A Capture-Recapture Design Robust to Unequal Probability of Capture
Author(s): Kenneth H. Pollock
Source: *The Journal of Wildlife Management*, Vol. 46, No. 3 (Jul., 1982), pp. 752-757
Published by: Wiley on behalf of the Wildlife Society
Stable URL: https://www.jstor.org/stable/3808568
Accessed: 11-07-2019 22:39 UTC

SHORT COMMUNICATIONS

A CAPTURE-RECAPTURE DESIGN ROBUST TO UNEQUAL PROBABILITY OF CAPTURE

KENNETH H. POLLOCK, Department of Statistics, North Carolina State University, Box 5457, Raleigh, NC 27650.

This author believes that the design of capture-recapture experiments deserves more attention from statisticians and biologists. Study design should be oriented around satisfaction of as many model assumptions as practically possible so that a simple and reasonably efficient model can be used for estimation.

At the basis of many capture-recapture sampling models is the assumption that all animals are equally likely to be caught in each sample (The Equal Catchability Assumption). This assumption is often violated in wildlife populations (Seber 1973:81) and 2 general types of alternatives exist (Pollock 1981):

(1) Heterogeneity: The probability of capture in any sample is a property of the animal and may vary over the population. That is, animals may vary in capture probabilities according to age, sex, social status, and many other factors.

(2) Trap response: The probability of capture in any sample depends on the animal's prior history of capture. That is, animals may become "trap shy" or "trap happy" depending upon the type of trapping method used. Either 1 or both of these 2 types of alternatives may be acting in a particular animal population.

The traditional capture-recapture model used by biologists for closed populations (populations closed to additions or deletions) in short-term studies is the Schnabel Model (Schnabel 1938) that requires The Equal Catchability Assumption. In recent years there has been substantial research on models for closed populations that allow heterogeneity and/or trap response of the capture probabilities. Otis et al. (1978) published an important monograph on these models that allows their routine use by biologists.
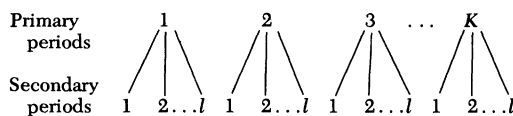
The capture-recapture model becoming used by biologists for open populations in long-term studies is the Jolly-Seber Model (Seber 1973). This model requires The Equal Catchability Assumption and the complexity of open population models is likely to preclude general models that allow heterogeneity and/or trap response.

During the preparation of a review of capture-recapture methods (Pollock 1981), I realized that statisticians have drawn a sharp distinction between closed and open population models that is perhaps rather artificial. Here I describe a design for long-term studies that is robust to heterogeneity and/or trap response. It allows an analysis that uses methodology from closed and open population models. There is a brief examination of its robustness properties using simulation and an example is given in detail to illustrate the methodology for biologists.

THE DESIGN

Description

Consider the following representation of a capture-recapture sampling experiment:

Primary periods: $1 \quad 2 \quad 3 \ldots K$

Secondary periods: $1 \quad 2 \ldots l \quad 1 \quad 2 \ldots l \quad 1 \quad 2 \ldots l \quad 1 \quad 2 \ldots l$

where we have K primary sampling pe-

riods (for example, years) and within each 1 of these we have l secondary sampling periods that are close to each other in time (for example l consecutive days of trapping).

The biologist is interested in the population size for each of the primary sampling periods $(N_1, N_2, \ldots, N_K)$ assuming that the population is constant over the secondary sampling periods within a primary sampling period. He will also be interested in survival and birth rates between the primary sampling periods $(\phi_1, \phi_2, \ldots, \phi_{K-1}$ and $B_1, B_2, \ldots, B_{K-1})$.

Estimation of Population Parameters

Assuming that the population is approximately closed over the secondary sampling periods within a primary period, then 2 estimation procedures are possible.

Jolly-Seber.—Under this procedure all the secondary sampling periods within a primary sampling period would be "pooled." By "pooled" I mean that we are just interested if an animal is captured at least once or uncaptured in the primary sampling period. Arnason and Baniuk (1978) also suggested pooling for the purpose of increasing precision when capture probabilities are low. Thus, the Jolly-Seber estimates of population sizes $(N_1, \ldots, N_K)$, survival rates $(\phi_1, \ldots, \phi_{K-1})$, and birth rates $(B_1, \ldots, B_{K-1})$ would be calculated using the pooled data.

Cormack (1973) and Seber (1973:196) have described these estimators in detail. Suppose to begin with that M_i, the number of marked animals in the population just before the ith sample, is known for all values $i = 2, \ldots, K$ (there are no animals marked at the time of the first sample, so $M_1 = 0$). An estimator of N_i, the

population size at time i, is the Petersen estimator

$$\hat{N}_i = n_i M_i / m_i, \qquad (1)$$

which is based on equating the ratios of marked-to-total animals captured in the ith sample (m_i/n_i) to the ratio in the population (M_i/N_i). An estimator of the survival rate from sample i to sample $i + 1$ is

$$\hat{\phi}_i = M_{i+1}/(M_i - m_i + R_i), \qquad (2)$$

where R_i is the number of the n_i animals captured in the ith sample that are released. All unmarked animals are marked before release. An estimator of the recruitment from time i to $i + 1$ is

$$\hat{B}_i = \hat{N}_{i+1} - \hat{\phi}_i(\hat{N}_i - n_i + R_i), \qquad (3)$$

which is simply the estimated difference between the population size at time $i + 1$ and the expected number of survivors from i to $i + 1$).

To complete this outline we need an estimator of the number of marked animals at each sample time (M_i) that are unknown parameters in an open population. This can be obtained by equating the 2 ratios

$$Z_i/(M_i - m_i) \simeq r_i/R_i, \qquad (4)$$

which are the future recovery rates of 2 distinct groups of marked animals. $(M_i - m_i)$ are the marked animals not captured at i, and R_i are the animals captured at i and released for possible recapture. Note that Z_i and r_i are the members of $(M_i - m_i)$ and R_i, that are recaptured again at least once. The estimator of M_i is thus given by

$$\hat{M}_i = m_i + R_i Z_i / r_i. \qquad (5)$$

It should be emphasized that $\hat{M}_i$ is only defined for $i = 2, \ldots, K - 1$ and thus $\hat{N}_i$

in eq. (1) is only defined for $i = 2, \ldots,$ $K - 1$; $\hat{\phi}_i$ in eq. (2) for $i = 1, \ldots, K - 2$; $\hat{B}_i$ in eq. (3) for $i = 2, \ldots, K - 2$. Approximate variances and bias adjustment formulae have been presented by Seber (1973:202).

Heterogeneity and/or trap response will have a large effect on the population size estimators (eq. (1)) because the sample ratio (m_i/n_i) will no longer accurately reflect the population ratio (M_i/N_i). Also, the marked population estimators (eq. (5)) will not be so affected by unequal catchability because the 2 ratios in eq. (4) will both tend to be influenced similarly. The survival rate estimators (eq. (2)) that are simply ratios of the M_i's will also be less affected by unequal catchability than the population size estimators. Cormack (1972) suggested this using an intuitive argument and Carothers (1973) documented it using a simulation for populations where heterogeneity of capture probabilities occurs.

Modified Procedure.—I devised a procedure that attempts to minimize the influence of unequal catchability on the estimators by exploiting the 2 levels of sampling. Because survival rate estimators are not so affected, they will be estimated exactly as in the Jolly-Seber procedure that assumes equal catchability. These survival estimators will only be available for $i = 1, 2, \ldots, K - 2$.

It is suggested that population size estimators for each primary sampling period $(N_1, \ldots, N_K)$ be obtained using closed population models allowing for unequal catchability and based only on the captures and recaptures within a primary sampling period. The easiest way for a biologist to obtain these estimates is to use program CAPTURE developed by Otis et al. (1978). It considers a range of different models allowing for heteroge-

neity and/or trap response and also gives an objective method of choosing the appropriate model to use. Here estimators are available for all primary periods $(i = 1, \ldots, K)$ whereas under the Jolly-Seber approach estimators are only available for $i = 2, \ldots, K - 1$.

The birth rate estimators can be obtained from eq. (3), but now the population size estimators used are those described above. It is possible to estimate $B_1, \ldots, B_{K-2}$ whereas under Jolly-Seber it is only possible to estimate $B_2, \ldots, B_{K-2}$. If we had an estimator of ϕ_{K-1} then B_{K-1} would also be estimable. In some experiments it may be reasonable to estimate $\hat{\phi}_{K-1}$ by $\hat{\phi}_{K-2}$ or perhaps by an average of $\hat{\phi}_1, \ldots, \hat{\phi}_{K-2}$.

The approximate variances of the $\hat{\phi}_i$ and $\hat{N}_i$ estimators are available (Seber 1973, Otis et al. 1978). The variances of the $\hat{B}_i$ are given by

$$\begin{aligned}
\mathrm{var}(\hat{B}_i) = {}&\mathrm{var}(\hat{N}_{i+1}) + \phi_i{}^2\mathrm{var}(\hat{N}_i) \\
&+ N_i{}^2\mathrm{var}(\hat{\phi}_i) + \mathrm{var}(\hat{\phi}_i)\mathrm{var}(\hat{N}_i) \\
&+ (n_i - R_i)^2\mathrm{var}(\hat{\phi}_i).
\end{aligned} \quad (6)$$

Simulation

A small simulation study is presented to illustrate the robust properties of the modified procedure. Let us consider a population that is sampled for 3 years and within each year there are 5 sampling periods close enough together so that closure can be assumed. Let us assume that the population is subject to heterogeneity alone (Model M_h of Otis et al. 1978) and that each animal has the same capture probability for all 15 (3 × 5) sampling periods. The 7 trials we consider are from Otis et al. (1978:128) and have differing degrees of heterogeneity (Table 1).

For each trial 2 different birth and death processes acting on the population are considered:

Table 1. Description of trials of Model M_h (Heterogeneity Model) used in the simulation study.

Trial	N	$P_j, j = 1, 2, \ldots, N$
1	400	$P_j = 0.05, j = 1,200; P_j = 0.15, j = 201,300; P_j = 0.50, j = 301,400$[a]
2	400	$P_j = 0.01, j = 1,100; P_j = 0.05, j = 101,200; P_j = 0.10, j = 201,300; P_j = 0.20, j = 301,400$
3	400	$P_j = 0.10, j = 1,100; P_j = 0.20, j = 101,200; P_j = 0.25, j = 201,300; P_j = 0.30, j = 301,400$
4	400	$P_j = 0.01, j = 1,50; \ \ P_j = 0.15, j = 51,200; P_j = 0.25, j = 201,300; P_j = 0.30, j = 301,400$
5	400	$P_j = 0.20, j = 1,100; P_j = 0.30, j = 101,200; P_j = 0.40, j = 201,300; P_j = 0.50, j = 301,400$
8	100	$P_j = 0.05, j = 1,40; \ \ P_j = 0.10, j = \ \ 41,80; \ \ P_j = 0.30, j = \ \ 81,100$
12	100	$P_j = 0.10, j = 1,40; \ \ P_j = 0.20, j = \ \ 41,80; \ \ P_j = 0.30, j = \ \ 81,100$

[a] In this trial there are 400 animals. The first 200 have a capture probability of 0.05, the next 100 a capture probability of 0.15, and the final 100 a capture probability of 0.50 on each sampling occasion.

1. There are no births or deaths so that the population stays constant from year to year.

2. There is a survival rate of 0.5 for all animals in the population and births exactly equal deaths to keep the population constant in size from year to year.

Using the simulation procedure of PO-PAN-2 (Arnason and Baniuk 1978), approximate means and standard errors were obtained for $\hat{N}_2$, the Jolly-Seber estimator given by eq. (1). These values were each based on 99 simulation runs.

If we consider the 5 sampling periods within Year 2, the appropriate estimate of N_2 is the jackknife estimator $(\hat{N}_h)$, and Otis et al. (1978) have derived approximate means and standard errors for our trials using simulation. Comparison of the 2 estimators reveals that when the population has no births and deaths, the

Jolly-Seber estimates have smaller standard errors than the jackknife estimators (Table 2). In the more realistic case when survival is 0.5 and births equal deaths, the standard errors of both estimators are similar. Under both types of birth and death process the Jolly-Seber estimators typically have a much larger bias than the jackknife estimators.

EXAMPLE

Data from a capture-recapture study on a population of American alligator (*Alligator mississippiensis*) at Lake Ellis Simon, North Carolina between 1976 and 1979 (Fuller 1981) were used to illustrate that the suggested methodology has advantages.

Data on between-year captures and recaptures (Table 3) were used to obtain the Jolly-Seber estimates of population size in Year 2 (1977) and Year 3 (1978)

Table 2. Comparison of jackknife and Jolly-Seber estimates in the simulation study.

Trial	N	$\hat{N}_h$[a]	SE	$\hat{N}_2$[b]	SE	$\hat{N}_2$[c]	SE
1	400	331.06	23.99	250.99	13.22	244.38	22.96
2	400	298.06	24.52	257.08	26.23	258.14	52.74
3	400	461.22	32.64	367.06	12.44	369.04	35.79
4	400	417.08	30.91	338.07	12.07	338.27	29.19
5	400	443.95	19.71	386.89	4.94	387.83	17.81
8	100	84.66	14.32	70.17	10.49	72.93	30.74
12	100	107.57	16.80	89.98	7.83	84.13	16.54

[a] Jackknife estimate used for Model M_h (Heterogeneity Model) by Otis et al. (1978).
[b] Jolly-Seber estimate when the survival rate (ϕ) is 1 and there are no births so that the population is constant from year to year.
[c] Jolly-Seber estimate when the survival rate (ϕ) is 0.5 and births balance deaths to keep the population constant from year to year.

J. Wildl. Manage. 46(3):1982

Table 3. Alligator capture histories for between year captures.

1976	1977	1978	1979
$n_1{}^a = 20$	$n_2 = 51$	$n_3 = 41$	$n_4 = 25$
$R_1{}^b = 17$	$R_2 = 49$	$R_3 = 37$	$R_4 = 25$
	$X_{11} = 9$	$X_{111}{}^c = 4$	$X_{1111} = 1$
	$X_{01} = 42$	$X_{101} = 1$	$X_{1011} = 1$
		$X_{011} = 18$	$X_{0111} = 6$
		$X_{001} = 18$	$X_{0011} = 5$
			$X_{1001} = 0$
			$X_{1101} = 2$
			$X_{0101} = 4$
			$X_{0001} = 5$

[a] $n_1 = 20$, the number of alligators captured in Year 1.
[b] $R_1 = 17$, the number of alligators released in Year 1.
[c] $X_{111} = 4$, the number of alligators first captured in Year 1 that were recaptured Year 2 and again in Year 3.

and survival estimates. In 1978 there were enough within-year captures and recaptures to apply the closed population models of Otis et al. (1978). The model selection procedure in CAPTURE chose Model M_h, the heterogeneity model and hence the jackknife estimator $\hat{N}_h$ was computed. The Jolly-Seber population estimates for 1978 and 1977 were compared with the jackknife estimator for 1978 (Table 4). The jackknife estimator is double the Jolly-Seber estimate, which is consistent with the hypothesis that there is strong heterogeneity of capture probabilities for this population. Fuller (1981) also suggests that the jackknife estimate is more consistent with other independent estimates of the population size based on other sampling techniques.

Lefebvre et al. (1982) in a study of hispid cotton rat (*Sigmodon hispidus*) populations did a similar comparison of a closed population model estimate and the Jolly-Seber estimate over a series of

Table 4. Comparison of estimators for the alligator data.

Estimator, year	Estimate	SE	95% CI
Jolly-Seber, 1977	56	6.7	43–69
Jolly-Seber, 1978	69	12.6	44–94
Jackknife, 1978	140	28.5	84–197

months. Their results showed the estimates to be similar.

DISCUSSION

The example and simulation study suggest that the design and modified estimation procedure given in this paper should have practical importance. The major practical difficulty in using this design is its large size. Otis et al. (1978) suggest that a closed population model requires 5–10 sampling periods with average capture probabilities of at least 0.1 per period for reasonable results. Thus, the smallest practical design would be 3 primary periods each containing 5 secondary sampling periods. If, however, this is the design necessary for robust estimation of population parameters it is important for the biologist to be aware of it. In the past there have been too many inadequate studies because statisticians have not given biologists good design guidelines.

Acknowledgment.—I thank M. Fuller for use of the alligator data.

LITERATURE CITED

ARNASON, A. N., AND L. BANIUK. 1978. POPAN-2, a data maintenance and analysis system for recapture data. Charles Babbage Res. Cent., St. Pierre, Manitoba. 269pp.

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

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

———. 1973. Commonsense estimates from capture-recapture studies. Pages 225–254 in M. S. Bartlett and R. W. Hiorns, eds. The mathematical theory of the dynamics of biological populations. Academic Press, London, U.K.

FULLER, M. 1981. Characteristics of an American alligator (*Alligator mississippiensis*) population in the vicinity of Lake Ellis Simon. M.S. Thesis. North Carolina State Univ., Raleigh. 110pp.

LEFEBVRE, L. W., D. L. OTIS, AND N. R. HOLLER. 1982. Comparison of open and closed models for cotton rat population estimates. J. Wildl. Manage. 46:156–163.

OTIS, D. L., K. P. BURNHAM, G. C. WHITE, AND D.
    R. ANDERSON. 1978. Statistical inference for
    capture data from closed populations. Wildl.
    Monogr. 62. 135pp.
POLLOCK, K. H. 1981. Capture-recapture models:
    a review of current methods, assumptions and
    experimental design. Pages 426–435 in C. J.
    Ralph and J. M. Scott, eds. Estimating numbers
    of terrestrial birds. Stud. Avian Biol. 6.

SCHNABEL, Z. E. 1938. The estimation of the total
    fish population of a lake. Am. Math. Mon.
    45:348–352.
SEBER, G. A. F. 1973. The estimation of animal
    abundance and related parameters. Griffin,
    London, U.K. 506pp.

THE METHOD OF BOUNDED COUNTS: WHEN DOES IT WORK?

R. D. ROUTLEDGE, Department of Mathematics, Simon
Fraser University, Burnaby, British Columbia V5A 1S6,
Canada.

The problem addressed by the method of bounded counts, also called the outer-bound method by Giles (1978:44–45), is as follows: an accurate estimate is required of the number of individuals in a population such as brown bears (*Ursus arctos*) in a subarctic drainage basin. Repeated independent efforts are made to census the population. In any given census, the observers will not likely detect all individuals. Erickson and Siniff (1963) reported that the proportion of bears observed in aerial surveys of their Alaskan study area rarely exceeded 50% on any given survey. Overton (1971:426), Seber (1973:58), and Giles (1978:44–45), following Regier and Robson (1966), have proposed using the following method of bounded counts to construct an estimate of population size based solely on the numbers observed in m repeated, incomplete counts.

Let N represent the (unknown) population size, and let $X_{(1)} \le X_{(2)} \le \ldots \le X_{(m)}$ represent the numbers observed, arranged in increasing order. One might use $X_{(m)}$, the largest number seen, as an estimate of N. However, as long as no individual can be counted twice in any given survey, $X_{(m)}$ cannot exceed N. In contrast, $X_{(m)}$ may frequently be much smaller than N. Thus it is a biased estimate of N. Robson and Whitlock (1964)

proposed that the bias be reduced by increasing $X_{(m)}$ to $\hat{N} = X_{(m)} + [X_{(m)} - X_{(m-1)}]$. (Their original proposal was presented in the context of estimating a truncation point of a distribution.)

$\hat{N}$ would be an unbiased estimator of N if the X's (before ordering) were independently distributed according to the uniform distribution on the interval from zero to N. In this instance, the expected positions of the m ordered values would divide the interval into m equal parts. Thus the expected distances between $X_{(m)}$ and N, and between $X_{(m-1)}$ and $X_{(m)}$ would be equal. Adding to $X_{(m)}$, its distance from $X_{(m-1)}$, would remove the bias. Robson and Whitlock (1964) prove (given technical restrictions) that even if all possible values of X are not equally likely, the bias in $\hat{N}$ will be negligible if m is suitably large. They also provide a formula for $(1 - \alpha) \times 100\%$ confidence limits on N. The lower limit is $N_L = X_{(m)}$, and the upper is $N_U = X_{(m)} + [X_{(m)} - X_{(m-1)}] \times (1 - \alpha)/\alpha$. The actual probability that these limits should contain N is exactly $(1 - \alpha)$ if the X's are uniformly distributed. Otherwise, the probability can be guaranteed (given technical restrictions) to be close to $(1 - \alpha)$ only if m is sufficiently large.

ANALYSIS

How large must m be? The upper practical limit on m is usually about 5 or 10.