK99M19	152	156	174	180	96	102	102	106	135	137	200	204
51	**2**	AK99M19	152	156	174	180	96	102	102	106	135	137	200	204
51	**3**	AK99M19	152	156	174	180	96	102	102	106	135	137	200	204
52	**1**	AK99M26	150	154	174	180	98	104	98	104	137	141	196	200
52	**2**	AK99M26	150	154	174	180	98	104	98	104	137	141	196	200
52	**3**	AK99M26	150	154	174	180	98	104	98	104	137	141	196	200
53	**1**	AK99F10	150	150	174	178	100	104	104	106	135	143	200	200
53	**2**	AK99F10	150	150	174	178	100	104	104	106	135	143	200	200
53	**3**	AK99F10	150	150	174	178	100	104	104	106	135	143	200	200
54	**1**	AK99F13	146	152	172	174	94	100	102	104	143	145	196	200
54	**2**	AK99F13	146	152	172	174	94	100	102	104	143	145	196	200
54	**3**	AK99F13	146	152	172	174	94	100	102	104	143	145	196	200
55	**1**	AK99F27	146	150	172	176	94	94	102	104	x	x	200	200
55	**2**	AK99F27	146	150	172	176	94	94	102	104	x	x	200	200
55	**3**	AK99F27	146	150	172	176	94	94	102	104	x	x	200	200
56	**1**	AK00F1	x	x	174	178	x	x	104	104	x	x	198	198
56	**2**	AK00F1	x	x	x	x	x	x	98	100	x	x	x	x
56	**3**	AK00F1	x	x	174	176	x	x	98	98	x	x	x	x
57	**1**	AK00M3	150	150	x	x	96	102	102	102	141	141	x	x
57	**2**	AK00M3	x	x	x	x	x	x	102	102	x	x	x	x
57	**3**	AK00M3	x	x	x	x	x	x	102	104	x	x	x	x
58	**1**	BC99F6	152	152	172	176	96	96	x	x	137	145	200	200
58	**2**	BC99F6	150	152	172	176	96	96	x	x	137	145	196	198
58	**3**	BC99F6	152	152	172	176	96	96	x	x	137	137	200	200

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

59	**1**	BC00M11	150	150	x	x	102	102	106	106	x	x	x	x
59	**2**	BC00M11	x	x	x	x	x	x	106	106	x	x	x	x
59	**3**	BC00M11	x	x	x	x	x	x	106	106	x	x	x	x
60	**1**	BC00M9	x	x	x	x	x	x	x	x	x	x	x	x
60	**2**	BC00M9	x	x	x	x	x	x	104	104	x	x	x	x
60	**3**	BC00M9	x	x	x	x	x	x	x	x	x	x	x	x
61	**1**	AK99F4	152	152	178	180	100	100	103	103	x	x	200	200
61	**2**	AK99F4	152	154	172	178	94	100	102	104	x	x	x	x
61	**3**	AK99F4	152	152	172	178	94	100	104	110	x	x	x	x
62	**1**	AK99F17	150	150	172	180	98	100	98	104	139	143	200	200
62	**2**	AK99F17	150	150	172	180	98	100	98	104	139	143	200	200
62	**3**	AK99F17	150	150	172	180	98	100	98	104	139	143	200	200
63	**1**	AK99FX	154	154	176	180	94	98	104	104	147	147	196	200
63	**2**	AK99FX	150	150	176	180	94	98	104	104	141	147	196	200
63	**3**	AK99FX	150	154	176	180	94	98	104	104	141	147	196	200
64	**1**	AK99M6	154	154	172	180	98	100	98	106	141	141	x	x
64	**2**	AK99M6	150	154	x	x	98	100	x	x	139	139	x	x
64	**3**	AK99M6	x	x	166	180	98	102	x	x	139	143	x	x
65	**1**	AK99M9	152	152	172	178	94	98	x	x	143	143	204	204
65	**2**	AK99M9	152	156	172	172	94	98	102	104	x	x	196	204
65	**3**	AK99M9	150	154	172	178	94	98	102	104	x	x	204	204
66	**1**	AK00F3	154	154	x	x	98	98	100	100	139	139	204	204
66	**2**	AK00F3	152	152	x	x	96	96	100	104	x	x	x	x
66	**3**	AK00F3	x	x	x	x	x	x	x	x	141	141	x	x
67	**1**	AK00F3	x	x	168	170	x	x	100	102	x	x	x	x
67	**2**	AK00F3	x	x	x	x	x	x	102	106	x	x	x	x
67	**3**	AK00F3	x	x	x	x	x	x	104	104	x	x	x	x
68	**1**	AK00F5	x	x	172	172	x	x	96	100	x	x	x	x
68	**2**	AK00F5	x	x	x	x	x	x	104	104	x	x	x	x
68	**3**	AK00F5	x	x	x	x	x	x	100	104	x	x	x	x
69	**1**	AK00M4	x	x	x	x	x	x	104	106	x	x	x	x
69	**2**	AK00M4	x	x	x	x	x	x	102	104	x	x	x	x

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

69	**3**	AK00M4	x	x	x	x	x	x	104	104	x	x	x	x
70	**1**	BC00M2	150	150	170	172	x	x	98	102	x	x	200	200
70	**2**	BC00M2	x	x	x	x	x	x	102	102	x	x	x	x
70	**3**	BC00M2	x	x	x	x	x	x	104	104	x	x	x	x
71	**1**	BC00M13	150	152	170	172	100	100	106	106	x	x	x	x
71	**2**	BC00M13	x	x	x	x	x	x	104	104	x	x	x	x
71	**3**	BC00M13	x	x	172	172	x	x	x	x	x	x	x	x
72	**1**	BC00M2	x	x	x	x	x	x	98	98	x	x	204	204
72	**2**	BC00M2	x	x	x	x	x	x	102	104	x	x	196	196
72	**3**	BC00M2	x	x	x	x	x	x	100	102	x	x	x	x
73	**1**	BC00F8	152	152	172	172	98	98	100	104	147	147	196	196
73	**2**	BC00F8	x	x	166	168	x	x	104	106	147	147	196	204
73	**3**	BC00F8	x	x	x	x	x	x	106	106	x	x	x	x
74	**1**	BC00F18	x	x	x	x	x	x	104	104	x	x	x	x
74	**2**	BC00F18	x	x	168	170	x	x	104	104	x	x	x	x
74	**3**	BC00F18	x	x	168	170	x	x	95	95	x	x	x	x
75	**1**	YK99M2	156	156	x	x	94	96	100	104	145	145	196	196
75	**2**	YK99M2	x	x	178	178	x	x	102	106	x	x	196	196
75	**3**	YK99M2	x	x	170	170	x	x	100	104	x	x	x	x
76	**1**	YK00M1	150	150	172	172	96	96	102	104	x	x	x	x
76	**2**	YK00M1	x	x	x	x	x	x	102	104	x	x	x	x
76	**3**	YK00M1	x	x	x	x	x	x	102	104	x	x	x	x
77	**1**	YK99M5	148	152	174	174	100	100	102	108	x	x	196	196
77	**2**	YK99M5	158	158	174	174	94	100	102	104	145	145	204	204
77	**3**	YK99M5	x	x	174	174	x	x	x	x	x	x	x	x
78	**1**	YK99M6	150	156	172	172	94	98	104	104	149	149	200	200
78	**2**	YK99M6	x	x	172	172	98	98	92	104	x	x	198	200
78	**3**	YK99M6	x	x	172	174	x	x	100	102	x	x	x	x
79	**1**	YK00M3	152	152	x	x	102	102	100	106	x	x	x	x
79	**2**	YK00M3	x	x	x	x	x	x	102	106	x	x	x	x
79	**3**	YK00M3	x	x	178	178	x	x	104	104	x	x	x	x
80	**1**	YK00M7	x	x	x	x	98	98	100	102	x	x	x	x

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

80	**2**	YK00M7	x	x	176	176	96	98	104	106	x	x	x	x
80	**3**	YK00M7	x	x	x	x	96	100	104	104	x	x	x	x
81	**1**	YK00F9	x	x	x	x	x	x	100	104	x	x	x	x
81	**2**	YK00F9	x	x	172	172	x	x	104	104	x	x	x	x
81	**3**	YK00F9	x	x	x	x	x	x	102	106	x	x	x	x
82	**1**	YK00F10	x	x	x	x	96	98	98	98	x	x	x	x
82	**2**	YK00F10	x	x	x	x	94	96	100	108	x	x	x	x
82	**3**	YK00F10	x	x	172	172	x	x	104	104	x	x	x	x
83	**1**	YK00F11	x	x	x	x	98	98	x	x	x	x	x	x
83	**2**	YK00F11	x	x	176	176	104	104	102	106	x	x	x	x
83	**3**	YK00F11	x	x	166	170	x	x	104	104	x	x	x	x
84	**1**	YK00F16	x	x	x	x	96	96	98	104	x	x	x	x
84	**2**	YK00F16	x	x	x	x	94	94	102	104	x	x	x	x
84	**3**	YK00F16	x	x	x	x	x	x	100	102	x	x	x	x
85	**1**	YK00F20	148	148	x	x	x	x	98	98	x	x	x	x
85	**2**	YK00F20	x	x	x	x	x	x	102	104	x	x	x	x
85	**3**	YK00F20	x	x	x	x	x	x	100	104	x	x	x	x
86	**1**	AK99F17	150	158	172	180	98	100	98	104	139	143	200	200
86	**2**	AK99F17	150	150	172	180	98	100	x	x	139	143	200	200
86	**3**	AK99F17	150	150	172	180	98	100	x	x	139	143	200	200
87	**1**	AK99M6	150	154	172	180	98	100	x	x	x	x	200	204
87	**2**	AK99M6	150	154	172	180	98	100	98	106	x	x	200	204
87	**3**	AK99M6	150	154	172	180	98	100	98	106	x	x	200	204
88	**1**	AK99FX	146	150	176	180	94	98	104	104	141	147	196	200
88	**2**	AK99FX	146	150	176	180	94	98	104	104	141	147	196	200
88	**3**	AK99FX	146	150	176	180	94	98	104	104	x	x	196	200
89	**1**	AK99M26	150	154	174	180	98	104	98	104	137	141	196	200
89	**2**	AK99M26	150	154	174	180	98	104	x	x	137	141	196	200
89	**3**	AK99M26	150	154	174	180	98	104	98	104	137	141	196	200
90	**1**	BC99M3	152	154	x	x	98	104	102	104	141	141	200	200
90	**2**	BC99M3	150	150	172	174	100	104	104	104	x	x	x	x
90	**3**	BC99M3	152	152	172	172	x	x	102	104	x	x	x	x

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

91	**1**	BC99F15	152	152	172	172	96	98	104	104	137	143	196	196
91	**2**	BC99F15	152	152	172	172	96	98	104	104	137	143	196	196
91	**3**	BC99F15	152	152	172	172	96	98	104	104	137	143	196	196
92	**1**	BC00M2	x	x	x	x	x	x	x	x	x	x	x	x
92	**2**	BC00M2	x	x	x	x	x	x	x	x	x	x	x	x
92	**3**	BC00M2	x	x	x	x	x	x	x	x	x	x	x	x
93	**1**	BC00F8	152	152	172	176	x	x	102	104	x	x	200	200
93	**2**	BC00F8	x	x	176	176	x	x	98	104	x	x	x	x
93	**3**	BC00F8	x	x	172	176	x	x	104	106	x	x	x	x
94	**1**	YK99M2	156	156	176	176	100	100	x	x	141	141	200	200
94	**2**	YK99M2	148	164	176	176	98	98	98	104	x	x	200	200
94	**3**	YK99M2	150	150	176	176	x	x	104	104	x	x	200	200
95	**1**	YK00M3	x	x	x	x	x	x	102	102	141	141	198	198
95	**2**	YK00M3	x	x	176	176	x	x	102	102	143	143	194	198
95	**3**	YK00M3	x	x	x	x	x	x	102	102	x	x	x	x
96	**1**	YK00F10	146	152	x	x	98	98	x	x	x	x	196	200
96	**2**	YK00F10	x	x	x	x	93	93	x	x	x	x	196	200
96	**3**	YK00F10	x	x	x	x	x	x	x	x	x	x	x	x
97	**1**	YK00F11	162	162	x	x	x	x	x	x	143	143	198	198
97	**2**	YK00F11	x	x	x	x	x	x	100	106	x	x	x	x
97	**3**	YK00F11	x	x	x	x	x	x	100	104	x	x	x	x
98	**1**	YK00F20	158	158	178	178	x	x	104	104	141	141	196	196
98	**2**	YK00F20	x	x	x	x	x	x	102	102	x	x	202	202
98	**3**	YK00F20	x	x	x	x	x	x	102	102	x	x	x	x
99	**1**	YK00F3	x	x	172	178	102	102	x	x	x	x	198	198
99	**2**	YK00F3	x	x	x	x	x	x	98	98	x	x	196	200
99	**3**	YK00F3	x	x	x	x	x	x	100	102	x	x	194	194
100	**1**	YK00F15	150	150	178	182	x	x	102	102	x	x	196	200
100	**2**	YK00F15	x	x	x	x	x	x	96	98	x	x	192	194
100	**3**	YK00F15	x	x	x	x	x	x	98	104	x	x	x	x

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

TABLE C2. Microsatellite genotypes from plucked hair samples of lynx released in 2003. Each sample was genotyped 3 times after DNA extraction. An 'x' indicates no alleles amplified at that locus.

Sample	Run	Lynx ID	Lc106	Lc106	Lc109	Lc109	Lc110	Lc110	Lc111	Lc111	Lc118	Lc118	Lc120	Lc120
1	1	QU03F1	98	108	x	x	x	x	x	x	x	x	x	x
1	2	QU03F1	x	x	174	176	98	104	144	148	x	x	x	x
1	3	QU03F1	x	x	x	x	x	x	148	150	137	137	196	200
2	1	QU03F2	98	110	172	172	98	100	x	x	143	143	196	196
2	2	QU03F2	98	110	172	172	98	100	x	x	143	143	196	196
2	3	QU03F2	98	110	172	172	98	100	x	x	143	143	196	196
3	1	QU03F3	x	x	178	180	94	102	150	150	x	x	196	204
3	2	QU03F3	98	104	178	180	94	102	150	150	x	x	x	x
3	3	QU03F3	98	104	178	180	94	102	150	150	139	141	196	200
4	1	QU03F4	106	106	172	180	x	x	x	x	x	x	202	204
4	2	QU03F4	106	106	172	180	x	x	x	x	143	143	200	204
4	3	QU03F4	106	106	172	180	x	x	x	x	143	143	200	202
5	1	QU03F5	104	110	172	178	100	102	150	152	x	x	x	x
5	2	QU03F5	x	x	172	178	100	104	150	150	139	143	x	x
5	3	QU03F5	x	x	x	x	98	102	150	152	x	x	196	200
6	1	QU03F6	104	106	172	180	94	94	150	152	141	147	196	200
6	2	QU03F6	104	106	172	180	94	94	150	152	141	147	196	200
6	3	QU03F6	104	106	172	180	94	94	150	152	141	147	x	x
7	1	QU03F7	104	104	174	180	94	100	x	x	137	139	200	204
7	2	QU03F7	104	104	174	180	94	100	150	150	137	139	200	204
7	3	QU03F7	104	104	174	180	94	100	150	150	137	139	200	204
8	1	QU03M1	x	x	176	180	96	100	150	152	143	143	200	204
8	2	QU03M1	x	x	176	178	94	96	150	152	143	143	204	204
8	3	QU03M1	104	108	174	180	x	x	150	152	141	143	200	204
9	1	QU03M2	100	104	x	x	x	x	x	x	x	x	196	196
9	2	QU03M2	x	x	x	x	x	x	146	150	141	141	196	204

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

9	**3**	QU03M2	x	x	176	178	x	x	150	152	x	x	194	196
10	**1**	QU03M3	102	108	178	180	94	98	146	154	139	141	200	204
10	**2**	QU03M3	102	108	178	180	94	98	x	x	139	141	196	200
10	**3**	QU03M3	102	108	178	180	94	100	150	154	139	141	200	204
11	**1**	QU03M4	102	108	172	172	x	x	x	x	141	141	x	x
11	**2**	QU03M4	100	100	x	x	98	102	x	x	145	145	196	204
11	**3**	QU03M4	102	102	174	174	94	98	148	150	x	x	x	x
12	**1**	QU03M5	102	104	174	180	94	104	150	154	143	143	200	200
12	**2**	QU03M5	102	104	174	180	94	104	150	150	141	143	200	202
12	**3**	QU03M5	102	102	174	180	94	100	150	154	143	143	x	x
13	**1**	MB03M01	98	106	172	174	98	100	150	150	x	x	x	x
13	**2**	MB03M01	98	106	x	x	x	x	x	x	x	x	196	200
13	**3**	MB03M01	x	x	172	174	x	x	148	148	x	x	204	204
14	**1**	BC03F01	x	x	174	180	x	x	x	x	x	x	196	196
14	**2**	BC03F01	104	104	x	x	100	100	146	150	139	141	x	x
14	**3**	BC03F01	104	104	174	174	x	x	x	x	139	141	196	200
15	**1**	BC03F02	102	104	172	174	98	106	150	150	143	145	200	200
15	**2**	BC03F02	102	104	172	174	98	106	150	150	143	145	200	200
15	**3**	BC03F02	102	104	172	174	98	106	150	150	143	145	200	200
16	**1**	BC03F03	102	104	172	178	104	106	x	x	143	147	200	200
16	**2**	BC03F03	102	104	172	178	104	106	150	150	141	147	200	200
16	**3**	BC03F03	102	104	x	x	98	104	x	x	x	x	200	200
17	**1**	BC03F04	104	104	x	x	96	96	150	150	x	x	x	x
17	**2**	BC03F04	x	x	x	x	100	100	150	150	x	x	200	200
17	**3**	BC03F04	102	102	x	x	98	98	x	x	141	143	196	200
18	**1**	BC03F05	100	102	174	174	98	100	152	154	141	143	196	200
18	**2**	BC03F05	100	102	174	174	98	100	152	154	141	143	196	200
18	**3**	BC03F05	100	102	x	x	98	100	150	152	141	143	196	200
19	**1**	BC03F06	98	100	176	180	94	94	x	x	139	143	200	204
19	**2**	BC03F06	98	100	176	180	94	94	x	x	139	143	200	204
19	**3**	BC03F06	x	x	176	180	x	x	x	x	139	143	200	204
20	**1**	BC03F07	102	104	174	176	x	x	152	152	x	x	x	x

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

20	**2**	BC03F07	104	104	x	x	94	96	150	150	137	143	x	x
20	**3**	BC03F07	x	x	x	x	x	x	x	x	x	x	x	x
21	**1**	BC03F08	94	98	180	180	104	104	150	150	139	143	200	202
21	**2**	BC03F08	x	x	x	x	104	104	150	152	139	143	200	200
21	**3**	BC03F08	94	98	180	180	104	104	150	152	139	143	x	x
22	**1**	BC03F09	102	104	174	182	94	98	x	x	141	143	200	200
22	**2**	BC03F09	102	104	x	x	94	98	x	x	141	143	x	x
22	**3**	BC03F09	102	104	174	182	94	98	x	x	141	143	200	200
23	**1**	BC03F10	102	104	176	180	94	100	150	152	143	145	196	196
23	**2**	BC03F10	102	104	x	x	94	100	150	152	143	145	196	196
23	**3**	BC03F10	102	104	x	x	x	x	150	152	143	145	196	204
24	**1**	BC03M01	x	x	x	x	94	96	150	150	141	145	x	x
24	**2**	BC03M01	x	x	x	x	x	x	150	150	143	145	x	x
24	**3**	BC03M01	x	x	176	176	102	102	150	150	143	145	200	202
25	**1**	BC03M02	x	x	x	x	94	100	150	150	141	145	x	x
25	**2**	BC03M02	98	102	172	172	x	x	148	152	x	x	x	x
25	**3**	BC03M02	x	x	x	x	x	x	150	152	x	x	x	x
26	**1**	BC03M03	x	x	x	x	x	x	150	152	137	139	x	x
26	**2**	BC03M03	96	102	172	172	x	x	x	x	139	141	x	x
26	**3**	BC03M03	100	102	x	x	x	x	148	150	143	143	x	x
27	**1**	BC03M04	104	110	172	180	x	x	x	x	141	143	198	202
27	**2**	BC03M04	x	x	172	172	x	x	148	150	137	145	202	202
27	**3**	BC03M04	x	x	172	172	x	x	150	150	141	143	200	202
28	**1**	BC03M05	100	100	174	176	x	x	x	x	x	x	x	x
28	**2**	BC03M05	100	106	176	178	x	x	x	x	137	141	x	x
28	**3**	BC03M05	102	110	174	176	x	x	148	150	x	x	x	x
29	**1**	BC03M06	x	x	172	176	96	98	150	152	137	141	196	200
29	**2**	BC03M06	102	106	172	176	96	98	150	150	137	141	196	200
29	**3**	BC03M06	106	106	172	176	96	98	150	152	137	141	196	200
30	**1**	BC03M07	102	106	178	180	98	98	150	154	x	x	196	200
30	**2**	BC03M07	102	106	178	180	98	98	150	154	145	145	196	200
30	**3**	BC03M07	102	106	178	180	98	98	150	154	145	145	196	200

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

31	**1**	BC03M08	102	106	174	180	x	x	152	152	143	145	196	204
31	**2**	BC03M08	102	106	174	180	100	104	x	x	x	x	196	204
31	**3**	BC03M08	102	106	174	180	100	104	152	152	141	147	196	200
32	**1**	BC03M09	102	106	176	180	x	x	152	152	141	143	196	202
32	**2**	BC03M09	96	100	172	176	x	x	x	x	143	143	198	200
32	**3**	BC03M09	x	x	x	x	100	100	x	x	143	143	196	200
33	**1**	BC03M10	102	104	174	180	104	104	x	x	143	143	196	200
33	**2**	BC03M10	102	104	174	180	x	x	x	x	141	141	196	200
33	**3**	BC03M10	102	104	174	176	x	x	x	x	137	141	196	200

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.