A Simple Population Estimate Based on Simulation for Capture-Recapture and Capture-Resight Data

Steven Minta, Marc Mangel

A SIMPLE POPULATION ESTIMATE BASED ON SIMULATION FOR CAPTURE–RECAPTURE AND CAPTURE–RESIGHT DATA[1]

Steven Minta[2]

Department of Wildlife and Fisheries Biology, University of California, Davis, California 95616 USA

AND

Marc Mangel

Department of Zoology, University of California, Davis, California 95616 USA

Abstract.   The use of capture–resight data for population estimation has seldom been exploited. It offers potential flexibility and advantages to the design of biological investigations in which a population estimate is required. Presently, the Petersen model is the only method for estimating closed populations using capture–resight data. A simple Monte Carlo simulation method can lead to a full probability distribution for the population. From this probability distribution, one can compute maximum likelihood estimates and a likelihood interval on the population. The shape and asymmetry of the distribution and width of likelihood intervals are determined by sampling heterogeneity and sample size. The method is simple and can be used by anyone with access to a microcomputer. Since it is data-intensive, estimates based on small data sets (including capture–recapture) with few animals can be quickly calculated. The method is especially applicable to species and habitats in which capture–resight, radiotelemetry, or other tracking data can be obtained and to situations in which nonrandom catchability or sightability is likely after the initial capture. The technique successfully estimated populations of badgers, bison, and crested porcupines. We compare observed with theoretical sighting distributions to examine the effects of model and sampling biases.

Key words: animal tracking; capture—recapture; capture–resight; experimental design; mark–recapture; mark–resight; Monte Carlo simulation; population methods; population-size estimation; radiotelemetry; sample bias.

INTRODUCTION

In contrast to theoretical and statistical progress, there has been limited progress estimating fundamental population characteristics in the field. The increasingly efficient methods developed by biometricians in the last two decades have not markedly improved design guidelines or decreased research cost and effort (Eberhardt 1978, Arnason and Baniuk 1980, Pank 1981, Pollock 1982, Skalski 1985, Seber 1986). However, we do know more completely when and how an estimate is likely to be reasonable. We are more aware of how critical assumptions can be, and that we must always assess deviations from them (Cormack 1972, Carothers 1973, Roff 1973*a, b,* Burnham and Anderson 1979, Buckland 1982, Pollock and Raveling 1982). This has become the major analytical task, and coping with these methodological assumptions has so increased statistical and procedural complexity that large software systems are almost mandatory for an acceptably thorough analysis (Nichols et al. 1981, Brownie and Pollock

1985). Unfortunately, these highly developed systems require large data sets to justify their use and are, admittedly, not for the one-shot or occasional user (Arnason and Baniuk 1978, 1980, Pollock 1981). There are potentially many occasional users, since most researchers have diverse objectives and costs, of which population estimation is but one (Skalski et al. 1984, Greenwood et al. 1985). The occasional user may not understand key points fully and therefore be unable to use the methods effectively (Nichols et al. 1981). There is a need for: (1) simpler and less expensive population estimators for the wide range of applications and users, and (2) alternative estimators for accommodating the variety of experimental conditions and deviations from assumptions.

In this paper we introduce modifications of capture–recapture (mark–recapture), especially capture–resight (mark–resight) and radiotelemetry, which can be applied to a variety of organisms. We then introduce a Monte Carlo method for capture–resighting and show how this simulation method can lead to a maximum likelihood estimate and confidence levels on the population. All capture–recapture methods require independence of initial marking and subsequent recapture/ resighting probabilities. However, variance estimators

based on a Petersen model are biased when used with aggregated recapture or resighting data from animals with heterogeneous recapture or sighting probabilities. By using all of the information contained in individual capture probabilities, the simulation approach yields (1) variance estimates that should be nearly unbiased, and (2) asymmetrical confidence intervals that in practice are likely to perform better than normal approximations.

CAPTURE–RECAPTURE

Some type of capture–recapture has been applied to nearly all forms of animal life but its usefulness may be much overrated (Roff 1973a, Carothers 1979, White et al. 1982). The method never purported to be universally applicable, but in the absence of competing methods biologists use what is available and familiar. Unfortunately, the capturing, marking, and recapturing of animals is a highly nonrandom process and can severely disrupt a population. Although catchability is the major consideration of capture–recapture models with which we can partially contend, it is also the source of greatest confounding variation (Otis et al. 1978, Carothers 1979, Mares et al. 1981, Pollock 1981, Buckland 1982, Seber 1982, Nichols and Pollock 1983). There is a long list of sources of variable catchability, and the resulting matrix of possible interactions is bewildering. Differential catchability has plagued researchers who must use capture–recapture estimation. The annotated bibliography of Tepper (1967) and reviews by Roff (1973b), Smith et al. (1975), Tanaka (1980), and Seber (1982) cover many sources of variation. Most of the progress in capture–recapture experimental designs and statistical methods has evolved to overcome the problems associated with nonrandom catchability. While precision and realism have increased, so has complexity, both theoretical and practical. However, less sophisticated but still robust methods may actually perform better in application (i.e., Jolly–Seber vs. Lincoln–Petersen estimates: Blower et al. 1981, Mares et al. 1981, Jolly 1982, Seber 1982, Montgomery 1985).

CAPTURE–RESIGHT

Another approach to more efficient population estimation is to accommodate violations of the assumption of equal catchability. Numerous assumptions can be relaxed or avoided if an animal's captures are independent, for instance, by different modes of capture. Most biometricians strongly recommend the use of more than one capture method (Burnham and Overton 1979, Nichols et al. 1981) and Arnason and Baniuk (1978) suggest it should be standard practice. Visually "recapturing" animals is such an alternative capture mode. If an animal can be "recaptured" by observation, that is, without actually being exposed to more than the initial physical capture and handling, then even more catchability assumptions are met (e.g., ho-

mogeneous behavioral response and capture probabilities).

The advantages of capture–resighting have long been known, and it has been applied to an array of organisms. Pollock (1981) considers capture–resight to be a design that has great potential, and its promise and potential utility have also been extolled by Arnason and Baniuk (1978, 1980), Eberhardt et al. (1979), Nichols et al. (1981), White et al. (1982), and Brownie and Robson (1983). For many species and circumstances, capture–resight is cheaper than capture–recapture. Overall field effort is reduced and after an intense initial capture/marking period, fewer and less trained personnel are needed to make reobservations (Arnason and Baniuk 1978, Pollock 1981). More data can be accumulated in less time and resightings can be made while achieving other research objectives. Overall, the method is less disruptive to the animals and their environment. In addition, since the general consensus is that two or more independent population estimates be attempted (Smith et al. 1975, Berthold 1976, Connolly 1981, Bergerud 1982, Seber 1982, Skalski and Robson 1982), capture–resight can serve as an alternative estimator. For many investigations, it will be the preferred estimator.

Most capture–resight estimates used have been the Lincoln–Petersen (Petersen) type. In those cases individual marks need not be discerned, only marked vs. unmarked. Multiple capture–resight calls for individual identification. Consequently, most of the animals and their marks have been highly observable (e.g., large size, diurnal, high exposure in habitat) and countable (e.g., temporally stationary, solitary or in small aggregations). If the goal of capture–resight is in mind, more highly distinguishable markers can often be planned (e.g., Patterson 1978, Hill and Clayton 1985). Certainly, animals must be likely to be sighted and counted at some period of their diel and seasonal life cycle. Capture–resight requires independence among captures and observations of animals. Animals need not be sighted for capture–resight, they need only be detected as individuals. For instance, low-profile terrestrial animals might be detected through radiotelemetry and tracking (see Monte Carlo Simulation; Badgers in Wyoming).

Radiotelemetry has been infrequently used to estimate populations. Seber (1986) considers it to be one of the few important technological advances that are relevant to population estimation. The potential advantage of having an individually marked portion of a population that (often) can be located and sighted has not been exploited in the application of rigorous population estimates. Since animals need only be detected for capture–resight estimates, and since radiotelemetry provides remote detection without interference, then as a type of capture–resight, radiotelemetry can be employed to "recapture" animals as part of an experimental design for recapture or resight. When an

animal is sighted during a sampling interval and independently of its marked/unmarked status, it is counted as either a "marked" or unmarked individual. Thus, after an animal is sighted it can be counted as a marked individual if it transmits a telemetry signal that can be used to identify it. Gauthier and Theberge (1985) calculated a Petersen estimate of group-living caribou with this method.

The application of capture–resight (and radiotracking) is constrained by a variety of assumptions and methods, and at least some computational complexity for multiple capture–resight data. Open-population models easily incorporate either the strategy of marking animals before sampling of recaptures or the strategy of continuously tagging all unmarked animals captured throughout the entire sampling period. Although it is not a statistical requirement for the closed model, continuous marking is the only design accommodated by many capture–recapture programs, including those of Otis et al. (1978). However, in a capture–resight design, unmarked animals that are visually "captured" are not likely to be marked but are still recorded as unmarked animals and used in analysis of population size. Consequently, the investigator is confined to a Petersen estimator for closed populations in which, following the initial capture and marking, a second independent sample of sightings is recorded.

MODEL AND SAMPLING BIASES

White et al. (1982) classified biases as model biases and small-sample biases. Model biases are the more serious of the two because important assumptions of the sampling system are not correct. Sample-size biases diminish as sample size increases; model (and sampling) biases do not go away. The critical assumption of equal catchability in capture–recapture models requires that there be population and sampling homogeneity. Equal catchability can be decomposed into three distinct assumptions: (1) all animals have the same probability of being caught in the first sample, (2) catching and tagging do not affect future catchability, and (3) the second sample is a simple random sample. The Petersen model is robust to departures from these assumptions, in large part because it works with the mean capture rate but without any measure of the variability. Seber (1970) shows that the estimate, $\hat{N}$, can still be used even when assumption 3 is not true and a systematic rather than a random second sample is taken, provided there is uniform mixing of marked and unmarked individuals and all animals are equally catchable in the second sample. If a different sampling method is used for each sample, as in capture–resight, then Seber (1970) concludes that when assumption 2 is true, assumptions 1 and 3 can be replaced by the assumption that the catchabilities in the two samples are independent. In other words, variation in catchability could exist for both samples without introducing bias if the sources of selectivity in the two samples

were independent (as in capture–resight); however, individual heterogeneity of sighting probabilities will cause the variance of the estimate to be underestimated (Seber 1970).

Theoretically, the Petersen model produces an unbiased estimator in that replications will converge on the population parameter, N, such that $E(\hat{N}) = N$. That is, random replications of a random sampling regime will produce estimates centrally distributed about N. White et al. (1982) note that bias is a conceptual quantity because we usually have only one set of data and can compute only one value of $\hat{N}$ from the data. Thus, bias relates strictly to the performance of an estimator. Yet in application, the researcher who must make a single population estimate is interested in the performance of that single estimate, and one way to gain an estimate closer to N is to adhere to assumptions. In addition, the estimated variance of $\hat{N}$ will be more reliable. Sightings should randomly or systematically represent the entire area of the population to allow for nonrandom spatial distribution and population mixing. This is especially true if $\hat{N}$ is to be converted to a density estimate which is sensitive to heterogeneities. The sampling design should strive to allow equal probability of sighting individuals. Nonetheless, if assumptions are unavoidably but knowingly violated, and the researcher has only one set of data, then the statistical concept of average performance of an estimator is less meaningful and useful than the practical performance provided by an estimator that can account for population and sampling heterogeneities. Cox (1958:85–89) and Hurlbert (1984) discuss the similar problem of randomization in single small experiments and the ad hoc approach of procedural adjustment. Hurlbert (1984) quotes Cox's (1958:88) succinct summary and it is worth requoting:

> . . . to adapt arrangements we suspect are bad, simply because things will be all right in the long run, is to force our behavior into the Procrustean bed of a mathematical theory. Our object is the design of individual experiments that work well: good long-run properties are concepts that help us in doing this, but the exact fulfillment of long-run mathematical conditions is not the ultimate aim.

When the characteristics of the animal, environment, or experimental design preclude a single sighting sample, multiple applications of the Petersen estimator are computed where the numbers of animals sighted during separate surveys provide the second sample. Each survey or distinct sample time produces a new Petersen estimate which is combined with the others. Bartmann et al. (1987) and G. C. White and R. A. Garrott (unpublished manuscript) develop the likelihood under similar assumptions to the Schnabel model (model M_t, Otis et al. 1978) and reiterate the assumptions for combining these estimators: the individual sighting surveys should be independent and each in-

dividual animal should have the same probability of being sighted as every other individual on a particular occasion. Each survey should have equal effort, although the effects of unequal effort have not been fully explored. Bartmann et al. (1987) found that the majority of estimates were smaller than true population values, suggesting a negative bias (see also Eisenberg 1972, Rice and Harder 1977, Mares et al. 1981). They caution users that heterogeneity may cause some bias in the estimates and conclude that for small populations, a large proportion (>45%) should be marked before reliable estimates and confidence intervals can be obtained. We reach similar conclusions from a general model (see Appendix) and we suggest that marking such a large proportion may not be practicable for many species.

The method we propose combines multiple sighting surveys when surveys are not necessarily independent and when surveys are not equal in area or effort. The second sample need not be made up of discrete surveys as may happen with resightings that cannot be conveniently grouped into sample times. In other words, sampling may be continuous over any period of time the population remains closed. However, random or systematic coverage of the entire area within that period is assumed. Individual identification of marked animals is required. More importantly, whereas the variance for the Petersen estimate is symmetrical and invariant to population and sampling heterogeneity, the Monte Carlo method produces likelihood intervals that allow for asymmetry in the sampling distribution of the estimator and that vary according to the magnitude of the sample size and model/sampling biases. There are two issues that must be addressed. Here we concentrate on resighting heterogeneity and how that information can be incorporated into population estimates. Both the standard Petersen method and our method assume that initial captures are independent of resights. To do otherwise requires an additional layer of assumptions that we prefer not to make and feel is difficult to justify.

MONTE CARLO SIMULATION

Populations can be efficiently estimated via a data-intensive Monte Carlo simulation. Maximum likelihood estimators with likelihood intervals can be derived from a simple program that can be run on a microcomputer. The general capture–recapture assumptions hold: (1) Animals are captured, marked, and randomly or systematically resighted. (2) There is geographic and demographic closure, no tag loss, and proper recording of observations. (3) The initial captures are a random sample of the population, and marked animals are identical to unmarked animals in detection or observability. The assumption of independence between capture probability of the initial marking vs. subsequent resightings is critical. Over the sampling period, the frequency of resightings of each marked

animal will be accumulated and all unmarked animals will be pooled into a category of "total sightings of unmarked animals." (When the sampling period has ended, the resighting frequencies can be simply interpreted to resighting probabilities.) If the individuals in the marked sample are representative of the sample of unmarked individuals, then we can partition the sightings of unmarked animals into frequencies similar to those frequencies known for marked individuals. The randomized process by which we accomplish this "partitioning" is through Monte Carlo simulation.

The appeal of this capture–resight design and simulation analysis lies in the complete extraction of information from the data structure which allows an accurate assessment of the variance of the estimate. Unequal visual "catchability" within or among samples is not a concern. For example, if there are marked individuals who appear "sight-happy," perhaps due to extreme mobility or dependence among sampling occasions, then we account for them and assume such individuals have an equal probability of occurring in the unmarked segment of the population. By simulating the distribution of all sighting frequencies, the data on all animals detected in the population are used in their entirety.

The details and a justification of the Monte Carlo simulation are provided in the Appendix. Here we will present the general ideas and illustrate them with three examples: badgers snowtracked in Wyoming, bison resighted on Santa Catalina Island, and crested porcupines resighted in the Negev Desert. We will describe the badger example in detail to illustrate clearly how the assumptions of both the single and combined Petersen estimates are violated.

An initial subset of N_k of animals is marked so that individuals can be identified. Sighting is then conducted over some period of time. Let S denote the total number of sightings. This number is composed of S_k sightings of marked animals and $S_u = S - S_k$ sightings generated by unmarked animals. The fundamental question answered by the Monte Carlo simulation is: how many different animals out of the $N_u = N - N_k$ animals were sighted in the S_u sightings?

To answer this question, one can think of the animals as "generating" sightings. The first set of inputs in the Monte Carlo simulation is then

$$P_i = \text{fraction of marked animals that}$$
$$\text{generated } i \text{ resightings.} \quad (1)$$

The P_i, $i = 0, 1, 2, \ldots, N_m$, where $N_m =$ maximum number of resightings, is thus empirically determined. The 0 category must be verified by telemetry or independent data collection. The 0 category refers to animals that are never seen, although they are marked and known to be present. This is not contradictory; verification of the 0 category is equivalent to establishing an individual's presence within the area of estimation, but not while using the sampling protocol

specified for population estimation. For example, an animal may be located regularly for other sampling purposes (say, by telemetry for evaluation of habitat use) and is always located within the area. We assume there will be a similar proportion of unmarked animals corresponding to each P_i category, including $i = 0$. Thus, verification of the 0 category amounts to testing the closure assumption; if closure exists, then all animals marked but not seen again fall in the 0 category.

From the set $\{P_i\}$ and the value S_u, each iteration of the Monte Carlo simulation generates a feasible number N_u of animals that could generate S_u sightings. We can think of $N_u = \sum_i N_{ui}$ where N_{ui} is the number of unmarked animals that generated i resightings. Our method implicitly computes N_{u0}, the number of unmarked animals that did not generate any resightings, from information on P_0. The possible range for N_u is $1 \le N_u \le S_u$, where the lower limit applies if all sightings are generated by one animal, and the upper limit applies if each unmarked animal generates exactly one sighting. Neither of these limits is very likely. The output from the Monte Carlo simulation is a frequency distribution for N_u. The outputs are

f_j = fraction of Monte Carlo iterations in
which $N_u = j$.   (2)

Thus the numbers $\{f_j\}$ are a frequency distribution for the unknown number of animals. From the $\{f_j\}$, one can compute maximum likelihood estimates (essentially the Petersen estimate) and $\approx$ 95% likelihood intervals (see Appendix). The following examples will illustrate how the method is used.

For each example we will compare the Monte Carlo maximum likelihood estimate and its 95% likelihood interval with the Petersen estimate and its 95% confidence interval. We use Bailey's binomial model for the Petersen estimate because animals are merely observed (repeatedly) and not actually recaptured; therefore, sampling is with replacement (Seber 1982:61):

$$N^P = \frac{N_k(S + 1)}{(S_k + 1)} . \qquad (3)$$

$$V\hat{a}r(N^P) = \frac{N_k^2(S + 1)(S_u)}{(S_k + 1)^2(S_k + 2)} . \qquad (4)$$

$$95\% \text{ CI} = N^P \pm 1.96 V\hat{a}r(N^P)^{1/2} . \qquad (5)$$

The Petersen estimate uses the aggregated sightings; thus the variance is unaffected by the distribution of sightings among marked animals. However, if there is extreme heterogeneity (inequality) among animals in their sightability, then we would have less confidence in our estimate and would expect greater variance around the point estimate. Similarly, if there was extreme homogeneity (equality) among animals in their sightability, then our confidence in the estimate would increase and the variance should decrease accordingly.

The Petersen estimate, since it is based on means and aggregated data, will not respond at all to the level of homogeneity in the sighting distribution. The extreme case would be to mark 10 animals and later obtain 50 sightings. When the marked animals are each sighted about the same number of times, we trust the estimate. When a single individual is seen 50 times and no other animal is resighted, we know we have a model or sampling problem. Biological data fall between these hypothetical extremes. Because the Monte Carlo method accounts for differential sightability by using the exact distribution of sightings (not the summed sightings), we expect those differences to correspondingly increase or decrease likelihood intervals with little effect upon the maximum likelihood estimate.

To evaluate the effects of heterogeneous sightability, we will compare the Monte Carlo and Petersen results from the empirical distribution (i observed sightings) with Monte Carlo results from theoretical distributions derived from the parameters of the empirical distribution. We will generate the "equal" distribution of extreme homogeneity in sightings and the "unequal" case of extreme heterogeneity, as in the above example. For the equal case, sightings of marked animals are set as close to the mean of the empirical distribution, $S_B = \frac{S_k}{N_k}$, as integer additivity will allow. To generate the unequal distribution, sightings are split about the empirical mean so that generated sightings fall as close as possible to the minimum and maximum values for observed sightings. For comparison of randomly distributed sightings, binomial sightings are generated by the binomial distribution function with $n = S_k$ and $p = N_k^{-1}$. If $S_k \ge 70$, then the Poisson approximation is used with $\lambda = S_B$ because p is small, and as n becomes large, the binomial rapidly converges to the Poisson. To allow the reader to compare the $\{P_i\}$, we calculate a standard χ^2 goodness-of-fit test for some paired distributions. When sample sizes of frequency classes were less than unity, they were grouped. Because of small frequencies, goodness-of-fit results and inferences should be interpreted as indicating trend only.

Badgers in Wyoming

In fall of 1984, North American badgers (*Taxidea taxus*) were snowtracked in a 15 km² area on the National Elk Refuge, Jackson, Wyoming. The size and shape of the target area were dictated by topographic and plant community features that created a relatively isolated area of high badger density. Fifteen of the badgers were radiotagged and known to be occupying or overlapping the area. During the 2-mo tracking period there was no death or emigration of radiotagged individuals, and radiotagged badgers outside the target area did not immigrate. One badger emigrated near the end of the sampling period. During daylight and under suitable weather conditions, the target area was searched

TABLE 1. Number of resightings of marked badgers. Identifiers refer to numbered individuals within age and sex classes: A = adult, J = juvenile, F = female, M = male.

Badger identifier	Number of resightings
AF1	11
JM1	9
JF1	7
JF2	7
AF2	6
AF3	6
AF4	5
AF5	5
AM1	5
AM2	3
AM3	2
JM2	1
AM4	1
AF6	0
AM5	0

for badger snowtracks. A total of 107 tracks could be followed to a terminal hole, where the badger would be inactive in an underground burrow. All telemetry frequencies were then scanned to determine whether the badger was "marked" or "unmarked." Radiotelemetry revealed that 68 of the tracks were generated by marked badgers. The actual distribution of resightings is shown in Table 1.

Why is the single or combined Petersen not adequate for this data set? Asynchronous and synchronous temporal events occurred throughout autumn, affecting population and sampling heterogeneity. The whole or a part of the population exhibited entirely different movement in early autumn compared to late autumn, sometimes correlated and other times not (e.g., with abiotic variables). Spatial and temporal heterogeneity also occurred in experimental protocols due to practical circumstances. The area sampled varied from day to day and week to week because of weather and snowpack conditions. In effect, the continuous sampling of tracks throughout autumn appears arbitrary: any single sampling occurrence produced few tracks, and badgers may or may not have moved during the interval between sampling occurrences. Location and size of sampled areas are variable from sampling occurrence to sampling occurrence, as are the intervals among sampling occurrences. We strived for systematic sampling throughout autumn, but a number of samples were dependent and correlated in space and time. Furthermore, heterogeneities do not necessarily "average out" in any sense of the word, and population mixing is not random or uniform either as a population property or as an outcome of the sampling scheme. Badgers preferentially occupied home ranges within favored habitat in the census area. Thus, by combining population and sampling heterogeneities we can see how one badger may be "seen" disproportionately more often and at different times than another, sometimes in the same area, and at other times not.

The $\{P_i\}$ are shown in Table 2. Note the following

about the $\{P_i\}$: (1) If there are no marked animals observed to generate i tracks, then $P_i = 0$. (2) It is implicitly assumed that none of the unmarked animals generate more than a maximum of N_m tracks. This limitation is common in data-intensive estimation procedures (e.g., Efron 1982). The output from 10 000 iterations of the Monte Carlo simulation is shown in Fig. 1 for the empirical distribution (observed sightings). The maximum likelihood estimate (MLE) for N_u is $\hat{N}_u = 9$ badgers and the $\approx 95\%$ likelihood interval of the MLE is $[6 \leq N_u \leq 13]$ badgers.

Table 2 and Fig. 1 contain the $\{P_i\}$ and resulting simulations for the three sighting distributions with different levels of homogeneity. As expected, $\hat{N}_u = 9$ and $\hat{N} = 24$ for the equal and binomial distributions, except $\hat{N}_u = 8$ and $\hat{N} = 23$ for the unequal distribution. The variance of the empirical distribution ($\sigma^2 = 10.25$) is double that of the binomial generated distribution ($\sigma^2 = 4.23$) because a few badgers are seen more than expected and a few others are seen less than expected (goodness of fit: $\chi^2 = 14.00$, df $= 6$, $P = .031$). Therefore, the likelihood intervals increase: equal distribution $[23 \leq N \leq 25]$, binomial $[22 \leq N \leq 27]$, and unequal $[19 \leq N \leq 32]$.

For purposes of comparison, the uncorrected Petersen estimate of N_u is computed as follows. There were 68 resightings generated by 15 badgers, giving $\frac{68}{15} = 4.53 = S_B$ sightings per badger. Therefore, the Petersen estimate for the unmarked population is $N_u^P = \frac{S_u}{S_B} = \frac{39}{4.53} = 8.6$. The Petersen estimate for the population, using Bailey's binomial model, is $N^P = 23.5$ with a 95% CI of $[20.2 \leq N \leq 26.8]$. Compare this to the Monte Carlo results of $\hat{N} = 9 + 15 = 24$ with 95% likelihood interval of $[21 \leq N \leq 28]$. For the badger example,

TABLE 2. P_i for marked badgers of Table 1. The mean of all three distributions of sightings, S_B, is 4.53 sightings per marked badger. Distributions are explained in Table 3 footnote.

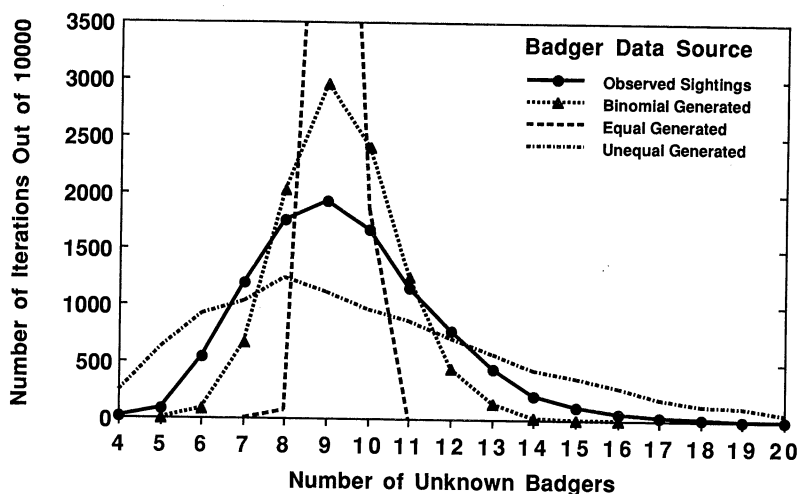| i	Observed P_i	Generated by badgers	Probability of resighting		
			Binomial P_i	Equal P_i	Unequal P_i
0	2/15	AF6, AM5	.01	0	7/15
1	2/15	AM4, JM2	.04	0	2/15
2	1/15	AM3	.11	0	0
3	1/15	AM2	.17	0	0
4	0	None	.19	7/15	0
5	3/15	AF4, AF5, AM1	.18	8/15	0
6	2/15	AF2, AF3	.13	0	0
7	2/15	JF1, JF2	.08	0	0
8	0	None	.05	0	0
9	1/15	JM1	.02	0	0
10	0	None	.01	0	0
11	1/15	AF1	.00	0	6/15

Fig. 1. Monte Carlo computation of the distribution of unknown badgers in Wyoming during the autumn of 1984, in 10 000 iterations. The empirical distribution of field data is referred to as observed sightings. The three other derived data sources were generated from the parameters of that observed distribution of sightings (see Table 3 footnote).

the estimates are close but the likelihood interval is slightly asymmetrical about the estimate.

Bison on Santa Catalina Island

In 1976 and 1977, mature bison cows (*Bison bison*) were ear-marked or radiocollared and resighted from vehicles and on foot (Lott and Minta 1983a, b). In 1976, there were 2003 resightings during a 26-d period following the summer breeding; 187 were generated by 16 marked individuals and 1816 were generated by unmarked individuals. In 1977, there were 2096 resightings in a 27-d period at the same time of year as in 1976; 139 were generated by 14 marked individuals and 1957 were generated by unmarked individuals. The P_i are shown in Tables 3 and 4.

Figs. 2 and 3 show the empirical and derived distributions of N_u in 1976 and 1977, generated by the Monte Carlo simulation. The bison on this island had been independently estimated by direct enumeration and found not to vary from ≈ 175 mature cows in these two years (D. Lott, *personal communication*). We will allow these censuses to represent the true population parameter of censused cows, $N = 175$, and therefore $N_u = 175 - 16 = 159$. Note that an island is an ideal source for testing this kind of method since if the observation period is short enough, closure of the population is virtually guaranteed.

The 1976 sighting data gave close estimates of the known number of censused cows, $N = 175$. $N^P = 170.6$ with CI of $[147.4 \leq N \leq 193.7]$. The Monte Carlo method produced $\hat{N} = 172$ for all choices of the $\{P_i\}$ with likelihood interval $[162 \leq N \leq 182]$ for the observed data, $[171 \leq N \leq 174]$ for equally distributed sightings, $[165 \leq N \leq 179]$ for Poisson distribution, and $[156 \leq N \leq 190]$ for unequal sightings. The variance of the observed sightings ($\sigma^2 = 23.84$) is double that of the Poisson ($\sigma^2 = 11.69$), again indicating het-

erogeneous sightability among marked animals in the tail of the empirical distribution, but with minor difference (goodness of fit: $\chi^2 = 11.36$, df $= 8$, $P = .186$).

In 1977, highly clumped sightings (Table 4) led to

TABLE 3. P_i for bison on Santa Catalina Island in 1976. $S_B = 11.69$ sightings per marked cow.

	Probabilty of resighting*			
i	Observed P_i	Poisson P_i	Equal P_i	Unequal P_i
0	0	.000	0	0
1	0	.000	0	0
2	1/16	.001	0	0
3	0	.002	0	7/16
4	0	.007	0	0
5	1/16	.015	0	0
6	1/16	.030	0	1/16
7	1/16	.050	0	0
8	0	.073	0	0
9	2/16	.094	0	0
10	1/16	.110	0	0
11	0	.117	5/16	0
12	1/16	.114	11/16	0
13	1/16	.102	0	0
14	1/16	.085	0	0
15	1/16	.067	0	0
16	3/16	.049	0	0
17	1/16	.033	0	0
18	0	.032	0	0
19	0	.013	0	0
20	1/16	.008	0	8/16
21	0	.004	0	0
22	0	.002	0	0
23	0	.001	0	0

* For the equal case, sightings of marked animals are set as close to the mean of the empirical distribution as integer additivity will allow. To generate the unequal distribution, sightings are split about the empirical mean so that generated sightings fall as close as possible to the minimum and maximum values for observed sightings. For comparison of randomly distributed sightings, binomial sightings are generated by the binomial distribution function. As n becomes large, the binomial rapidly converges to the Poisson.

TABLE 4. P_i for bison on Santa Catalina Island in 1977. S_B = 9.22 sightings per marked cow. Distributions are explained in Table 3 footnote.

i	Observed P_i	Poisson P_i	Equal P_i	Unequal P_i
0	0	.000	0	0
1	0	.000	0	0
2	1/14	.002	0	0
3	0	.008	0	0
4	1/14	.020	0	0
5	0	.039	0	0
6	0	.065	0	1/14
7	2/14	.092	0	6/14
8	0	.114	0	0
9	0	.126	1/14	0
10	1/14	.125	13/14	0
11	4/14	.113	0	0
12	0	.093	0	0
13	5/14	.071	0	7/14
14	0	.051	0	0
15	0	.033	0	0
16	0	.021	0	0
17	0	.012	0	0
18	0	.007	0	0
19	0	.003	0	0
20	0	.002	0	0
21	0	.001	0	0

overestimation of the cow population (N_u = 161, N = 175): N^P = 209.7, [176.3 ≤ N ≤ 243.1]; Monte Carlo $\hat{N}$ = 211, [203 ≤ N ≤ 221]. The generated distributions gave an $\hat{N}$ = 212 with the following likelihood intervals: equal [211 ≤ N ≤ 213], Poisson [204 ≤ N ≤ 221], unequal [204 ≤ N ≤ 220]. Comparing Poisson with observed sightings, heterogeneity does not originate from a minority of marked animals either being "sight-happy" or "sight-shy" in the tail of the empirical distribution. Indeed, the variance of the empirical distribution (σ^2 = 11.92) is hardly inflated relative to the Poisson (σ^2 = 9.93); consequently, the distributions of

N_u for observed and Poisson-generated sightings are very close. Because 9 of 14 marked cows were seen either 11 or 13 times, the sampling regime is highly suspicious (goodness of fit: χ^2 = 28.91, df = 8, P = .0005). Finally, the closeness of the likelihood intervals for the observed and unequal distributions is a clear demonstration of the extreme sampling heterogeneity in 1977.

PORCUPINES ON THE NEGEV DESERT

Indian crested porcupines (*Hystrix indica*) were captured and marked near Kibbutz Sede Boqer during a behavioral study of the porcupine (Alkon and Saltz, *in press*). Although every animal that was captured in a trap was radiotagged, we can subdivide the data in order to apply the Monte Carlo method. A total of S = 34 captures were observed during a 225-d sampling period in which 23 different animals were captured, tagged, and recaptured. We arbitrarily picked 7 animals that generated S_k = 9 resightings and treated the remaining 25 resightings as if they were generated by unmarked individuals. Table 5 shows the $\{P_i\}$. Fig. 4 shows the results of the Monte Carlo simulation. The MLE is $\hat{N}_u$ = 21 for a total of $\hat{N}$ = 28 porcupines with 95% likelihood interval of [21 ≤ N ≤ 35].

The Petersen estimate is smaller, N^P = 24.5, with 95% CI of [12.3 ≤ N ≤ 36.7]. Monte Carlo results of the three derived sighting distributions give an identical MLE of $\hat{N}$ = 27 with the following likelihood intervals: equal [25 ≤ N ≤ 30], binomial [21 ≤ N ≤ 35], and unequal [19 ≤ N ≤ 39]. Note the equality of the observed likelihood intervals with the binomial. We expect this, even for the small sample size, because the observed sightings are binomially distributed (Table 5, goodness of fit: χ^2 = 0.22, df = 2) and the variances are equal (σ^2_{obs} = 1.06, σ^2_{bin} = 1.10). In this case, both the Petersen and our method overestimate the number of animals present.
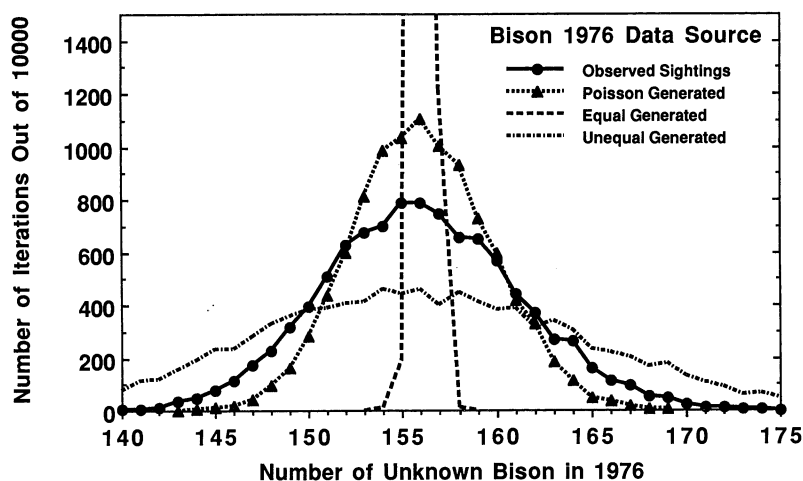

Fig. 2. Monte Carlo computation of the distribution of unknown bison on Santa Catalina Island in 1976, in 10 000 iterations. The true number of "unknown" bison was 159.
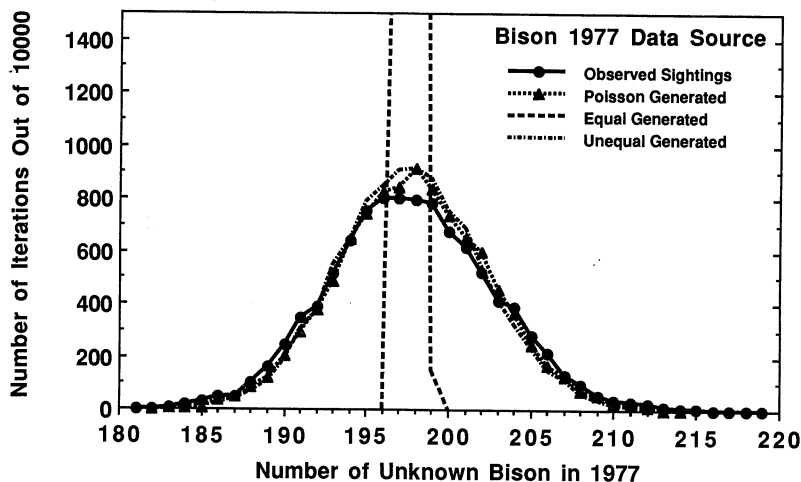
Fig. 3. Monte Carlo computation of the distribution of unknown bison on Santa Catalina Island in 1977, in 10 000 iterations. The true number of "unknown" bison was 159.

DISCUSSION

The results from any estimation procedure should be accompanied by an appraisal of sampling biases that may have arisen from violations of assumptions or from the execution of the experimental protocol (see Bowen and Sargeant 1983 for an excellent example). Since biases are unavoidable in population estimation, the real task is to identify the kinds and then gauge their severity. For any estimate requiring the marking of individuals, a suite of biases will most likely occur because the initial "marked" sample will not be a random sample of the target population. We can statistically partition this heterogeneous response to the capture/marking method into environmental variation, variation among population subgroups, and variation among individuals.

Environmental variation, such as unusual weather events that cannot be anticipated and controlled, might best be dealt with by post hoc adjustment (standard Jolly–Seber model or model M_t of Otis et al. 1978) or elimination of anomalous or outlier data. Individual variation, which may make certain animals more likely to be captured and marked, is most often explained by characteristics such as increased aggressiveness, higher mobility, being more exploratory and naive, dispersal behavior, etc. These attributes can probably be effectively accounted for by the variation among population subgroups based on the interaction of age and sex, for example, curious and naive juveniles or aggressive and more mobile resident males. Differences in capture probability associated with some identifiable morphological characteristic can be dealt with via stratification. The remaining individual variation not adequately explained by age and sex is precisely the type of variation that the Monte Carlo method adjusts for.

When the sample of marked animals is age or sex biased the investigator will not be accurately estimating the true population, because those marked animals are not identical to the remaining "unknown" population segment. If the biased sample is large enough and there is already a rough idea of the sex and age ratio of the population, then the researcher can randomly choose from the biased marked sample until the expected ratios are approximated. (The unchosen, surplus marked individuals from this adjustment are then "returned" to the unknown population segment.) Of course, if age and sex are readily identifiable in the unknown segment, then separate subpopulation estimates can be calculated as in the bison example (mature cows) and in Gauthier and Theberge (1985; adults–subadults).

Age and sex capture bias was evident in the badger data but age and sex of the unknown badger segment could not be determined from their tracks. Thus, instead of estimating subpopulations we were forced to lump all individuals for a single estimate. In this population there was an actual preponderance of adult males that was magnified by their greater susceptibility to being trapped and the greater trap-shyness of female adults. The conclusion that adult males were more susceptible is based partly on analysis of 109 independent captures of 74 badgers over 3 yr, a portion of which was implanted for the 1984 estimate. The adult sex ratio of all captured badgers was 1.75:1 males to females ($n = 77$) for adults and 0.85:1 ($n = 24$) for juveniles (young of the year). Juveniles comprised only

TABLE 5. P_i for Indian crested porcupines in the Negev Desert. $S_B = 1.29$ sightings per marked porcupine.

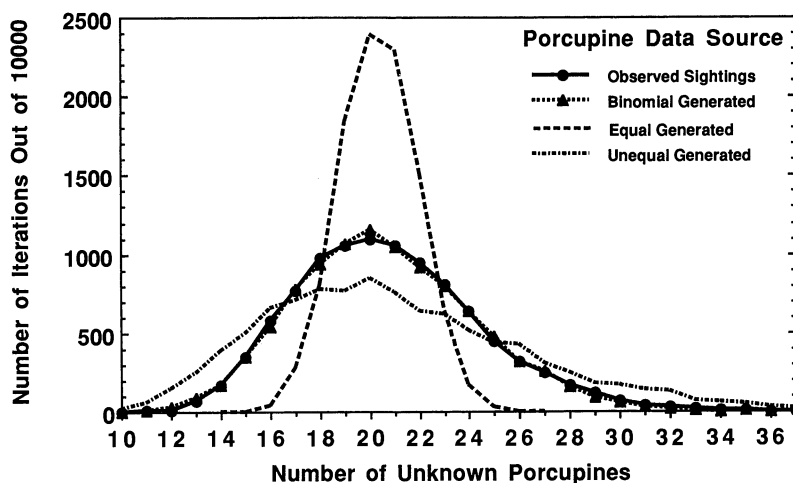| i	Observed P_i	Binomial P_i	Equal P_i	Unequal P_i
0	2/7	.250	0	4/7
1	2/7	.375	5/7	0
2	2/7	.250	2/7	0
3	1/7	.097	0	3/7
4	0	.024	0	0
5	0	.004	0	0

Fig. 4. Monte Carlo computation of the distribution of Indian crested porcupines in the Negev Desert, in 10 000 iterations.

24% ($n = 101$) of the population age structure. Of the 15 radiotagged animals in 1984, the sex ratio was 0.83:1 for adults and 1.00:1 for juveniles, which were 27% of that "known" segment. We expected the adult sex ratios to differ: the best time of the year to trap badgers for research purposes was when males of this population were most likely to enter traps and when females were least likely to do so. The severity of this bias is difficult to determine but other data points to a more realistic adult sex ratio of 1.5:1. A counterbalancing bias is that there were probably additional radiotagged adult male badgers present but not detectable with the telemetry because they were in prolonged bouts of torpor deep underground (radio signals are rapidly attenuated by frozen wet earth and snow). The differential activity levels due to age and sex can readily be seen in Table 1. During the sampling period juveniles and adult females were most active (and moved farthest) while adult males spent more time belowground and moved shorter distances when aboveground. Given the opposing directions of these biases, we made no adjustment of the data but consider the discussion worth while to illustrate the importance of understanding how the experimental layout and biases influence estimates (after Begon 1983, Hurlbert 1984).

Lack of adherence to the sampling design for the bison data helps explain the disparity between population estimates for 1976 and 1977. Sampling effort and coverage were consistent from year to year, however: (1) Some areas were not sampled as frequently as others. (2) The pattern of environmental heterogeneity was not the same each year, thus the attraction to bison of areas is different among years. And (3), bison movements and activity varied year to year on an individual basis (Lott and Minta 1983a, b). Since the data were not intended for the purposes of a population estimate, the sampling design was not entirely systematic and as a result sampling biases occurred. This is evident from the distribution of P_i's in Table

4, which is indicative of nonrandom sampling that resulted in many cows being resighted far more than others.

The porcupine example is particularly interesting for two reasons. First, the simulation MLE, $\hat{N}_u = 21$, is consistent with the direct enumeration by Alkon and Saltz (in press) and with the depletion method (Mangel and Beder 1985) used by P. U. Alkon and M. Mangel (unpublished manuscript). Second, the porcupine estimate is based on a relatively small sample yet leads to a moderately tight distribution for the number of porcupines. We are thus hopeful about the application of this method for use with other small data sets.

The method we have described is compatible with any capture–recapture design for which there are initial captures followed by an independent set of multiple recaptures in a closed population (see Arnason and Baniuk 1978 for open populations). Capture–resight, particularly when used with telemetry, offers promise for cost-effective population estimation. To date, the co-development of these techniques has emphasized estimating visibility biases for aerial surveys (e.g., Bartmann et al. 1986, Pollock and Kendall 1987, Samuel et al. 1987), mostly of big-game ungulate species. Researchers who have derived estimates of ungulates directly from telemetry and resight data have had to contend with a high initial tagging rate, large samples, high costs, and combining independent Petersen-type estimators (e.g., Rice and Harder 1977, Mackie et al. 1981, Bartmann et al. 1987, Kufeld et al. 1987).

We expect to see more capture–resight and telemetry/capture–resight applications to diverse species and circumstances. The use of telemetry is widespread, and even when telemetry is not part of an experimental design, the cost of maintaining radiotagged animals in a study area can be offset by the valuable additional data collected on movement, home range, habitat use, migration, mortality, etc. The Monte Carlo method requires a relatively small proportion of the population

to be marked. Furthermore, the flexibility of the method can accommodate continuous monitoring and accumulation of sightings over larger windows of time using various means of sighting (both from the air and ground) while collecting data for other objectives than population estimation. Sampling heterogeneity is accounted for in the magnitude and asymmetry of the likelihood intervals. Of particular interest is the ability to obtain a second independent estimate with slight additional effort. For example, one estimate can be derived through aerial surveys of marked animals (e.g., Mackie et al. 1981, Samuel et al. 1987) and another from Monte Carlo estimation based on resightings.

Many other "recapture" techniques can be used since recapture can be of any form as long as individual identification is possible: retrapping, resighting, telemetry, or sign (e.g., tracks). The badger example combined telemetry and sign (tracks and holes), but sign alone could potentially lead to an estimate; for example, K. S. Smallwood and E. L. Fitzhugh (*unpublished manuscript*) have used discriminant function analysis to successfully differentiate individual mountain lions (*Felis concolor*) from quantifiable attributes of naturally occurring track impressions.

Conclusion

Mark–resight methods for population estimation, particularly involving telemetry, have many advantages. Radiotelemetry allows the collection of information previously unavailable, particularly information concerning differential catchability (or sightability, in this case). The new simulation method for evaluation of capture–resight data is easy to use, data intensive (so that one does not ignore information contained in the capture–resight data), and has as outputs the maximum likelihood population estimate and likelihood interval information. The method may also be used on any capture–recapture data that meet the conditions and assumptions. Capture–resight and simulation methods are not meant to replace other estimation procedures. We stress that each can be an alternative choice that may conveniently fit certain types of animals, habitats, research objectives, and experimental designs.

We envision three directions for future work on this methodology: (1) the extension of the methodology to continuous marking (visual or telemetric) throughout the sampling period and (2) the explicit consideration of open vs. closed populations. (3) It may be possible to compute quantities such as relative biases of the Petersen and simulated variance estimators and efficiency of the Petersen $\hat{N}$ relative to the simulation $\hat{N}$ (J. Nichols, *personal communication*). These extensions will make the simulation approach a handy tool for every ecologist's tool kit. We appeal to biologists and statisticians to continue innovating simple methods and modifying existing ones to accommodate the wide diversity of organisms and habitats.

Literature Cited

Alkon, P. U., and D. Saltz. *In press.* Potatoes and the nutritional ecology of crested porcupines in a desert biome. Journal of Applied Ecology.

Arnason, A. N., and L. Baniuk. 1978. POPAN-2, a data maintenance and analysis system for mark–recapture data. Charles Babbage Research Centre, St. Pierre, Manitoba, Canada.

Arnason, A. N., and L. Baniuk. 1980. A computer system for mark–recapture analysis of open populations. Journal of Wildlife Management 44:325–332.

Bartmann, R. M., L. H. Carpenter, R. A. Garrott, and D. C. Bowden. 1986. Accuracy of helicopter counts of mule deer in pinyon–juniper woodland. Wildlife Society Bulletin 14:356–363.

Bartmann, R. M., G. C. White, L. H. Carpenter, and R. A. Garrott. 1987. Aerial mark–recapture estimates of confined mule deer in pinyon–juniper woodland. Journal of Wildlife Management 51:41–46.

Begon, M. B. 1983. Abuses of mathematical techniques in ecology: applications of Jolly's capture–recapture method. Oikos 40:155–158.

Bergerud, A. T. 1982. Caribou. Pages 268–270 *in* D. E. Davis, editor. CRC handbook of census methods for terrestrial vertebrates. CRC, Boca Raton, Florida, USA.

Berthold, P. 1976. Censuses in ornithology: survey and critical review. Journal für Ornithologie 117:1–69.

Blower, J. G., L. M. Cook, and J. A. Bishop. 1981. Estimating the size of animal populations. George Allen and Unwin, London, England.

Bowen, W. D., and D. E. Sargeant. 1983. Mark–recapture estimates of harp seal pup (*Phoca groenlandica*) production in the northwest Atlantic. Canadian Journal of Fisheries and Aquatic Science 40:728–742.

Brownie, C., and K. H. Pollock. 1985. Analysis of multiple capture–recapture data using band-recovery methods. Biometrics 41:411–420.

Brownie, C., and D. S. Robson. 1983. Estimation of time-specific survival rates from tag-resighting samples: a generalization of the Jolly-Seber model. Biometrics 39:437–453.

Buckland, S. T. 1982. A mark–recapture survival analysis. Journal of Animal Ecology 51:833–847.

Burnham, K. P., and D. R. Anderson. 1979. The composite dynamic method as evidence for age-specific waterfowl mortality. Journal of Wildlife Management 43:356–366.

Burnham, K. P., and W. S. Overton. 1979. Robust estimation of population size when capture probabilities vary among animals. Ecology 60:927–937.

Carothers, A. D. 1973. The effects of unequal catchability on Jolly-Seber estimates. Biometrics 29:79–100.

———. 1979. Quantifying unequal catchability and its effect on survival estimates in an actual population. Journal of Animal Ecology 48:863–869.

Connolly, G. E. 1981. Assessing populations. Pages 287–345 *in* O. C. Wallmo, editor. Mule and black-tailed deer of North America. University of Nebraska Press, Lincoln, Nebraska, USA.

Cormack, R. M. 1972. The logic of capture–recapture estimates. Biometrics 28:337–343.

Cox, D. R. 1958. Planning of experiments. Wiley, New York, New York, USA.

Eberhardt, L. L. 1978. Appraising variability in population studies. Journal of Wildlife Management 42:207–238.

Eberhardt, L. L., D. G. Chapman, and J. R. Gilbert. 1979. A review of marine mammal census methods. Wildlife Monographs 63:1–46.

Efron, B. 1982. The jackknife, the bootstrap and other resampling plans. SIAM, Philadelphia, Pennsylvania, USA.

Eisenberg, R. M. 1972. A critical look at the use of Lincoln index-type models for estimating population density. Texas Journal of Science 23:511–517.

Gauthier, D. A., and J. B. Theberge. 1985. Population characteristics of the Burwash caribou herd in the southwestern Yukon estimated by capture–recapture analysis. Canadian Journal of Zoology 63:516–523.

Greenwood, R. J., A. B. Sargeant, and D. H. Johnson. 1985. Evaluation of mark–recapture for estimating striped skunk abundance. Journal of Wildlife Management 49:332–340.

Hill, S. B., and D. H. Clayton. 1985. Wildlife after dark: a review of nocturnal observation techniques. Occasional Paper Number 17, Bell Museum of Natural History, University of Minnesota, Minneapolis, Minnesota, USA.

Hudson, D. J. 1971. Interval estimation from the likelihood function. Journal of the Royal Statistical Society B33:256–262.

Hurlbert, S. H. 1984. Pseudoreplication and the design of ecological field experiments. Ecological Monographs 54:187–211.

Jolly, G. M. 1982. Mark–recapture models with parameters constant in time. Biometrics 38:301–321.

Kufeld, R. C., D. C. Bowden, and D. L. Schrupp. 1987. Estimating mule deer density by combining mark–recapture and telemetry data. Journal of Mammalogy 68:818–825.

Lott, D. F., and S. C. Minta. 1983a. Random individual association and social group instability in American bison (*Bison bison*). Zeitschrift für Tierpsychologie 61:153–172.

Lott, D. F., and S. C. Minta. 1983b. Home ranges of American bison cows on Santa Catalina Island, California. Journal of Mammalogy 64:161–162.

Mackie, R. J., K. L. Hamlin, and D. F. Pac. 1981. Census methods for mule deer. Pages 97–106 *in* F. L. Miller and A. Gunn, editors. Symposium on census and inventory methods for populations and habitats. Forest, Wildlife, and Range Experiment Station, University of Idaho, Moscow, Idaho, USA.

Mangel, M., and J. H. Beder. 1985. Search and stock depletion: theory and application. Canadian Journal of Fisheries and Aquatic Sciences 42:150–163.

Mares, M. A., K. E. Streilein, and M. R. Willig. 1981. Experimental assessment of several population estimation techniques on an introduced population of eastern chipmunks. Journal of Mammalogy 62:315–328.

Menkens, G. E., and S. H. Anderson. 1988. Estimation of small-mammal population size. Ecology 69:1952–1959.

Montgomery, W. I. 1985. Interspecific competition and the comparative ecology of two congeneric species of mice. Pages 126–187 *in* L. M. Cook, editor. Case studies in population biology. Manchester University Press, Manchester, England.

Nichols, J. D., B. R. Noon, S. L. Stokes, and J. E. Hines. 1981. Remarks on the use of mark–recapture methodology in estimating avian population size. Studies in Avian Biology 6:121–136.

Nichols, J. D., and K. H. Pollock. 1983. Estimation methodology in contemporary small mammal capture–recapture studies. Journal of Mammalogy 64:253–260.

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.

Pank, L. F. 1981. Summarizing remarks: estimating birds per unit area. Studies in Avian Biology 6:162.

Patterson, I. J. 1978. Tags and other distant-recognition markers for birds. Pages 54–62 *in* B. Stonehouse, editor. Animal marking: recognition marking of animals in research. MacMillan, London, England.

Pollock, K. H. 1981. Capture–recapture models: a review of current methods, assumptions, and experimental design. Studies in Avian Biology 6:426–435.

———. 1982. A capture–recapture design robust to unequal probability of capture. Journal of Wildlife Management 46:752–757.

Pollock, K. H., and W. L. Kendall. 1987. Visibility bias in aerial surveys: a review of estimation procedures. Journal of Wildlife Management 51:502–510.

Pollock, K. H., and D. G. Raveling. 1982. Assumptions of modern band-recovery models, with emphasis on heterogeneous survival rates. Journal of Wildlife Management 46:88–98.

Rice, W. R., and J. D. Harder. 1977. Application of multiple aerial sampling to a mark–recapture census of white-tailed deer. Journal of Wildlife Management 41:197–206.

Roff, D. A. 1973a. On the accuracy of some mark–recapture estimators. Oecologia (Berlin) 12:15–34.

———. 1973b. An examination of some statistical tests used in the analysis of mark–recapture data. Oecologia (Berlin) 12:35–54.

Samuel, M. D., E. O. Garton, M. W. Schlegel, and R. G. Carson. 1987. Visibility bias during aerial surveys of elk in northcentral Idaho. Journal of Wildlife Management 51:622–630.

Seber, G. A. F. 1970. The effects of trap response on tag recapture estimates. Biometrics 26:13–22.

———. 1982. The estimation of animal abundance and related parameters. Second edition. Griffin, London, England.

———. 1986. A review of estimating animal abundance. Biometrics 42:267–292.

Skalski, J. R. 1985. Construction of cost functions for tag-recapture research. Wildlife Society Bulletin 13:273–283.

Skalski, J. R., and D. S. Robson. 1982. A mark and removal field procedure for estimating population abundance. Journal of Wildlife Management 46:741–751.

Skalski, J. R., M. A. Simmons, and D. S. Robson. 1984. The use of removal sampling in comparative censuses. Ecology 65:1006–1015.

Smith, M. H., R. H. Gardner, J. B. Gentry, D. W. Kaufman, and M. H. O'Farrell. 1975. Density estimations of small mammal populations. Pages 25–53 *in* F. B. Golley, K. Petrusewicz, and L. Ryszkowski, editors. Small mammals: their productivity and population dynamics. Cambridge University Press, New York, New York, USA.

Tanaka, R. 1980. Controversial problems in advanced research on estimating population densities of small rodents. Researches in Population Ecology, Supplement 2:1–67.

Tepper, E. E., editor. 1967. Statistical methods in using mark–recapture data for population estimation. Biobibliography Number 4, Department Library, United States Department of the Interior, Washington, D.C., USA.

Tidman, K. R. 1984. The operation's evaluation group. Naval Institute Press, Annapolis, Maryland, USA.

White, G. C., D. R. Anderson, K. P. Burnham, and D. L. Otis. 1982. Capture–recapture and removal methods for sampling closed populations. Los Alamos National Laboratory, Los Alamos, New Mexico, USA.

APPENDIX

I. Motivation for the Monte Carlo Simulation

The situation to be considered is the following. A closed population of animals is initially sampled and some relatively small fraction of the animals is tagged so that their presence can be identified. Following the initial tagging, the population is observed over a long period of time. The problem is then to estimate the total population.

For definiteness, consider the badger problem in which the population is observed directly through telemetry and indirectly through badger snowtracks. The simplest case is one in which each member of the population makes exactly the same number of tracks. Let N denote the fixed total number of badgers in the population and let n denote the fixed number of badgers that are tagged. If each badger (they are all identical in this first assumption) makes a fixed number of tracks t, the number of tracks made by marked badgers is

$$T_M = nt \tag{A.1}$$

and the number of tracks made by unmarked badgers is

$$T_U = (N - n)t. \tag{A.2}$$

The analogue of the Petersen estimate for the number of untagged badgers is determined by

$$\bar{B} = \frac{\text{(number of unknown tracks)}}{\text{(number of known tracks per badger)}} \tag{A.3}$$

$$= \frac{T_U}{T_M/n} = \frac{(N - n)t}{nt/n}$$

$$= N - n. \tag{A.4}$$

When T_M and T_U are random variables, then

$$E(T_M) = nt \tag{A.1b}$$

$$E(T_U) = (N - n)t, \tag{A.2b}$$

and we can use the large-sample approximation

$$E(\hat{B}) = E\left[\frac{T_U}{T_M/n}\right] \approx \frac{E(T_U)}{E(T_M)/n} = N - n. \tag{A.4b}$$

There are two types of heterogeneity that will affect this simple analysis. The first is population heterogeneity: not all individuals will make exactly the same number of tracks. The second is sampling heterogeneity when the initial tagging is performed and within and among the resighting periods.

First consider heterogeneity of the population. To make things as simple as possible, suppose that there are only two types of badgers. One type makes t_1 tracks and the other type makes t_2 tracks. Suppose that there are N_1 of the first type of animals and N_2 of the second type of animals in the population, with $N_1 + N_2 = N$, the total population size. Let n_i denote the number of tagged animals of type i. The number of tracks made by marked animals is now

$$T_M = n_1 t_1 + n_2 t_2 \tag{A.5}$$

and the number of tracks made by unmarked animals is now

$$T_U = (N_1 - n_1)t_1 + (N_2 - n_2)t_2. \tag{A.6}$$

The estimate for the number of unmarked animals still follows the same reasoning as Eq. A.3 and is thus

$$\bar{B} = \frac{T_U}{T_M/(n_1 + n_2)} \tag{A.7}$$

$$= \frac{[(N_1 - n_1)t_1 + (N_2 - n_2)t_2][n_1 + n_2]}{n_1 t_1 + n_2 t_2}.$$

The usual Petersen estimate implicitly assumes that the initial sampling is perfect. That is, suppose that the probability of capturing an animal in the initial sampling is p (this may be small, especially if the initial tagging period is short, as it typically is). The implicit assumption about sampling heterogeneity in the Petersen estimate is that

$$n_i = pN_i. \tag{A.8}$$

If this assumption is used in Eq. A.7, one obtains

$$\bar{B} = \frac{[N_1(1 - p)t_1 + N_2(1 - p)t_2][p(N_1 + N_2)]}{pN_1 t_1 + pN_2 t_2} \tag{A.9}$$

$$= (1 - p)(N_1 + N_2) = (1 - p)N.$$

Since the initial number of animals tagged is pN, the total estimate for the population is unbiased. However, this procedure requires that the assumption of perfect sampling hold.

In most circumstances, Eq. A.8 will not hold. Instead, if the probability of capturing an animal for the initial sampling is p, one should think of the n_i as random variables. Assuming random trapping in the initial phase, and that all animals are equally vulnerable, the random variable n_i will have a binomial distribution with parameters N_i and p. One then needs to consider the expected value of the estimate $\bar{B}$. It is given by

$$E\{\bar{B}\} = \sum_{m_1} \sum_{m_2} Pr\{n_1 = m_1, n_2 = m_2\} B(m_1, m_2), \tag{A.10}$$

where $Pr\{n_1 = m_1, n_2 = m_2\}$ is the product of two binomial expressions and

$$B(m_1, m_2) = \frac{[(N_1 - m_1)t_1 + (N_2 - m_2)t_2](m_1 + m_2)}{m_1 t_1 + m_2 t_2}. \tag{A.11}$$

The expression $B(m_1, m_2)$ is essentially the Petersen estimate, using the m_i. The usual approach would be to follow the so-called "Seber delta method" (which goes back to the Operations Evaluation Group in World War II [Tidman 1984]) and compute $E\{\bar{B}\}$ in a Taylor series of the nonlinear estimate $B(m_1, m_2)$. A second approach, especially appropriate for this simple problem, is to compute the expectation exactly, using Eq. A.10. When doing this, one needs to be careful about the limits of the summation. In particular, the case in which $m_1 = m_2 = 0$ must be excluded. Table A1 shows the results for the computation of the entire population size, which is

$$E\{\hat{N}\} = \sum_{m_1} \sum_{m_2} Pr\{n_1 = m_1, n_2 = m_2\} \cdot [B(m_1, m_2) + m_1 + m_2]. \tag{A.12}$$

As a check on the computations, one can set $t_1 = t_2$ in Eq. A.12; the expected value of the population is then $N_1 + N_2$ in all cases.

What information can be gleaned from Table A1? The estimate becomes unbiased for either large values of p (in which case most of the population is trapped initially!) and large values of N (in which case the central limit theorem smooths out fluctuations in the initial sampling process). It is clear, however, that the usual Petersen estimate may be highly biased if either the total population is small (and 40 animals may not be particularly small) or the initial probability of capture for the tagging is small ($p = 0.1$ in a population of 100 animals corresponds to tagging about 10 animals; this is not unreasonable). The apparent positive bias in the estimate is not a consistent feature of the analysis. For example, if one extends Eq. A.12 to the case of three classes of animal, the estimated population may be above or below the true population, as shown in Table A2.

TABLE A1. Population estimates for two variable subpopulations based on constant, heterogeneous tracking by members of those subpopulations.

Population parameters				Probability of capture during initial sampling, p						
N_1	N_2	t_1	t_2	.05	.10	.15	.20	.25	.30	.35
				Population estimate						
20	20	9	1	79	58	49	45	43	42	42
10	30	9	1	77	63	54	49	46	45	43
30	10	9	1	63	48	43	42	41	41	41
50	50	9	1	134	109	105	103	102	102	101
20	80	9	1	146	120	111	108	105	104	103
80	20	9	1	111	102	101	101	101	101	100

TABLE A2. Population estimates for three variable subpopulations based on constant, heterogeneous tracking by members of those subpopulations.

Population parameters		Probability of capture during initial sampling, p				
(N_1, N_2, N_3)	(t_1, t_2, t_3)	.05	.10	.15	.20	.25
		Population estimate				
(5, 10, 15)	(0, 5, 10)	38	37	35	34	33
(15, 10, 5)	(0, 5, 10)	22	24	26	28	29
(15, 8, 7)	(0, 5, 10)	24	26	28	29	30
(7, 8, 15)	(0, 5, 10)	37	36	35	34	33

What conclusion should be drawn? The easiest way out, to paraphrase William Feller, is to take the approach of the statistical formalist: do not try to do anything unless one is certain that either p or N is sufficiently large for the Petersen estimate to work. On the other hand, there are many applied and practical problems in which "not doing anything" is an inappropriate answer; some kind of population estimate is needed. It is for those problems, in which sampling heterogeneity causes the Petersen estimate to possibly be highly biased (and one will never know what the true situation is) that the method developed by us is applicable. Another motivation for our method is provided by Menkens and Anderson (1988).

II. DESCRIPTION OF THE MONTE CARLO SIMULATION

The inputs to the Monte Carlo program are (1) S_u = number of sightings of unmarked animals, and (2) the frequencies $\{P_i\}$ derived from the empirical distribution of sightings. In each draw of the program, a random number of sightings, $\tilde{s}$, is picked as follows. First one draws a random number, Y, uniformly distributed on [0, 1]. The number of sightings $\tilde{s}(Y)$ is then such that

$$\sum_{i=0}^{s(Y)} P_i \le Y < \sum_{i=0}^{s(Y)+1} P_i. \qquad (A.13)$$

During the simulation

F_j = number of times that j unmarked animals generate the S_u sightings
T = number of sightings in the current iteration
T_s = specified stopping rule
N_I = number of iterations of the Monte Carlo program.

The simulation proceeds as follows (for 10 000 iterations):

1) Initialize: $N_I = 0$, $\{F_j\} = 0$, $1 \le j \le S_u$, input S_u, $\{P_i\}$.
2) Initialize each iteration: $T = 0$, $N_u = 0$.
3) Draw a random number of sightings $\tilde{s}$ from the empirical distribution.
4) Set $T = T + \tilde{s}$, $N_u = N_u + 1$.
5) Is $T > T_s$?
   (a) Yes: Step 6.
   (b) No: Step 3.
6) Increase F_{N_u} by 1.
   Increase N_I by 1.
7) Does $N_I = 10\,000$?
   (a) Yes: Step 8.
   (b) No: Step 2.

8) Define f_j by $f_j = \dfrac{F_{N_u}}{10\,000}$ for $j = 1, \ldots S_u$.

For the examples we have presented, we set the stopping rule, T_s, equal to S_u; therefore, the bootstrap distribution generated is conditional on the observed value of S_u. In some situations, variability associated with S_u, particularly due to small sample sizes, could change estimators and likelihood intervals. To allow variation in S_u, our program employs two user-specified stopping rules, a lower and an upper T_s. Since Step 5 allows T to exceed T_s, we recommend using S_u as the upper T_s, and some number greater than $S_u - N_m$ (where N_m is maximum

number of resightings for any marked animal) as the lower T_s. To examine sensitivity of the simulation to S_u variability, a good candidate for lower T_s might be $S_u - N_{med}$ where N_{med} is the median of the distribution of resightings. We found little difference between $T_s = S_u$ and $T_s = S_u - N_{med}$ for the text examples. In the extreme case of $T_s = S_u - N_m$, we found that for the observed and binomial/Poisson distributions in the text, N_u estimates and their likelihood intervals averaged only 11.3% less (SD = 12.7%) than estimates from $T_s = S_u$. The width of the likelihood intervals remained nearly constant, which translates into a slight increase in the interval, relative to the decreased estimator.

The output of the algorithm is a distribution $\{f_j\}$ where f_j is the fraction of iterations in which j animals generated the S_u tracks. The MLE for N_u is the value of j for which f_j is the largest. Probability intervals containing a fraction α of the total probability are found by solving for i,

$$\sum_{j=j^*-i}^{j^*+i} f_j \ge \alpha, \qquad (A.14)$$

where j^* is the MLE for N_u.

Values of N_u that are a fraction $j\alpha$ as likely as the MLE are found by solving for the closest j such that

$$f_j = \alpha f_{j^*}. \qquad (A.15)$$

Hudson (1971) examined "likelihood intervals" that are derived directly from the likelihood function simply by drawing a horizontal line across the graph of the likelihood function. Given the MLE estimate of $\hat{\theta}$, a "likelihood interval" for parameter θ $I(\theta; \beta | j)$, is defined by

$$I(\theta; \beta | J) = \{\theta \cdot l(\theta | j) \ge e^{-\beta} \cdot l(\hat{\theta} | j)\}, \qquad (A.16)$$

and leads to an asymptotic 95.4% confidence interval for large samples when $\beta = 2$. It is reasonably approximated by 0.954 in small samples. The choice of $\beta = 2$ depends neither on the distribution of j, nor the sample size, nor on the stopping rule, nor whether the sample is complete or censored, and so on, but only on the asymptotic confidence coefficient (Hudson 1971). For our estimate, the approximate 95% likelihood interval is contained by $\{j_1 \le j^* < j_2\}$ where $e^{-\beta} \cdot f_{j^*} = 0.135 f_{j^*}$ $= f_{j1} = f_{j2}$. Graphically, the horizontal line drawn through the distribution of j at $0.135 f_{j^*}$ intersects k_1 and k_2.

For a copy of the Basic program (IBM compatible), send blank floppy disk to S. Minta.