Population estimation with sparse data: the role of estimators versus indices revisited

Kevin S. McKelvey and Dean E. Pearson

Abstract: The use of indices to evaluate small-mammal populations has been heavily criticized, yet a review of small-mammal studies published from 1996 through 2000 indicated that indices are still the primary methods employed for measuring populations. The literature review also found that 98% of the samples collected in these studies were too small for reliable selection among population-estimation models. Researchers therefore generally have a choice between using a default estimator or an index, a choice for which the consequences have not been critically evaluated. We examined the use of a closed-population enumeration index, the number of unique individuals captured (M_{t+1}), and 3 population estimators for estimating simulated small populations ($N = 50$) under variable effects of time, trap-induced behavior, individual heterogeneity in trapping probabilities, and detection probabilities. Simulation results indicated that the estimators produced population estimates with low bias and high precision when the estimator reflected the underlying sources of variation in capture probability. However, when the underlying sources of variation deviated from model assumptions, bias was often high and results were inconsistent. In our simulations, M_{t+1} generally exhibited lower variance and less sensitivity to the sources of variation in capture probabilities than the estimators.

Résumé : L'utilisation d'indices pour évaluer les populations de petits mammifères est très critiquée et pourtant les études publiées entre 1996 et 2000 sur les petits mammifères démontrent que l'utilisation d'indices est toujours la méthode la plus courante d'évaluation des populations. Dans la littérature, 98 % des échantillons recueillis pour les études étaient trop petits pour permettre le choix judicieux d'un modèle d'estimation de population. Les chercheurs se retrouvent donc devant un choix à faire entre une méthode par défaut ou un indice, choix dont les conséquences n'ont pas été évaluées de façon formelle. Nous avons examiné les résultats de l'utilisation d'un indice de dénombrement d'une population fermée, du nombre d'individus particuliers capturés (M_{t+1}), et de 3 estimateurs de population pour estimer de petites populations simulées ($N = 50$) soumises à des effets divers du temps, du comportement relié au piégeage, de l'hétérogénéité dans la probabilité de capture des individus et des probabilités de détection. Les résultats de cette simulation ont démontré que les estimateurs donnent des évaluations qui comportent peu d'erreur et qui sont d'une grande précision lorsque l'estimateur utilisé reflète les sources de variation sous-jacentes des probabilités de capture. Cependant, lorsque les sources de variation sous-jacentes s'éloignent des présuppositions du modèle, les chances d'erreur sont souvent élevées et les résultats sont changeants. Dans nos simulations, M_{t+1} a généralement une faible variance et manifeste moins de sensibilité aux sources de variation des probabilités de capture que les autres estimateurs.

[Traduit par la Rédaction]

Introduction

Since the early days of wildlife research, workers studying population biology have debated the use of indices versus statistical models for evaluating population size. This debate appears to be particularly rooted in the small-mammal literature, owing to the ubiquity of capture–recapture studies within this field (Hilborn et al. 1976; Otis et al. 1978; Jolly and Dickson 1983; Nichols and Pollock 1983; Nichols 1986; Efford 1992; Rosenberg et al. 1995; Slade and Blair 2000; Tuyttens 2000). For instance, Nichols and Pollock (1983) identified numerous arguments made by mammalogists for using indices rather than capture–recapture models and rebutted all of them on the basis of simulations and statistical theory. Despite these admonitions, researchers continue to use indices as a common tool for estimating population size (Montgomery 1987; Slade and Blair 2000).

Most published comparisons of estimators and indices have focused on large ($N \geq 100$) well-sampled populations, and have generally concluded that estimators are preferable to indices because they result in less bias (Otis et al. 1978; White et al. 1982; Jolly and Dickson 1983; Nichols and Pollock 1983; Pollock et al. 1990; Efford 1992). Such comparisons, however, are largely academic to the extent that they ignore small samples and small populations, which constitute a large proportion of the samples obtained in small-mammal studies (Slade and Blair 2000). Additionally, for relative comparisons, accuracy is less important than precision, and many ecological questions are relative in nature.

Relationship between indices and estimators

The use of indirect measurements, or indices, is ubiquitous in ecology because many parameters of interest cannot be directly measured. In some cases, such as capture–recapture studies, attempts are made to convert indirect measurements

Received February 12, 2001. Accepted August 15, 2001. Published on the NRC Research Press Web site at http://cjz.nrc.ca on September 21, 2001.

K.S. McKelvey[1] and D.E. Pearson. Rocky Mountain Research Station, U.S. Department of Agriculture Forest Service, P.O. Box 8089 Missoula, MT 59807, U.S.A.

[1]Corresponding author (e-mail: kmckelvey@fs.fed.us).

 DOI: 10.1139/cjz-79-10-1754

into measures of desired, but unmeasured, parameters through the use of models. When model forms are uncertain, and model parameters are imprecise estimates, there are always costs associated with these conversions, and the issue of whether or not a particular conversion is desirable is seldom straightforward.

Indices are obtained whenever the desired (D) and measured (I) metrics differ but are presumed to be functionally related. For our purposes, an index is an indirect measure for which the relationship between D and I is not quantified. When an attempt is made to convert an index into an estimator by quantifying this functional relationship, there are several sources of error: sampling error associated with measuring the index and related function parameters, the degree to which the functional form is correct, and the extent to which D and I are actually related. In general we measure I because D is difficult to measure, and for this reason the functional form and parameterization of the conversion will generally be estimated from smaller datasets with relatively high variance. Thus, unless the parameterization data were collected representatively, they may not reflect the population to which they are applied. Introduced sampling error associated with using poorly measured correction parameters can occlude the correlation between D and I, and using an incorrect functional form and (or) non-representative parameter estimates will introduce unknown levels of bias. Hence, when we choose to convert an index, we do so on the assumption that increased variability and the introduction of unknown levels of bias are more than offset by the benefits associated with obtaining estimates of D.

As an example, down wood is considered to be an important biological parameter (Bull et al. 1997; Pearson 1999), but is difficult to directly measure in terms of volume or mass. Hence, it is common to measure it indirectly using line-intercept methods. Having made the decision to indirectly measure wood, one has several choices: to convert the intercept data into volume or mass using either locally or literature-derived functions (e.g., Brown and See 1981), or simply to use the intercept data directly. Clearly, the least error will be associated with using the data directly, therefore the decision as to whether to convert the data is based on the reliability of the equations and the importance of the conversion. If one is studying small mammals, then down wood mass or volume is simply an index of some unknown attributes of down wood that the organisms are using (Pearson 1999). Arguably, the conversion does not improve understanding and adds unknown error and bias. However, if one is estimating fire behavior, then conversion to mass may be critical. Converting an index into an estimate leads to improved understanding if enough information is available for both choosing a proper conversion function and reliably estimating its value. Additionally, the converted value must be in some way more biologically meaningful than the unconverted data.

Index conversion in small-mammal capture–recapture studies

In trapping studies of closed populations, we can seldom reliably capture all of the animals in an area. We therefore collect data for an index: the number of unique individuals captured, M_{t+1} (see Otis et al. 1978). Because $0 \leq M_{t+1} \leq N$, where N is the size of the sampled population, M_{t+1} is negatively biased. To convert M_{t+1} into N, various functions based on recapture data can be used. This approach has a number of potential problems. In capture–recapture analysis, the functional relationship between M_{t+1} and N can be neither measured nor verified in any absolute sense, and hence cannot be directly estimated. Instead the collected data are analyzed in order to understand the underlying sources of variation in capture probabilities, and this understanding is subsequently used to both define and control the conversion. For closed populations, these methods were codified in the computer program CAPTURE in the late 1970s (Otis et al. 1978; White et al. 1982). With relatively large sample sizes and high capture probabilities, CAPTURE employs discriminant function analysis to effectively select from 8 potential models addressing the important sources of variability in capture probabilities and produces population estimates with low bias and high precision (Otis et al. 1978; White et al. 1982). However, when samples are small and capture probabilities low, model selection is poor and population estimates are unreliable (Otis et al. 1978; Menkins and Anderson 1988; Hallett et al. 1991).

Because it is seldom feasible to test capture–recapture models directly, much of what we know concerning their behavior is based on simulation. Simulation studies have generally focused on comparisons of indices and estimators for large populations ($N \geq 100$: e.g., Otis et al. 1978; Jolly and Dickson 1983; Nichols and Pollock 1983; Pollock et al. 1990; Efford 1992; Tuyttens 2000). The use of closed-population models for small samples has been poorly studied (Menkins and Anderson 1988), especially with regard to comparisons with indices. One exception is Rosenberg et al. (1995), who specifically looked at jackknife (Burnham and Overton 1978) and Chao (1988) estimators for small populations and low capture probabilities. However, Rosenberg et al. (1995) only examined the case where the heterogeneity models they tested were consistent with the simulated sources of variability in capture probabilities. They did not examine the behavior of these models when other factors such as time and trap-induced behavior influence the sampling. A closer examination of the potential efficacy of indices and estimators for population estimation when data are sparse is therefore warranted.

We address 3 primary objectives with this research: (1) to determine the distribution of sample sizes reported in small-mammal population studies and relate this distribution to the feasibility of applying model-selection procedures, (2) to assess the efficacy of using the jackknife, null (Otis et al. 1978; White et al. 1982), and sample-coverage (Chao et al. 1992; Lee and Chao 1994) models and M_{t+1} as default population estimators when small sample sizes preclude effective model selection, and (3) assess the performance of these estimators and M_{t+1} for making relative comparisons between study populations, given sample sizes comparable to those observed in the small-mammal literature.

Methods

Literature review

We examined all articles published over the last 4 years (the first issues of 1996 through the most recent issues from 2000) from 4

Table 1. Parameters associated with each of the 8 population attributes simulated.

Simulated population attribute*	Average per-session detection probability	N	t	b	h
M_0	0.3–0.6	50	0.00	0.00	0.00
M_t	0.3–0.6	50	±0.35	0.00	0.00
M_b	0.3–0.6	50	0.00	–0.70	0.00
M_h	0.3–0.6	50	0.00	0.00	±0.30
M_{bh}	0.3–0.6	50	0.00	–0.70	±0.30
M_{th}	0.3–0.6	50	±0.35	0.00	±0.30
M_{tb}	0.3–0.6	50	±0.25	–0.50	0.00
M_{tbh}	0.3–0.6	50	±0.25	–0.50	±0.30

Note: Behavior and time effects were reduced when they were combined in the same model to allow reliable computation of the estimators.

*The subscripts *t*, *b*, and *h* refer to sources of variation in capture probability associated with time, trap-induced behavior, and individual capture heterogeneity, respectively; M_0 lacks these sources of variation.

journals that commonly publish small-mammal studies from North America, South America, and Europe: *Acta Theriologica*, *Canadian Journal of Zoology*, *Journal of Mammalogy*, and *The Journal of Wildlife Management*. We extracted population estimates or sample sizes from small-mammal papers that attempted to derive abundance or density. When data were presented for multiple species, multiple sampling locations, or multiple years, we treated each as a separate sample. Where multiple estimates were made for a species within a single year at the same location, we used the peak population sample period only. When averages were presented, values were rounded to the nearest whole number. Zeros were used when authors presented zeros as estimates or when animals were not detected in one sampling unit or period but were known to be present in others. In some cases estimates reflected species complexes rather than single species (e.g., *Microtus* spp. or *Sorex* spp.), resulting in a positive bias in the estimated individual species sample.

For each paper we also grouped the estimation methods into the following categories: the basic enumeration indices, described as the number of unique individuals captured (M_{t+1}, for closed populations) or the minimum number known alive index (MNKA, for open populations; Krebs 1966); relative-abundance indices such as the number of captures or unique individuals captured per 100 or 1000 trap-nights; CAPTURE, or a specific estimation approach if it was used as a default rather than derived from CAPTURE or other model-selection procedures. If multiple approaches were used to estimate abundance (e.g., CAPTURE for some species and an index for others), each method was included separately in the results.

Simulations

Population estimation

We simulated closed populations having 8 underlying population attributes as defined by Otis et al. (1978). These attributes include as independent sources of variation, probability of detection differing across time (*t*) or across individuals (*h*), and capture probability conditioned on previous capture, or behavioral response (*b*). Following Otis et al. (1978), simulated population attributes (see models in Otis et al. 1978) were characterized by their sources of variation: M_t has capture probabilities that vary across time; M_{tbh} has capture probabilities that vary across time, individuals, and capture history. M_0 refers to a "null model" in which capture probability is held constant. All simulations were based on a population (N) of 50 individuals and four trapping periods. Results were based on 2000 simulations of each combination of estimation model and population attribute. Reported coefficients of variation (CVs) are based on the simulation results rather than the CVs generated by CAPTURE or otherwise statistically derived.

The primary purpose of the simulations was to examine the relative efficacy of various estimation methods when confronted with an unknown set of population attributes. We therefore made the effects of heterogeneity, time, and behavior large. Additionally, we wanted at least one of the simulated attributes to be correct for each estimator; these simulations serve as tests of the maximum efficacy of each estimator, given the small samples and relatively low number of trapping periods. There are many ways to simulate the sources of variation associated with time, heterogeneity, and specific behavioral responses to trapping. Because estimation models change their behavior with changes in N, detection probability, and details of the underlying population attributes (see Boulanger and Krebs 1996), we do not claim that our results are exhaustive. In setting parameters for the simulations, we attempted to emulate real conditions based on the small-mammal literature and our trapping experience. However, the relationships of these simulated populations to the actual situations encountered during small-mammal trapping are unknown.

The primary concern associated with using an index such as M_{t+1} is that probabilities of detection vary across space and time, potentially confounding the results. To simulate this, we allowed the average per-session trapping probabilities to vary randomly between 0.3 and 0.6 for each simulation. Adding other factors such as time and heterogeneity did not change average probabilities of detection (or the probability of first detection in processes that included trap-conditioned behavior).

Heterogeneity was modeled as being intrinsic to each organism and constant over time. We wanted heterogeneity to be as large as possible while keeping the distribution constant and changing average probabilities of detection. We therefore assigned to each organism at the beginning of each simulation a specific probability of detection drawn from a uniform random deviate ± 0.30 and centered on the average capture probability; if the probability of detection was 0.31, then individual probabilities of detection varied randomly between 0.01 and 0.61. Time was modeled as a random multiplicative effect, constant across the population but varying with trap-night. We modeled trap-shy behavior as a multiplicative effect applied after first capture: it was fixed for all organisms that had been trapped at least once (Table 1). We did not model trap-happy behavior.

We explored the behavior of 3 estimation models: the null model, which assumes constant capture probability (Otis et al. 1978; White et al. 1982); the jackknife model (Burnham and Overton 1978, 1979), which assumes only heterogeneity; and a more recent model based on sample coverage (Chao et al. 1992; Lee and Chao 1994), which can correct for heterogeneity and time. We refer to this model as C_{th}. We use the estimator labeled C_2 in Chao et al. (1992), which includes second-order bias adjustment and generally gives population estimates slightly lower than does the estimator lacking bias adjustment (Chao et al. 1992; Lee and Chao 1994). In all cases we used the same algorithms to produce the sample file, but in the case of the null and jackknife estimators, we used the most recent version of CAPTURE (Rexstad and Burnham 1991) to estimate the results, whereas the C_{th} model was coded by the authors.

We chose to test the jackknife model, as it has been shown to be robust to violations in underlying assumptions (Otis et al. 1978; Burnham and Overton 1979; Boulanger and Krebs 1994, 1996), and therefore has the potential to serve as a default model (a model chosen by the researcher rather than through a statistically valid model-selection process) when samples are too small for effective model selection (Rosenberg et al. 1995). Menkins and Anderson (1988) suggest using the Lincoln–Petersen estimator (Seber 1982) when data are sparse. However, we chose to test the null instead of

the Lincoln–Petersen model because the null model functions as a default model in CAPTURE (the program tends to choose the null model when data are sparse (White et al. 1982)). Additionally, we tested the sample-coverage estimator (C_{th}) because it has the property of collapsing into simpler models when time or heterogeneity factors are absent. For instance, C_{th} collapses approximately into the null model when probabilities of capture are held constant (M_0), whereas the jackknife model shows more variable bias under the same conditions (Fig. 1).

Ordinal ranking

We performed a second test to compare the ability of the estimators to determine the ordinal ranking of 2 populations. For this test we used the same models, probabilities of detection, and four trapping periods that were used to test the accuracy of population estimation. However, in these simulations we examined the ability of the models to detect a 25% population decline, from 50 to 37 individuals. We looked at two scenarios. In the first scenario, population attributes and probabilities of detection were allowed to change randomly between the two sampling periods. In the second, we kept the population attributes fixed and restricted them to those appropriate for the tested estimators: tests between M_{t+1} and the jackknife and C_{th} models used the M_h population attribute for both sampling periods, and comparisons between M_{t+1} and the null model used M_0. Probability of detection was allowed to vary between sampling periods for all simulations. The estimator was deemed correct if the population estimate for the second sampling period was less than the estimate for the first sample, and incorrect otherwise.

Results

Literature review

Most researchers reporting results of small-mammal studies between 1996 and 2000 used indices (66.7%) instead of estimators (33.3%; Fig. 2). The most commonly used index was MNKA (35.3%), followed by M_{t+1} (23.5%) and relative-abundance indices (7.8%). CAPTURE was the only model-selection program used. Program MARK (White and Burnham 1999) was used once for estimating survival but never for selecting a statistical model or estimating population size. Sample sizes were generally well below those required for effective model selection using CAPTURE (Otis et al. 1978; Menkins and Anderson 1988; Fig. 3). Ninety-four percent of the samples were ≤50 and 98% were ≤100. Only the extreme tail of the distribution lay within the range of sample sizes necessary for effective model selection. This portion of the distribution contained only a few of the more common species of small mammals (e.g., *Microtus pennsylvanicus* and *Peromyscus maniculatus*), and even these relatively abundant species only attained the requisite sample sizes when their populations were particularly large. Some studies therefore achieved samples large enough for the use of population estimators for one or two sampling periods for one or more plots, but most samples were too small for effective model selection, particularly in multiyear studies. In some cases researchers used estimators for some of the data while employing indices for other species or times (e.g., Sullivan et al. 1998; Von Trebra et al. 1998; Hanley and Barnard 1999).

Simulations

Population estimation

The behavior of the 3 population estimators was strongly influenced by the underlying population attributes (Table 2). All 3 estimation models produced highly biased estimates when applied to simulated populations that exhibited behavioral responses to trapping (M_b, M_{bh}, M_{tb}, M_{tbh}). Trap-shy behavior produces relatively few recaptures, and the models interpreted this as an indication of large populations. The C_{th} model, in situations where it was appropriate, was consistently biased low. The uncorrected version (Lee and Chao 1994) of this estimator, while not presented here, is biased slightly high under the same circumstances. In our simulations the jackknife estimator was generally biased high, behavior that has been noted before (Boulanger and Krebs 1994, 1996; see also Table 3.14 in White et al. 1982). In addition to bias, all of the population estimates were, at least in certain circumstances, correlated with M_{t+1} (Fig. 4). The CV for M_{t+1} was lower in most cases than the CVs associated with the capture–recapture models, even though probability of detection was allowed to vary randomly between 0.30 and 0.60 for each simulation.

Ordinal ranking

When the population attributes varied, M_{t+1} outperformed the null, jackknife, and C_{th} estimators by 16.8, 12.8, and 18.9%, respectively (Table 3). When the population attributes were invariant and were appropriate for the estimator, performance was approximately equal, with the jackknife and null models slightly outperforming M_{t+1}.

Discussion

The role of indices versus estimators in studies of small-mammal populations remains unresolved. Despite arguments against the use of indices (e.g., Jolly and Dickson 1983; Nichols and Pollock 1983; Nichols 1986), they continue to be used twice as often as statistical models for evaluating small-mammal populations (Fig. 2). Examination of the distribution of samples collected in small-mammal studies strongly suggests that sample-size constraints are the real and unavoidable reason for the use of indices rather than estimators (Fig. 3). Although literature reviews are subjective, our results are consistent with those of others who reviewed estimation methods from earlier time periods (Montgomery 1987; Slade and Blair 2000), and the conclusion that sample size is the primary obstacle to using estimators is echoed by numerous authors of the papers we surveyed (e.g., Tattersall et al. 1997; Ellison and van Riper 1998; Nupp and Swihart 1998; Fernandez et al. 1999; Lewellen and Vessey 1999). We therefore examined the role of indices versus estimators in the context of the sample sizes available for studies of small-mammal populations.

The answer to the question of whether index conversion is prudent lies in the quality of the corrector, knowledge of the bias, and the uses to which the data are to be put. If an unbiased estimate of the true parameter can be obtained, the case for correction is improved. Statistical models produce population estimates with low bias and high precision when the models reflect the underlying population attributes (Otis et al. 1978; Jolly and Dickson 1983; Nichols and Pollock 1983; Efford 1992). Our results indicate that this is true even when sample sizes are small (see also Rosenberg et al. 1995). In contrast, enumeration indices such as M_{t+1} and MNKA are

Fig. 1. Population estimates for the 3 models tested when population attributes conform to the null model (M_0; see Table 1). (*a*) Null model. (*b*) C_{th} model. (*c*) Jackknife model. Probabilities of detection varied from 0.05 to 0.95, four trapping periods were simulated, and $N = 50$ for all simulations.

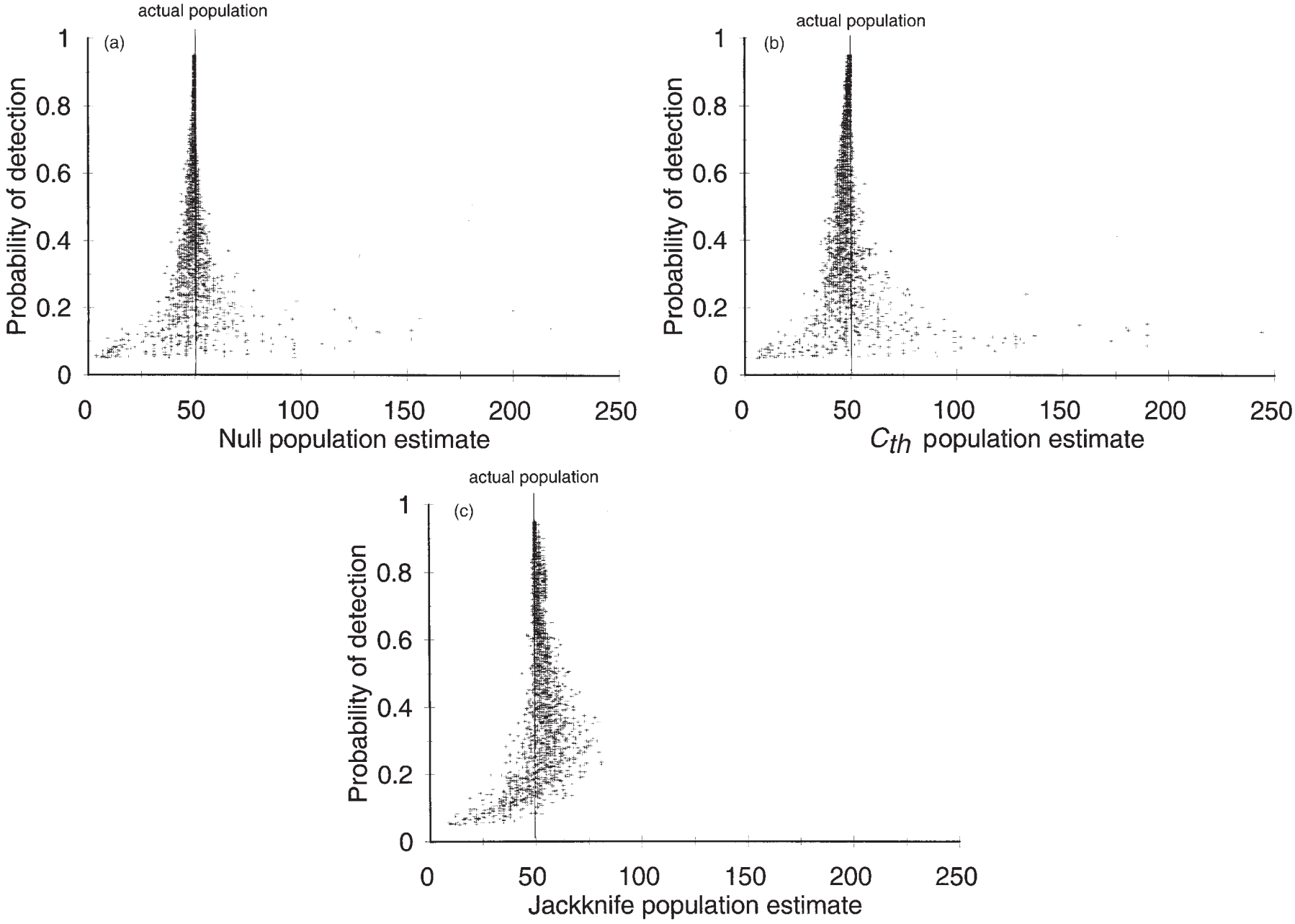


Fig. 2. Methods used to assess small-mammal abundance, based on research published in *Acta Theriologica, Canadian Journal of Zoology, Journal of Mammalogy,* and *The Journal of Wildlife Management* from 1996 through current issues in 2000. The population estimators are Jolly–Seber (Seber 1982), Lincoln–Petersen (Seber 1982), Manly–Parr (cited in Fernandez et al. 1999), and Schumacher–Eschmeyer (cited in Lofgren et al. 1996). Relindex refers to relative abundance indices such as individuals trapped per 100 trap-nights.

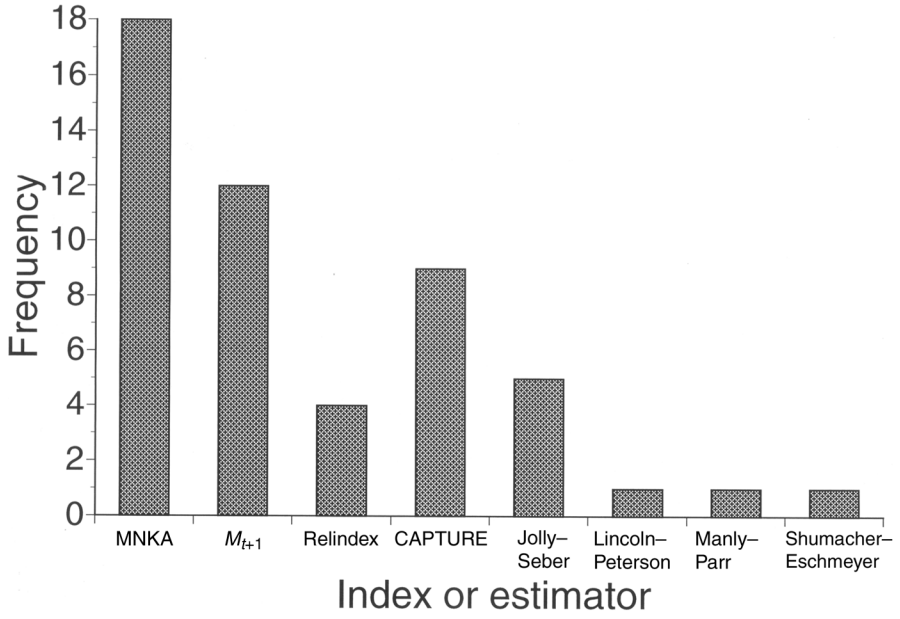
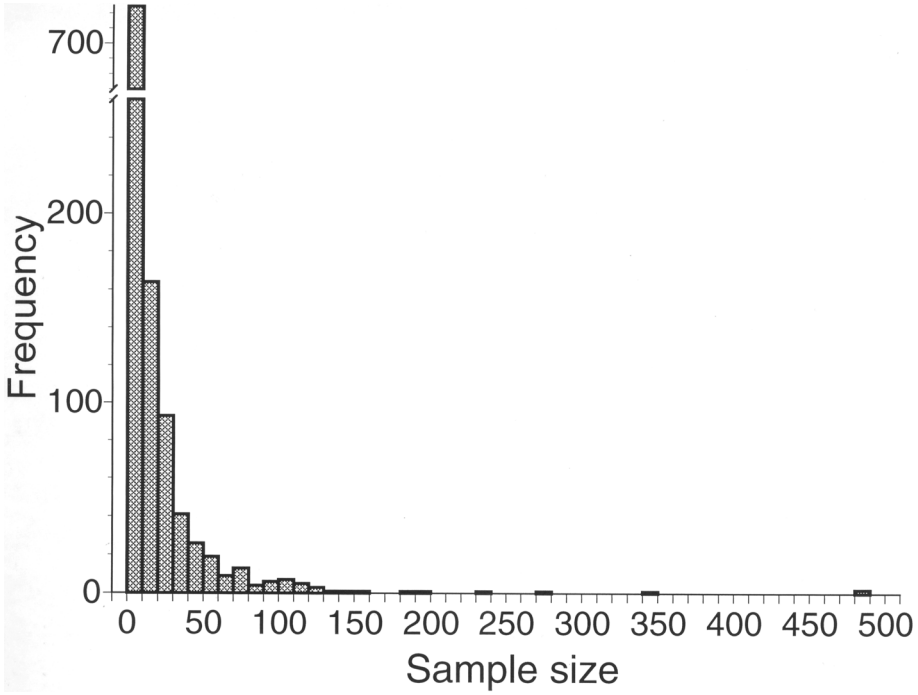


Fig. 3. Numbers of organisms sampled in small-mammal studies published in *Acta Theriologica, Canadian Journal of Zoology, Journal of Mammalogy,* and *The Journal of Wildlife Management* from 1996 through current issues in 2000. Data are limited to studies in which population size was critical to the analyses (time trend or comparative data).



negatively biased and can be misleading if capture probabilities vary greatly over space or time (Nichols and Pollock 1983; Nichols 1986). Moreover, uncorrected indices provide no measure of confidence as to their relationships to population size. As a result, many authors have argued strongly for the use of estimators rather than indices for estimating small-mammal populations (Otis et al. 1978; Jolly and Dickson 1983; Nichols and Pollock 1983; Nichols 1986; Efford 1992; Slade and Blair 2000; Tuyttens 2000).

We emphatically agree that statistical models rather than enumeration indices should be used whenever the underlying population attributes can be inferred or sample sizes are sufficient for effective model selection. However, in capture–recapture studies the sources of variation are rarely known

Fig. 4. Correlation between the tested capture–recapture estimators and M_{t+1}. (*a*) Null model. (*b*) C_{th} model. (*c*) Jackknife model. The M_h population attribute (Table 1) was simulated. Probabilities of detection varied from 0.3 to 0.6, four trapping periods were simulated, and N = 50 for all simulations.

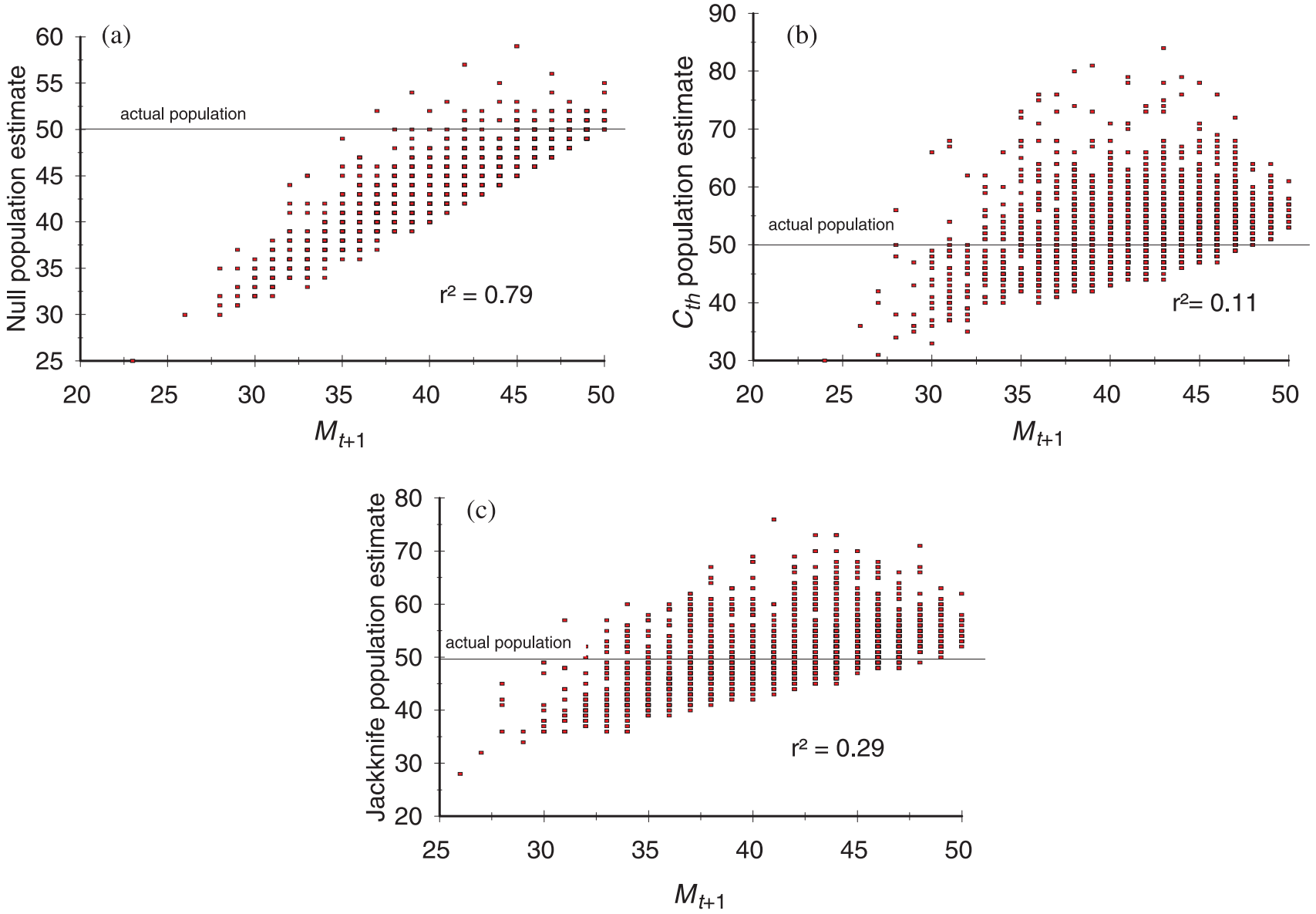


Table 2. Average population estimates (N = 50) made using the 3 tested models and M_{t+1} on the 8 population attributes.

Population attribute	Capture–recapture estimate			
	Null	Jackknife	C_{th}	M_{t+1}
M_0	49.64 (6.77)	55.21 (9.31)	47.95 (10.63)	44.86 (8.14)
M_t	50.29 (7.50)	55.63 (10.20)	48.36 (12.04)	43.72 (10.08)
M_b	91.78 (37.89)	86.13 (9.24)	94.39 (43.53)	44.78 (8.24)
M_h	44.64 (9.01)	50.94 (11.17)	45.15 (11.75)	41.81 (11.34)
M_{bh}	76.45 (21.95)	78.16 (11.67)	80.67 (34.39)	41.91 (10.99)
M_{th}	48.15 (7.79)	54.00 (10.65)	47.29 (12.63)	43.42 (11.34)
M_{tb}	66.18 (18.10)	72.58 (12.04)	65.32 (24.51)	44.69 (8.88)
M_{tbh}	60.89 (19.88)	68.53 (12.69)	60.88 (21.98)	43.32 (10.34)

Note: Numbers in parentheses are percent CV, derived from the simulation results.

Table 3. Results of ordinal ranking simulations.

Population attribute	Estimator/index value			
	Jackknife	C_{th}	Null	M_{t+1}
Variable attributes	0.835	0.772	0.788	0.963
M_h	0.961	0.952	na	0.953
M_0	na	na	0.995	0.987

Note: Values are the proportion of 2000 simulated sample-pairs in which the estimate for the second sample was lower than the estimate for the first sample. Per-session probability of detection was allowed to vary randomly between 0.30 and 0.60 in all simulations; na, not applicable.

and population estimators must be chosen by way of statistical extrapolation from a subset of the data. This process inherently has low power and small samples with low capture probabilities result in poor model selection and erratic model behavior, including unknown bias and poor population estimates with high variability. In contrast, the index M_{t+1} appears quite robust to changes in the underlying sources of variation and exhibits a known negative bias. Of the metrics we examined for small closed populations, M_{t+1} was the most robust to changes in the underlying population attributes. Even though we allowed capture probabilities to vary considerably (0.30–0.60), in most cases the CVs associated with M_{t+1} were the smallest (Table 2).

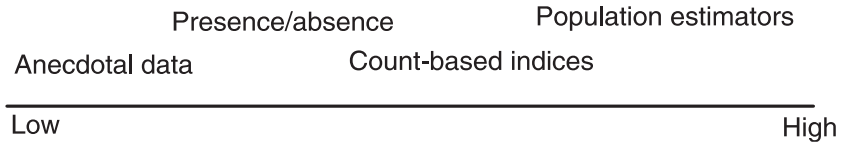
There are several reasons why M_{t+1}, while obviously biased low, showed the least sensitivity to changes in the underlying population attributes. One is that trap-induced behaviors have no impact on M_{t+1} because they have little or no impact on first captures. For M_{t+1} there are essentially two potential confounding factors: individual level heterogeneity and time. Of these two factors, heterogeneity will always cause a decrease in the number of organisms caught. Variation in detection rates across time can affect M_{t+1}, but the random effects that we simulated appeared to have negligible impact. Average M_{t+1} was slightly lower when heterogeneous capture probabilities were simulated (Table 2), but the differences were quite modest compared with the model-to-model variation observed when forcing the 3 estimators. The second reason why M_{t+1} is relatively stable is that it is bounded: $0 \leq M_{t+1} \leq N$. The population estimators, particularly C_{th} and the null estimator, are not bounded by N and can, because of sampling variance and (or) poor model choice, produce extremely errant estimates (Table 2).

Enumeration indices have been criticized as being poor population metrics because they exhibit negative bias and can be misleading when capture probabilities vary (Nichols and Pollock 1983; Nichols 1986). However, we found that M_{t+1} was considerably more robust than the population estimators we examined for detecting changes in small populations when the underlying population attributes were unknown (Table 3). Slade and Blair (2000) observed high correlations ($r^2 > 0.7$) between the indices M_{t+1} and MNKA and estimators for five species of small mammals. They concluded that indices were proportional to estimates, i.e., bias was relatively constant. Such high correlations between capture–recapture indices and estimators have been reported for numerous species (Lefebvre et al. 1982; Hallett et al. 1991; Manning et al. 1995; Morris 1996; Nupp and Swihart 1996; Waters and Zabel 1998). Since bias itself is not problematic for relative comparisons, M_{t+1} can be a valid metric for comparing populations in cases where capture probabilities are relatively constant.

We conclude that the relative merits of estimation and enumeration for comparing small-mammal populations vary with data quality. We define four zones of data quality that determine the potential value of data for population estimation (Fig. 5): an estimation zone of high data confidence where only population estimators should be employed, an index zone of decreased data confidence where appropriate population estimators cannot be reliably chosen, a presence/absence zone where the numerical values of indices cannot be validated but data can still yield reliable presence/absence information, and an anecdotal zone where data are too weak for any numerical analysis. We focus on the index zone because population estimation with large samples has been thoroughly addressed elsewhere and because most small-mammal data appear to fall within the index zone, a fact that has largely been ignored.

The index zone is separated from the estimation zone by the reliability of the model-selection algorithms. The model-selection procedures in CAPTURE are unreliable when applied to small samples (Otis et al. 1978; Menkins and Anderson 1988; Pollock et al. 1990; Manning et al. 1995). Even for moderate samples, $50 < N < 100$, Menkins and Anderson (1988) reported that selection of the appropriate model by CAPTURE was no better than random. When sources of variation associated with time, heterogeneity, and behavior are large, as in our simulations, CAPTURE does

Fig. 5. Heuristic representation of our opinion concerning the best methods for analyzing capture–recapture data, based on the quantity and quality of available data.



somewhat better (M_b was chosen as the primary or alternative model 67% of the time when it was correct and M_h was chosen 50% of the time), but is still unreliable. Given the inability to choose the proper model, the researcher has three choices: use a default estimator, use a relative index, or reject the data.

To argue in favor of using a default estimator one must argue either that the use of any estimator is generally better than using direct enumeration, or that some estimators work better under most conditions of heterogeneity, behavior, and time expected for a specific dataset. In this paper we define better as less bias and a lower CV. Based on these metrics, none of the estimators we tested were better than M_{t+1} when the underlying population attributes were unknown and the associated variation in capture probabilities far from null expectations. We therefore strongly caution against the use of default models under circumstances comparable to those we simulated. However, we do not know how well this understanding compares with the range of real-world small-mammal trapping situations. Manning et al. (1995) tested population estimators and M_{t+1} on known vole populations. For $N \approx 90$ and four trapping sessions, all of the estimators except Lincoln–Petersen produced lower bias and a lower residual sum of squares score (RSS) than M_{t+1} (Manning et al. 1995; M_{t+1} statistics are computed from Table 2), whereas for $N \approx 30$, only 3/11 of the tested models provided lower bias and only 1 had a lower RSS. Manning et al. (1995) chose the jackknife estimator as the best default model for their data. However, the M_h model performed more poorly than M_{t+1} when $N \approx 30$, having high positive bias and a high RSS even though these populations exhibited high capture probabilities (0.42–0.73 per trapping session), high individual heterogeneity, and little evidence of strong time or behavior patterns.

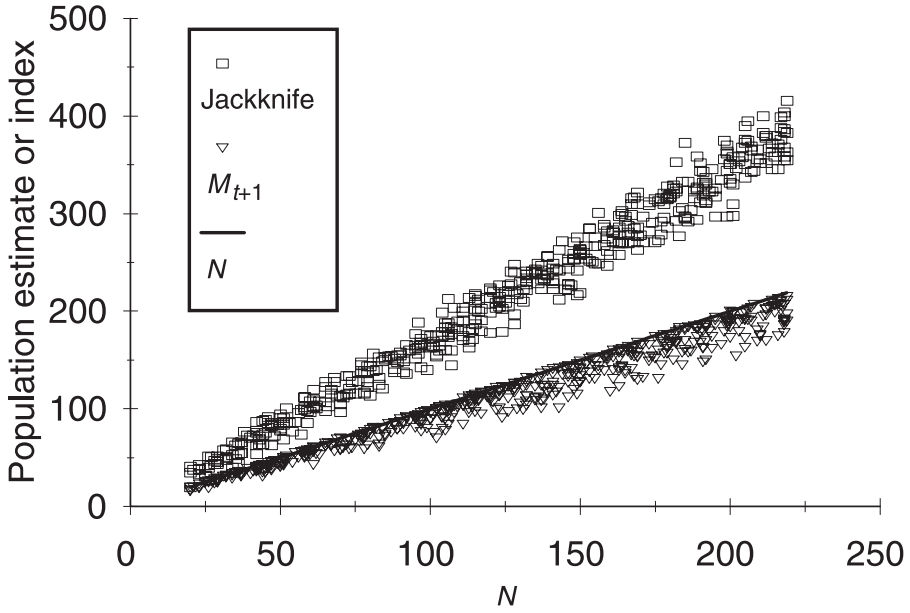
Some researchers might argue that data within this zone should not be used for population studies. However, this response is not very satisfying, given that the bulk of the small-mammal data reside here (Fig. 3). This approach implies that a coin toss provides as much or more information for making management decisions than understanding based on indices, and suggests that we have learned nothing by using abundance indices as surrogates for population estimates. We believe that index-based studies have provided a great deal of reliable information concerning small-mammal habitat use and responses to management. For instance, in the Rocky Mountain region of the United States, southern red-backed voles (*Clethrionomys gapperi*) prefer mature and late-seral forests to recent clearcuts and young forests (Ramirez and Hornocker 1981; Halvorson 1982; Scrivner and Smith 1984; Medin 1986; Hayward and Hayward 1995). This research has not only proved to be repeatable, but the outcome, that clear-cutting reduces densities of red-backed voles, has repeatedly proved to be true (for a review see Pearson 1999). Given the high level of correlation reported when counts and estimators are compared (e.g., Nupp and Swihart 1996; Waters and Zabel 1998; Slade and Blair 2000), it is unlikely that many of the understandings derived from the current index-based mammal literature would change greatly if estimators had been used.

In the index zone we consider M_{t+1} to be the default for comparing small-mammal populations under the assumption of closure. However, indices must be validated with regard to the assumption of constant probability of detection before they can be used for population comparisons. If indices cannot be validated, this is an indication that the quality of the data falls below the index zone and that the data should not be used to compare populations. Because the division between the estimation and index zones is data-specific and not discretely defined, the use of default estimators may sometimes be justified. However, the decision to use a default estimator rather than M_{t+1} should be based on quantifiable understanding which suggests that the chosen model is appropriate. Without this understanding, the use of an estimator and associated confidence intervals can give an impression of accuracy and precision that is generally invalid in this zone. We offer the following advice for validating indices or default estimators in the index zone.

(*i*) Replicate: enumeration indices should never be used without replication. An index derived from a single trapping session contains no measure of precision, and even relative comparisons between single-session indices are very questionable. With repetition, however, estimates of precision can be generated and relative comparisons evaluated.

(*ii*) Screen for differential capture probabilities: even with repetition, there is the possibility of systematic bias when using indices. For example, if capture probabilities were treatment-dependent, the use of indices could lead to spurious conclusions. We believe, therefore, that it is prudent to examine correlations between M_{t+1} and an estimator for indications of differential capture probabilities. High correlations (>0.80) suggest that the index and the estimator are proportional (Slade and Blair 2000), but researchers should be aware that M_{t+1} is generally correlated with estimators when samples are small (Fig. 4). Low correlation between an estimator and M_{t+1} indicates that at least one of the two is behaving poorly, and that caution should be used when interpreting the data. Changes in the regression slope between M_{t+1} and an estimator can also indicate changes in the underlying sources of variation in capture probabilities (Slade and Blair 2000). Although extremely high correlations between M_{t+1} and an estimator leave the researcher indifferent as to which metric is used for relative comparisons, a high correlation does not indicate accuracy (Fig. 6). Therefore, if an estimator is used, confidence intervals should not be pre-

Fig. 6. Correlation between M_{t+1} and the jackknife estimator, given the M_b population attribute (Table 1), four trapping sessions, $20 < N < 220$, and 500 simulations. Although correlation is very high between the estimator and the index ($r^2 = 0.95$), neither accurately reflects the true population.



sented, as they may be invalid. CAPTURE may also be used to estimate capture probabilities among populations. Although this approach is circular in that estimated capture probabilities are model-dependent, differential capture probabilities sufficient to invalidate understandings based on M_{t+1} will likely be detected even with poor model selection. In applying this approach, the model most consistently chosen within each treatment group should be used to estimate capture probabilities within that group.

(*iii*) Increase the number of trapping sessions: increasing the number of trapping sessions decreases the impact of differential probabilities of detection. With a 20% per-session probability of detection, null-model expectations at 4 and 10 trapping sessions are that 59 and 89%, respectively, of the population will be trapped. If the detection probability is 60%, at 4 and 10 sessions 97 and 100% of the population will be trapped. When M_{t+1} is used to compare these populations, bias is 38% at four trapping sessions but decreases to 11% at 10. However, when increasing the number of trapping sessions, researchers should consider that repeated captures deleteriously impact small mammals (e.g., Slade 1991).

(*iv*) Collect additional indices: multiple independent indices can be used to validate an index of abundance to provide further support for its use. In such cases, additional indices should be independently gathered, and not reliant on bait, so as to provide independent evidence for or against the validity of the primary count-based index. Fecal tracking boards (Emlen et al. 1957) or passive sampling of burrow entrances (Boonstra et al. 1992) are potential examples of methods of obtaining such data for small-mammal studies.

(*v*) Validate the use of default estimators within the index zone: a priori knowledge of a species from previous research or external verification of the underlying population attributes via telemetry could be used to justify the use of a default estimator (e.g., Hallett et al. 1991; Manning et al. 1995). In the absence of such information, a researcher must bring other data to bear to justify the use of a default estimator. CAPTURE can offer evidence as to the underlying population attributes under certain circumstances. Because model selection is based on evaluating trends in the recapture data, increasing the number of trapping sessions greatly improves the ability of CAPTURE to identify the underlying population attributes. In the case of our M_h simulation, increasing the number of trap-nights from 4 to 8 improved model selection from 50 to 81% correct. While insufficient for reliably choosing an estimator for a single study, nonrandom model choice in replicated studies can indicate the existence of a source of variability such as heterogeneity.

Clearly, if we could improve model selection, we could shrink the index zone, and the above approaches would yield even greater benefits. MARK (Burnham and Anderson 1998) uses different approaches to model selection, and may improve selection at smaller sample sizes. None of the published papers we reviewed used MARK (White and Burnham 1999) to estimate population size and, to our knowledge, MARK has not been directly compared with CAPTURE to determine whether this is the case. We suggest that additional work be conducted in the areas of model selection, telemetry, and perhaps study design to potentially expand the ability of researchers to validly use population estimators for the smaller samples that are predominant in population studies.

Lastly, we believe that there are a number of common practices which are not valid uses of capture–recapture data in the index zone:

(*i*) Using total number of captures as an index: total number of captures is sometimes used as an index, rather than M_{t+1}. Total captures is a very weak index that correlates poorly with population estimators (Slade and Blair 2000) and emphasizes behavioral differences among animals, especially if the index is used to make comparisons among

species. In general, indices should not be used to compare relative abundance among species. Strong evidence exists to indicate that capture probabilities vary sufficiently among species that indices are likely to fail in cross-species comparisons (Nichols 1986; Slade and Blair 2000).

(*ii*) Correcting enumeration indices according to effort: transforming raw capture–recapture indices into catch per unit effort indices assumes a linear relationship between capture and effort that is unsubstantiated. If trapping effort is so skewed that effort transformations are deemed necessary (Beauvais and Buskirk 1999), then data quality may be insufficient for evaluating populations.

(*iii*) Mixing methods within a study: some researchers have begun to mix estimators and indices within a study by applying indices to smaller samples and estimators to larger ones (Sullivan et al. 1998; Von Trebra et al. 1998; Hanley and Barnard 1999; Slade and Blair 2000). This approach will almost certainly result in differential bias within time trends or between treatments because the index, which is negatively biased, is used for small samples and an estimator, which can be positively biased (Table 2), is used for larger samples. Whether the choice is to use indices or estimators, the chosen method must be applied to all compared data, and the methods of analysis that are appropriate for the weakest compared dataset will determine the methods of analysis for all of the data.

Acknowledgements

We greatly appreciate the comments of Gregory D. Hayward, Mark S. Lindberg, L. Scott Mills, Erin C. O'Doherty, Yvette K. Ortega, and Jeffrey R. Waters on earlier drafts of the manuscript. Additionally, we thank two anonymous reviewers of the manuscript.

References

Beauvais, G.P., and Buskirk, S.W. 1999. Modifying estimates of sampling effort to account for sprung traps. Wildl. Soc. Bull. **27**: 39–43.

Boulanger, J.G., and Krebs, C.J. 1994. Comparison of capture–recapture estimators of snowshoe hare populations. Can. J. Zool. **72**: 1800–1807.

Boulanger, J.G., and Krebs, C.J. 1996. Robustness of capture–recapture estimators to sample bias in a cyclic snowshoe hare population. J. Appl. Ecol. **33**: 530–542.

Boonstra, R., Kanter, M., and Krebs, C.J. 1992. A tracking technique to locate small mammals at low densities. J. Mammal. **73**: 683–685.

Brown, J.K., and See, T.E. 1981. Downed, dead woody fuel and biomass in the northern Rocky Mountains. U.S. For. Serv. Gen. Tech. Rep. INT-117.

Bull, E.L., Parks, C.G., and Torgensen, T.R. 1997. Trees and logs important to wildlife in the interior Columbia River basin. U.S. For. Serv. Gen. Tech. Rep. PNW-391.

Burnham, K.P., and Anderson, D.R. 1998. Model selection and inference: a practical information–theoretic approach. Springer-Verlag, New York.

Burnham, K.P., and Overton, W.S. 1978. Estimation of the size of a closed population when capture probabilities vary among animals. Biometrika, **65**: 625–633.

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

Chao, A. 1988. Estimating animal abundance with capture frequency data. J. Wildl. Manag. **52**: 295–300.

Chao, A., Lee, S.-M., and Jeng, S.-L. 1992. Estimating population size for capture–recapture data when capture probabilities vary by time and individual. Biometrics, **48**: 201–216.

Efford, M. 1992. Comment—Revised estimates of the bias in the "minimum number alive" estimator. Can. J. Zool. **70**: 628–631.

Ellison, L.E., and van Riper, C., III. 1998. A comparison of small mammal communities in a desert riparian floodplain. J. Mammal. **79**: 972–985.

Emlen, J.T., Hine, R.L., Fuller, W.A., and Alfonso, P. 1957. Dropping boards for population studies of small mammals. J. Wildl. Manag. **21**: 300–314.

Fernandez, F.A.S., Dunstone, N., and Evans, P.R. 1999. Density-dependence in habitat utilization by wood mice in a Sitka spruce successional mosaic: the roles of immigration, emigration, and variation among local demographic parameters. Can. J. Zool. **77**: 97–405.

Hallett, J.G., O'Connell, M.A., Sanders, G.O., and Seidensticker, J. 1991. Comparison of population estimators for medium-sized mammals. J. Wildl. Manag. **55**: 81–93.

Halvorson, C.H. 1982. Rodent occurrence, habitat disturbance, and seed fall in a larch–fir forest. Ecology, **63**: 423–433.

Hanley, T.A., and Barnard, J.C. 1999. Spatial variation in population dynamics of Sitka mice in floodplain forests. J. Mammal. **80**: 866–879.

Hayward, G.D., and Hayward, P.H. 1995. Relative abundance and habitat associations of small mammals in Chamberlain Basin, central Idaho. Northwest Sci. **69**: 114–124.

Hilborn, R., Redfield, J.A., and Krebs, C.J. 1976. On the reliability of enumeration for mark and recapture census of voles. Can. J. Zool. **54**: 1019–1024.

Jolly, G.M., and Dickson, J.M. 1983. The problem of unequal catchability in mark–recapture estimation of small mammal populations. Can. J. Zool. **61**: 922–927.

Krebs, C.J. 1966. Demographic changes in fluctuating populations of *Microtus californicus*. Ecol. Monogr. **36**: 239–273.

Lee, S.-M., and Chao, A. 1994. Estimating population size via sample coverage for closed capture–recapture models. Biometrics, **50**: 88–97.

Lefebvre, L.W., Otis, D.L., and Holler, N.R. 1982. Comparison of open and closed models for cotton rat population estimates. J. Wildl. Manag. **46**: 156–163.

Lewellen, R.H., and Vessey, S.H. 1999. Estimating densities of *Peromyscus leucopus* using live-trap and nestbox censuses. J. Mammal. **80**: 400–409.

Lofgren, O., Hornfeld, B., and Eklund, U. 1996. Effects of supplemental food on a cyclic *Clethrionomys glareolus* population at peak density. Acta Theriol. **41**: 383–394.

Manning, T., Edge, W.D., and Wolff, J.O. 1995. Evaluating population-size estimators: an empirical approach. J. Mammal. **76**: 1149–1158.

Medin, D.E. 1986. Small mammal response to diameter-cut logging in an Idaho Douglas-fir forest. U.S. For. Serv. Res. Pap. INT-362.

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

Montgomery, W.I. 1987. The application of capture–mark–recapture methods to the enumeration of small mammal populations. Symp. Zool. Soc. Lond. No. 58. pp. 25–57.

Morris, D.W. 1996. Coexistence of specialist and generalist rodents via habitat selection. Ecology **77**: 2352–2364.

Nichols, J.D. 1986. On the use of enumeration estimators for interspecific comparisons, with comments on a "trappability" estimator. J. Mammal. **67**: 590–593.

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

Nupp, T.E., and Swihart, R.K. 1996. Effect of forest patch area on population attributes of whitefooted mice (*Peromyscus leucopus*) in fragmented landscapes. Can. J. Zool. **74**: 467–472.

Nupp, T.E., and Swihart, R.K. 1998. Effects of forest fragmentation on population attributes of white-footed mice and eastern chipmunks. J. Mammal. **79**: 1234–1243.

Otis, D.L., Burnham, K.P., White, G.C., and Anderson, D.R. 1978. Statistical inference from capture data on closed animal populations. Wildl. Monogr. No. 62.

Pearson, D.E. 1999. Small mammals of the Bitterroot National Forest: a literature review and annotated bibliography. U.S. For. Serv. Gen. Tech. Rep. RMRS-GTR-25.

Pollock, K.H., Nichols, J.D., Brownie, C., and Hines, J.E. 1990. Statistical inference for capture–recapture experiments. Wildl. Monogr. No. 107.

Ramirez, P., Jr., and Hornocker, M. 1981. Small mammal populations indifferent-aged clearcuts in northwest Montana. J. Mammal. **62**: 400–403.

Rexstad, E., and Burnham, K.P. 1991. User's guide for interactive program CAPTURE: abundance estimation of closed animal populations. Colorado Cooperative Fish and Wildlife Research Unit, Colorado State University, Fort Collins.

Rosenberg, D.K., Overton, W.S., and Anthony, R.G. 1995. Estimation of animal abundance when capture probabilities are low and heterogeneous. J. Wildl. Manag. **59**: 252–261.

Seber, G.A.F. 1982. The estimation of animal abundance and related parameters. 2nd ed. MacMillan, New York.

Scrivner, J.H., and Smith, H.D. 1984. Relative abundance of small mammals in four successional stages of spruce fir in Idaho. Northwest Sci. **53**: 171–176.

Slade, N.A. 1991. Loss of body mass associated with capture of *Sigmodon* and *Microtus* from northeastern Kansas. J. Mammal. **72**: 171–176.

Slade, N.A., and Blair, S.M. 2000. An empirical test of using counts of individuals captured as indices of population size. J. Mammal. **81**: 1035–1045.

Sullivan, T.P., Nowotny, C., Lautenschlager, R.A., and Wagner, R.G. 1998. Silvicultural use of herbicide in sub-boreal spruce forest: implications for small mammal population dynamics. J. Wildl. Manag. **62**: 1196–1206.

Tattersall, F.H., Macdonald, D.W., Manley, W.J., Gates, S., Feber, R., and Hart, B.J. 1997. Small mammals on one-year set-aside. Acta Theriol. **42**: 329–334.

Tuyttens, F.A.M. 2000. The closed-subpopulation method and estimation of population size from mark–recapture and ancillary data. Can. J. Zool. **78**: 320–326.

Von Trebra, C., Lavender, D.P., and Sullivan, T.P. 1998. Relations of small mammal populations to even-aged shelterwood systems in sub-boreal spruce forest. J. Wildl. Manag. **62**: 630–642.

Waters, J.R., and Zabel, C.J. 1998. Abundances of small mammals in fir forests in northeastern California. J. Mammal. **79**: 1244–1253.

White, G.C., and Burnham, K.P. 1999. Program MARK: survival estimation from populations of marked animals. Bird Study, **46**(Suppl.): 120–138.

White, G.C., Anderson, D.R., Burnham, K.P., and Otis, D.L. 1982. Capture–recapture and removal methods for sampling closed populations. Rep. LA-8787-NERP, Los Alamos National Laboratory, Los Alamos, N. Mex.