Environmental and Ecological Statistics **6**, 197–209 (1999)

A closure test for time-specific capture-recapture data

THOMAS R. STANLEY[1] and KENNETH P. BURNHAM[2]

[1]*U.S. Geological Survey, Midcontinent Ecological Science Center, 4512 McMurry Avenue, Fort Collins, Colorado 80525, U.S.A.*
[2]*Colorado Cooperative Fish and Wildlife Research Unit, Colorado State University, Fort Collins, Colorado 80523, U.S.A.*

Received July 1988; revised November 1998

The assumption of demographic closure in the analysis of capture-recapture data under closed-population models is of fundamental importance. Yet, little progress has been made in the development of omnibus tests of the closure assumption. We present a closure test for time-specific data that, in principle, tests the null hypothesis of closed-population model M_t against the open-population Jolly-Seber model as a specific alternative. This test is chi-square, and can be decomposed into informative components that can be interpreted to determine the nature of closure violations. The test is most sensitive to permanent emigration and least sensitive to temporary emigration, and is of intermediate sensitivity to permanent or temporary immigration. This test is a versatile tool for testing the assumption of demographic closure in the analysis of capture-recapture data.

Keywords: Darroch's model, Jolly-Seber model, model M_t, population estimation

1352-8505 © 1999 Kluwer Academic Publishers

1. Introduction

The assumption of demographic closure in capture-recapture studies (i.e., emigration, immigration, mortality or recruitment do not occur during the period of study) is of fundamental importance in the analysis of capture-recapture data under closed-population models for estimating population size (Otis *et al.*, 1978; Seber, 1982, 1992). Despite the importance of this assumption, little progress has been made in the development of an omnibus test for demographic closure. This is largely due to the fact that behavioral variation in capture probabilities is indistinguishable from failure of demographic or geographic closure (Otis *et al.*, 1978, p. 66; White *et al.*, 1982, p. 96).

Nonetheless, a number of partial tests of the closure assumption have been developed for capture-recapture data. For the open-population Jolly-Seber model, Balser (unpublished Ph.D. thesis, Cornell University, 1981) provided a test for temporary emigration that is asymptotically equivalent to the test suggested by Jolly (1982). For closed-population models, data on size distribution (e.g., length or weight) can be used in a Wilcoxon rank-sum test to test whether recruitment of young into a population has occurred, and this test can be extended to estimate the amount of recruitment that has occurred, thereby allowing recruits to be culled from the sample

1352-8505 © 1999 Kluwer Academic Publishers

(Robson and Flick, 1965). Robson and Regier (1968) describe a different, simpler test for recruitment based on contingency tables.

Probably the most widely used test for closure is one given by Otis *et al.* (1978). This test is implemented in CAPTURE, a computer program for analyzing closed-population capture-recapture data (White *et al.*, 1982), and is unaffected by heterogeneity in capture probabilities. However, Type I error rates for this test exceed nominal rates in the presence of time or behavioral variation in capture probabilities (White *et al.*, 1982), and the test is insensitive to temporary violation of closure occurring during the middle of the study (Otis *et al.*, 1978).

Pollock *et al.* (1974), using a general approach, developed a set of three tests for detecting mortality or recruitment in the presence of time variation in capture probabilities. Specifically, they present a test for recruitment allowing mortality, a test for mortality assuming no recruitment, and a test for mortality allowing recruitment. The test statistic for mortality assuming no recruitment is computationally intensive, so Pollock *et al.* (1974) derive a simplified (though less powerful) statistic based on asymptotic and distributional approximations. A test statistic for recruitment assuming no mortality was not presented, though they note one can be derived, and an overall test statistic for mortality and recruitment (i.e., closure) was not given.

In this paper we present a new test for closure developed under a null model allowing for time variation in capture probabilities that, in principle, builds on the work of Pollock *et al.* (1974). This test is chi-square, can be partitioned into informative components, and is a subset of the goodness-of-fit test for closed-population capture-recapture model M_t developed by Stanley and Burnham (1999). The power of the test in the presence of permanent or temporary emigration or immigration, and under stochastic time variation in capture probabilities, is evaluated empirically using Monte Carlo simulations.

2. Definitions

We assume a k-sample capture-recapture study where known losses on capture are permitted. On the jth capture occasion ($j = 1, \ldots, k$), unmarked individuals are marked before release and marked individuals have their marks recorded before release. The record of captures for a marked individual over k occasions is referred to as a capture history. A matrix of capture histories, where each row represents the capture history of a marked individual in the population, is referred to as a capture history matrix. The set of marked individuals released on the jth occasion constitute a cohort. The following notation, which is standard in the capture-recapture literature (Darroch, 1958; Jolly, 1965; Burnham *et al.*, 1987), is used in this paper:

N number of animals at risk of capture on the first capture occasion.
R_i number of animals released on the ith occasion ($i = 1, \ldots, k-1$).
m_i number of recaptures on the ith occasion ($i = 2, \ldots, k$), $m_1 = 0$.
r_i number of animals released on the ith occasion that were ever recaptured ($i = 1, \ldots, k-1$).
z_i number of recaptures of animals in cohorts released prior to the ith occasion that were subsequently recaptured after capture occasion i, but not recaptured on occasion i ($i = 2, \ldots, k-1$).
T_i number of recaptures of animals in cohorts released prior to or at release occasion $i-1$ that were subsequently recaptured at or after capture occasion i ($i = 2, \ldots, k-1$). $T_i = m_i + z_i$.
u_i number of unmarked individuals captured on the ith occasion ($i = 1, \ldots, k$).

u_i^* number of unmarked individuals captured on or after the ith occasion ($i = 1, \ldots, k-1$).
n_i number of individuals captured on the ith occasion ($i = 1, \ldots, k$).
$T_i^* = T_i + u_i^* (i = 2, \ldots, k-1), T_1^* = u_1^*$.
$Q_i = \sum_{j=1}^{i-1}(R_j - m_j)$, ($i = 2, \ldots, k-1$).

3. A test for closure

The closed-population capture-recapture model M_t is a special (constrained) case of the open-population Jolly-Seber model (Jolly, 1965; Seber, 1965; Lebreton *et al.*, 1992), where the constraint imposed is one of demographic and geographic closure. Hence, in principal, we could test for closure under the null model M_t by constructing a test with the Jolly-Seber model as a specific alternative. In Stanley and Burnham (1999) components of such a test are presented as part of a fully efficient goodness-of-fit test for model M_t. These components are chi-square, are independent, and, when added, give us a test for closure.

Beginning with the distribution of the minimal sufficient statistic (MSS) for the Jolly-Seber model (Burnham *et al.*, 1987, p. 177), constraints can be imposed to yield the MSS of the no-recruitment model (Pollock *et al.*, 1990) and a residual distribution. The residual distribution represents lack-of-fit of the no-recruitment model, and is the product of $k-2$ hypergeometric distributions (Pollock *et al.*, 1974; Burnham, 1997):

$$\prod_{i=2}^{k-1} \frac{\binom{T_i}{m_i}\binom{u_i^*}{u_i}}{\binom{T_i + u_i^*}{n_i}}.$$

These distributions lead to $k-2$, 2×2 contingency tables where, for example, on the ith occasion we have:

m_i	$T_i - m_i$
u_i	$u_i^* - u_i$

The $k-2$ χ^2 statistics computed from these 2×2 tables can be added to yield an overall test of the null hypothesis of the no-recruitment model versus the Jolly-Seber model, and is equivalent to the Pollock *et al.* (1974) test for recruitment allowing mortality. We will refer to this test as χ^2_{NR} and denote its degrees of freedom by df_{NR}, where nominally $df_{NR} = k-2$.

By further constraining the no-recruitment model to have no-mortality, we arrive at the closed-population model M_t (Darroch, 1958; Otis *et al.*, 1978). Stanley and Burnham (1999) present a statistic for testing the null hypothesis of model M_t against the no-recruitment model as a specific alternative. This test statistic, which we denote χ^2_{tNR}, is computed as:

$$\chi^2_{tNR} = \sum_{i=1}^{k-1}(r_i - R_i\hat{\lambda}_i)^2\left[\frac{1}{R_i\hat{\lambda}_i} + \frac{1}{R_i(1-\hat{\lambda}_i)}\right] + \sum_{i=1}^{k-1}(n_i - T_i^*\hat{\tau}_i)^2\left[\frac{1}{T_i^*\hat{\tau}_i} + \frac{1}{T_i^*(1-\hat{\tau}_i)}\right]$$

where, $\hat{\tau}_i = \hat{p}_i/(1 - \hat{q}_i\hat{q}_{i+1}\ldots\hat{q}_k)$, $\hat{\lambda}_i = 1 - \hat{q}_{i+1}\hat{q}_{i+2}\ldots\hat{q}_k$, $\hat{q}_i = (1 - \hat{p}_i)$, $\hat{p}_i = n_i/(\hat{N} - u_1^* + u_i^* + Q_i)$, and $\hat{N}$ is the maximum likelihood estimate of N estimated conditionally under null model M_t (formula (10.2) in Burnham, 1997). Because our simulations did not allow for losses

on capture, we estimated p_i as $\hat{p}_i = n_i/\hat{N}$ and, for simplicity, used the asymptotically equivalent unconditional maximum likelihood estimate of N. The nominal degrees of freedom for this test, denoted df_{tNR}, are $k-2$. However, when data are sparse, $\hat{\lambda}_i$ or $\hat{\tau}_i$ can equal 0 and the contribution to χ^2_{tNR} from the component is exactly zero. Thus, for each $\hat{\lambda}_i = 0$ or each $\hat{\tau}_i = 0$, we reduce the degrees of freedom by one. This test differs from the test for mortality assuming no recruitment described by Pollock *et al.* (1974).

The statistics χ^2_{NR} and χ^2_{tNR} are independent, and can be added to yield an overall χ^2 test for closure under the null model M_t. This test, which we denote χ^2_c and compute as $\chi^2_c = \chi^2_{NR} + \chi^2_{tNR}$, has nominal degrees of freedom $df_c = df_{NR} + df_{tNR}$.

We note that the sequence of nested models described above is only one of many that are possible, and that the statistic χ^2_c can be computed in an alternative manner. Once again we start with the distribution of the MSS for the Jolly-Seber model. However, instead of constraining this model to have no-recruitment, we constrain it to have no-mortality. Burnham (1997; also Pollock *et al.*, 1974) has shown that the resulting residual distribution is a product of hypergeometric distributions, and takes the form:

$$\prod_{i=2}^{k-1} \frac{\binom{Q_i - m_i}{z_i}\binom{R_i}{r_i}}{\binom{Q_{i+1}}{T_{i+1}}}.$$

These distributions lead to $k-2$, 2×2 contingency tables where, for example, on the ith occasion we have:

z_i	$Q_i - m_i - z_i$
r_i	$R_i - r_i$

The $k-2$ χ^2 statistics from these 2×2 tables can be added to yield an overall test of the null hypothesis of the no-mortality model versus the Jolly-Seber model, and is equivalent to the Pollock *et al.* (1974) test for mortality allowing recruitment. We will refer to this test as χ^2_{NM} and denote its degrees of freedom by df_{NM}, where nominally $df_{NM} = k-2$.

Finally, by further constraining the no-mortality model to have no-recruitment, we once again arrive at the closed-population model M_t. Stanley and Burnham (1999) present a statistic for testing the null hypothesis of model M_t against the no-mortality model as a specific alternative. This test statistic, which we denote χ^2_{tNM}, is computed as:

$$\chi^2_{tNM} = \sum_{i=1}^{k-1}(u_i - u_i^*\hat{\tau}_i)^2\left[\frac{1}{u_i^*\hat{\tau}_i} + \frac{1}{u_i^*(1-\hat{\tau}_i)}\right] + \sum_{i=2}^{k}(m_i - Q_i\hat{p}_i)^2\left[\frac{1}{Q_i\hat{p}_i} + \frac{1}{Q_i(1-\hat{p}_i)}\right], \tag{1}$$

where $\hat{\tau}_i$ and $\hat{P}_i$ are as defined above for χ^2_{tNR}. The nominal degrees of freedom for this test, denoted df_{tNM}, are $k-2$. However, when data are sparse, $\hat{\tau}_i$ or $\hat{p}_i$ can equal 0 and the contribution to χ^2_{tNM} from the component is exactly zero. Thus, for each $\hat{\tau}_i = 0$ or each $\hat{p}_i = 0$, we reduce the degrees of freedom by one. Given χ^2_{tNM}, an alternative form for the closure test χ^2_c is $\chi^2_c = \chi^2_{NM} + \chi^2_{tNM}$, with nominal degrees of freedom $df_c = df_{NM} + df_{tNM}$.

4. A simulation study of power

The power of the test statistic χ_c^2 to detect violation of demographic closure was evaluated empirically using Monte Carlo simulations. The basic procedure was to first specify the time-specific structure on the k capture probabilities, and second to superimpose permanent or temporary emigration or immigration. Once the time-specific structure and form of migration were specified, 1,000 $N \times k$ capture history matrices were simulated and χ_c^2 computed. We note that in the context of losses, mortality and permanent emigration are equivalent, and in the context of additions, recruitment and permanent or temporary immigration are equivalent. Hence, we did not do separate simulations for mortality and recruitment.

Models assigning a capture probability (p) to the ith individual on the jth occasion ($i = 1, \ldots, N; j = 1, \ldots, k$), took the form $p_{ij} = [1 + \exp(-\beta_0 - \beta_j)]^{-1}$, where β_0 is a scaling factor and β_j introduces stochastic time variation and is iid $N(0, \sigma_\beta^2)$. The random nature of β_j means that, for a particular model (i.e., constant β_0, σ_β^2), the $\{\beta_j\}$ change on each of the 1,000 repetitions. Levels for factors k, N, β_0 and σ_β^2 used in the simulations are in Table 1.

Emigration was modeled by selecting y individuals from a population of size N, where $y \sim \text{bin}(N, \pi)$, and allowing them to emigrate. For simulations with $N = 100$ (Table 1), we let π take on values of 0, 0.3, and 0.6, giving expected values for the number of emigrants (E[y]) of 0, 30, and 60, respectively. For simulations with $N = 300$ (Table 1), we let π take on values of 0, 0.167, and 0.333, giving for E[y] values of approximately 0, 50, and 100, respectively. An indicator variable, I_{ij} ($i = 1, \ldots, N; j = 1, \ldots, k$), was used to keep track of whether the ith individual was present ($I_{ij} = 1$) or not present ($I_{ij} = 0$) in the population at risk of capture on the jth occasion. Thus, if $I_{i,j-1} = 1$ and $I_{ij} = 0$, we know the ith individual emigrated during the open interval $(j-1, j)$. For brevity, we call j the "departure interval". Similarly, if $I_{i,j-1} = 0$ and $I_{ij} = 1$, we know the ith individual returned during the open interval $(j-1, j)$. In this case, we call j the "arrival interval".

Permanent emigration was modeled in two ways: (1) by specifying a single fixed departure interval for all emigrants, effectively creating a single pulse of emigrants, and (2) by randomly selecting an individual departure interval for each emigrant, effectively creating a continuous stream of emigrants. In the first case we say there was a "fixed" departure interval, and in the second case say there was a "random" departure interval. The random departure interval for an individual ($j; j = 1, \ldots, k$) was selected with probability $1/k$.

Similarly, temporary emigration was modeled in two ways: (1) by specifying a single fixed departure interval and a single fixed, later, arrival (i.e., return) interval for all emigrants, and (2)

Table 1. Levels of factors used in simulations for the number of capture occasions (k), true population size (N), a scaling factor (β_0), and stochastic time variation (σ_β^2). Within a row, every level of β_0 was simulated with every level of σ_β^2, resulting in 12 time-specific models per row and 48 models in total. Only results for levels in boldface type are presented in this paper.

k	N	β_0	σ_β^2
6	**100**	**−1.2**, −1.9, −2.6	**0, 0.5, 1.0, 1.5**
	300	**−1.2**, −1.9, −2.6	**0, 0.5, 1.0, 1.5**
12	**100**	**−1.6**, −2.3, −3.0	**0, 0.5, 1.0, 1.5**
	300	**−1.6**, −2.3, −3.0	**0, 0.5, 1.0, 1.5**

by randomly selecting an individual departure interval and an individual arrival (i.e., return) interval for each emigrant. The random departure interval of an individual $(j; j = 1, \ldots, k-1)$ was selected with probability $1/(k-1)$, and the random arrival (return) interval of the same individual $(j'; j' = j+1, \ldots, k)$ was selected with probability $1/(k-j)$.

Note that for both permanent and temporary emigration, a departure interval of $j = 1$ means emigration occurred before the first sampling occasion. We modeled emigration in this manner to emulate the biologically realistic situation where disturbance from setting up the trapping grid causes animals to emigrate.

In this study, for the y individuals selected to emigrate, the $\{I_{ij}\}$ for permanent emigration with a fixed departure interval was $I_{i1} = I_{i2} = 1, I_{i3} = \cdots = I_{i6} = 0$ for $k = 6$, and $I_{i1} = \cdots = I_{i4} = 1, I_{i5} = \cdots = I_{i12} = 0$ for $k = 12$. For temporary emigration with fixed departure and fixed arrival intervals, the $\{I_{ij}\}$ was $I_{i1} = I_{i2} = 1, I_{i3} = I_{i4} = 0, I_{i5} = I_{i6} = 1$ for $k = 6$, and $I_{i1} = \cdots = I_{i4} = 1, I_{i5} = \cdots = I_{i8} = 0, I_{i9} = \cdots = I_{i12} = 1$ for $k = 12$. For the $(N - y)$ individuals that did not emigrate, the $\{I_{ij}\}$ was $I_{i1} = \cdots = I_{i6} = 1$ for $k = 6$, and $I_{i1} = \cdots = I_{i12} = 1$ for $k = 12$.

Immigration was modeled by selecting y individuals from a "source" population of size 100, where $y \sim \text{bin}(100, \pi)$, and allowing them to immigrate into a "sink" population of size N. For $N = 100$ (Table 1), we let π take on values of 0, 0.3, and 0.6, giving for E[y] values of 0, 30, and 60, respectively. For $N = 300$ (Table 1), we let π take on values of 0, 0.5, and 1, giving for E[y] values of 0, 50, and 100, respectively. As was done for emigration, an indicator variable, I_{ij} $(i = 1, \ldots, N; j = 1, \ldots, k)$, was used to keep track of whether the ith individual on the jth occasion was present $(I_{ij} = 1)$ or not present $(I_{ij} = 0)$ in the population at risk of capture.

Permanent immigration was modeled in two ways: (1) by specifying a single fixed arrival interval for all immigrants, effectively creating a single pulse of immigrants, and (2) by randomly selecting an individual arrival interval for each immigrant, effectively creating a continuous stream of immigrants. The random arrival interval for an individual $(j; j = 1, \ldots, k)$ was selected with probability $1/k$.

Temporary immigration was also modeled in two ways: (1) by specifying a single fixed arrival interval and a single fixed departure interval for all immigrants, and (2) by randomly selecting an individual arrival interval and an individual departure interval for each immigrant. The random arrival interval of an individual $(j; j = 1, \ldots, k-1)$ was selected with probability $1/(k-1)$, and the random departure interval of the same individual $(j'; j' = j+1, \ldots, k)$ was selected with probability $1/(k-j)$.

Note that for both permanent and temporary immigration, a random arrival interval of $j = 1$ means immigration occurred before the first sampling occasion. We modeled immigration in this manner to emulate the biologically realistic situation where, for example, baiting traps or other activities associated with setting up the trapping grid causes animals to immigrate.

In this study, for the y individuals selected to immigrate, the $\{I_{ij}\}$ for permanent immigration with a fixed arrival interval was $I_{i1} = I_{i2} = 0, I_{i3} = \cdots = I_{i6} = 1$ for $k = 6$, and $I_{i1} = \cdots = I_{i4} = 0, I_{i5} = \cdots = I_{i12} = 1$ for $k = 12$. For temporary immigration with fixed arrival and fixed departure intervals, the $\{I_{ij}\}$ was $I_{i1} = I_{i2} = 0, I_{i3} = I_{i4} = 1, I_{i5} = I_{i6} = 0$ for $k = 6$, and $I_{i1} = \cdots = I_{i4} = 0, I_{i5} = \cdots = I_{i8} = 1, I_{i9} = \cdots = I_{i12} = 0$ for $k = 12$. For the $(100 - y)$ individuals in the source population that did not migrate into the sink population, the $\{I_{ij}\}$ was $I_{i1} = \cdots = I_{i6} = 0$ for $k = 6$, and $I_{i1} = \cdots = I_{i12} = 0$ for $k = 12$. For the N individuals in the sink population, the $\{I_{ij}\}$ was $I_{i1} = \cdots = I_{i6} = 1$ for $k = 6$, and $I_{i1} = \cdots = I_{i12} = 1$ for $k = 12$.

Once the time-specific structure (i.e., $\{p_{ij}\}$) and form of migration (i.e., $\{I_{ij}\}$) were specified,

Bernoulli trials, each with parameter $(I_{ij}p_{ij})$ $(i = 1, \ldots, N; j = 1, \ldots, k)$, were performed to determine whether individual i was captured on the jth occasion, and to generate the $N \times k$ capture history matrix. The test statistic χ_c^2 was computed from this capture history matrix. In all, 48 time specific models (Table 1) and eight forms of migration were investigated, giving a total of 384 cases. For each case we simulared 1.000 capture history matrices. Thus, a total of 384,000 capture history matrices were simulated to investigate the power of χ_c^2. In the interest of economy, we present only results for the factor levels in boldface type in Table 1. Results not reported are available from the first author.

In the results that follow, Pearson's chi-square statistics were computed for all 2×2 contingency tables. In cases where the expected value for one or more cells of a table was less than two, the statistic was not computed and that 2×2 table did not contribute to the degrees of freedom of χ_c^2. Hence, computed values for χ_{NR}^2 and χ_{NM}^2, as well as df_{NR} and df_{NM}, are based only on tables for which data were not sparse. In our simulations χ_{tNM}^2 was not computed using (1). Rather, we computed it using the following formula: $\chi_{tNM}^2 = \chi_{NR}^2 + \chi_{tNR}^2 - \chi_{NM}^2$, with $df_{tNM} = df_{NR} + df_{tNR} - df_{NM}$.

Finally, because simulation of capture probabilities was on a *logit* scale (i.e., $logit(p_{ij}) = \beta_0 + \beta_j$) and may be difficult to interpret, we will, in the results, refer to the expected median capture probability (EMCP). The EMCP is simply the expected median capture probability over all k capture occasions and is defined as $\text{EMCP} = (1 + \exp(-\beta_0))^{-1}$. The EMCP is useful in that it gives a more intuitive feel for the capture probabilities under discussion than does β_0 by itself. In Table 1 the EMCP for $\beta_0 = -1.2$, -1.9, -2.6, -1.6, -2.3, and -3.0 is 0.23, 0.13, 0.07. 0.17, 0.09, and 0.05, respectively.

5. Results

The test statistic, χ_c^2, was sensitive to permanent emigration of individuals (Table 2). The power of the test to detect permanent emigration with the fixed departure interval used was approximately twice the power of the test to detect permanent emigration with a random departure interval. For both fixed and random departure intervals, power increased as the number of capture occasions, k, increased. or the expected number of emigrants increased. With the exception of χ_{NR}^2 (which tests for immigration), all the component statistics of χ_c^2 were sensitive to emigration. The test statistic χ_{tNR}^2 was the most powerful of the component statistics. Generally speaking, under the time-specific null model (i.e., $E[y] = 0$) Type I error rates for all test statistics closely approximated nominal rates.

The power of χ_c^2 and the component statistics χ_{NR}^2, χ_{tNR}^2, χ_{NM}^2, and χ_{tNM}^2 to detect temporary emigration was poor (Table 2). In all the cases we simulated, power was less than 20%. It appears that χ_c^2 picks up temporary emigration with fixed intervals more strongly than temporary emigration with random intervals. However, the differences are not great. All the component statistics of χ_c^2 had approximately the same power to detect temporary emigration.

Permanent immigration was detected by χ_c^2 (Table 3), though less strongly than permanent emigration. The test was equally sensitive to permanent immigration with a fixed arrival interval and a random arrival interval. Power increased as the number of capture occasions, k, increased or the expected number of immigrants increased. The component statistics χ_{NR}^2 and χ_{tNM}^2 detected permanent immigration and had approximately equal power. The component statistics χ_{tNR}^2 and χ_{NM}^2 had no power to detect permanent immigration.

Temporary immigration was detected by χ_c^2 (Table 3). The power of χ_c^2 to detect temporary

Table 2. Empirical power ($\alpha = 0.05$) of χ^2_c and the component statistics in the presence of permanent or temporary emigration, and fixed (F) or random (R) departure or arrival intervals. The EMCP for $k = 6$ was 0.23 and for $k = 12$ was 0.17. Data are percentages averaged over four levels of stochastic time variation. Standard errors for individual entries can be approximated as $[p(1-p)/4000]^{0.5}$, where p is power.

			χ^2_c		χ^2_{NR}		χ^2_{iNR}		χ^2_{NM}		χ^2_{iNM}	
N	k	E[y]	F	R	F	R	F	R	F	R	F	R
(a) Permanent emigration												
100	6	0	5	6	5	6	5	5	6	6	5	6
		30	34	13	7	6	40	15	36	13	16	9
		60	71	27	6	7	76	30	61	23	44	17
	12	0	7	6	7	7	6	5	7	6	7	6
		30	59	26	6	7	64	29	61	22	28	17
		60	93	49	6	8	95	53	78	40	68	32
300	6	0	6	5	5	6	5	5	6	5	5	5
		50	35	12	5	6	42	14	37	11	13	9
		100	72	25	5	6	79	30	74	24	24	12
	12	0	6	6	6	7	6	5	7	6	6	6
		50	57	23	6	6	66	27	61	21	25	14
		100	94	51	6	6	97	58	95	50	44	23
(b) Temporary emigration												
100	6	0	6	6	5	6	5	5	6	6	5	7
		30	7	8	7	8	6	6	7	7	6	7
		60	9	8	9	8	7	5	9	8	8	7
	12	0	7	7	7	7	6	6	7	7	7	7
		30	11	9	10	8	9	7	10	7	9	9
		60	12	11	12	12	8	8	12	9	9	10
300	6	0	6	6	6	6	6	5	6	6	6	6
		50	9	7	7	6	7	6	8	6	7	6
		100	13	8	10	8	10	6	11	6	10	8
	12	0	5	6	6	6	5	6	6	7	6	6
		50	11	8	9	7	9	7	10	7	9	8
		100	18	12	14	11	12	9	14	9	14	10

immigration with fixed intervals was greater than with random intervals, and was also greater than the power to detect permanent immigration. The power of χ^2_c to detect temporary immigration with random intervals was relatively poor. Power to detect temporary immigration increased as the number of occasions, k, increased, or the expected number of immigrants increased. For fixed intervals, the power of the component statistics of χ^2_c were approximately equal. For random intervals, the power of χ^2_{NR} and χ^2_{iNM} were approximarely equal, whereas the component statistics χ^2_{iNR} and χ^2_{NM} were not sensitive to temporary immigration.

The data presented in Tables 2 and 3 are averages over the four levels of stochastic time variation simulated (Table 1). While Type I error rates were unaffected by σ^2_β, the power of χ^2_c and the component statistics tended to increase as σ^2_β increased. To illustrate, consider the case of permanent emigration with a fixed departure interval, $N = 100$, $k = 6$, and E[y] = 30 in Table 2.

Table 3. Empirical power ($\alpha = 0.05$) of χ^2_c and the component statistics in the presence of permanent or temporary immigration, and fixed (F) or random (R) departure or arrival intervals. The EMCP for $k = 6$ was 0.23 and for $k = 12$ was 0.17. Data are percentages averaged over four levels of stochastic time variation. Standard errors for individual entries can be approximated as $[p(1-p)/4000]^{0.5}$, where p is power.

			χ^2_c		χ^2_{NR}		χ^2_{iNR}		χ^2_{NM}		χ^2_{iNM}	
N	k	E[y]	F	R	F	R	F	R	F	R	F	R
(a) Permanent immigration												
100	6	0	6	6	6	6	6	6	7	7	6	6
		30	9	9	11	11	5	4	6	6	11	10
		60	15	16	18	19	5	5	6	6	18	18
	12	0	7	7	7	8	7	6	7	7	7	7
		30	15	12	20	18	5	4	6	7	17	13
		60	28	28	37	37	5	4	6	7	32	32
300	6	0	6	6	6	6	6	5	7	6	6	7
		50	9	8	12	10	5	4	5	5	12	11
		100	18	19	22	21	6	6	6	6	22	23
	12	0	5	6	6	6	5	6	6	6	6	6
		50	15	16	20	20	5	5	6	6	19	18
		100	35	34	44	43	5	5	6	6	42	41
(b) Temporary immigration												
100	6	0	6	6	7	7	5	5	6	6	6	6
		30	17	8	14	8	12	6	14	6	14	8
		60	36	11	27	12	23	7	27	7	24	11
	12	0	6	6	7	6	5	5	6	6	7	6
		30	26	10	21	10	17	8	20	9	19	10
		60	57	16	41	18	40	8	41	9	41	16
300	6	0	5	6	5	6	5	5	5	5	5	5
		50	18	7	13	7	13	6	14	6	14	7
		100	41	11	30	12	27	6	29	6	29	12
	12	0	6	6	6	6	5	6	6	6	6	6
		50	28	9	21	9	19	7	21	7	21	9
		100	65	18	47	20	46	9	46	9	49	19

If we partition these results by the lowest and highest levels of σ^2_β, we see that power for χ^2_c, χ^2_{NR}, χ^2_{iNR}, χ^2_{NM}, and χ^2_{iNM} is 22%, 7%, 29%, 24%, and 10%, respectively, when $\sigma^2_\beta = 0$, but increases to 42%, 6%, 46%, 44%, and 20%, respectively, when $\sigma^2_\beta = 1.5$.

The component statistics, χ^2_{NR}, χ^2_{iNR}, χ^2_{NM}, and χ^2_{iNM}, themselves consist of component statistics (subcomponents) that may provide information on where failure of the closure assumption is occurring. While we did not evaluate the subcomponents of χ^2_{iNR} or χ^2_{iNM}, we did evaluate the subcomponents (i.e., the individual 2×2 tables) of χ^2_{NR} and χ^2_{NM}, which we denote $\chi^2_{NR,j}$ and $\chi^2_{NM,j}$ ($j = 2, \ldots, k-1$). We present these results in Table 4 for the case of $k = 6$, $N = 300$, $\beta_0 = -1.2$ (EMCP = 0.23), and E[y] = 100, under various forms of migration.

Subcomponent $\chi^2_{NR,j}$ tests the null hypothesis that the relationship between marked and unmarked animals in the jth sample does not differ from the relationship between marked and

Table 4. Empirical power ($\alpha = 0.05$) for the subcomponent statistics of χ^2_{NR} and χ^2_{NM} in the presence of emigration or immigration, and fixed or random departure or arrival intervals. Results are presented for the case of $k = 6$, $N = 300$, $\beta_0 = -1.2$ (EMCP = 0.23), and E[y] = 100. Data are percentages averaged over four levels of stochastic time variation. Standard errors for individual entries can be approximated as $[p(1-p)/4000]^{0.5}$, where p is power.

		Fixed Interval				*Random Interval*			
Migration Type	*Sub-component*	$j = 2$	*3*	*4*	*5*	$j = 2$	*3*	*4*	*5*
(a) Emigration									
Permanent (for fixed departure	$\chi^2_{NR,j}$	5	5	6	6	6	6	6	6
$j = 3$)	$\chi^2_{NM,j}$	5	72	46	26	9	15	18	17
Temporary (for fixed departure	$\chi^2_{NR,j}$	5	5	14	5	7	8	8	6
$j = 3$, and for fixed arrival $j = 5$)	$\chi^2_{NM,j}$	6	15	5	5	6	6	5	5
(b) Immigration									
Permanent (for fixed arrival $j = 3$)	$\chi^2_{NR,j}$	34	5	5	6	14	17	13	9
	$\chi^2_{NM,j}$	5	6	5	6	5	5	5	6
Temporary (for fixed arrival $j = 3$,	$\chi^2_{NR,j}$	21	17	24	6	8	7	6	12
and fixed departure $j = 5$)	$\chi^2_{NM,j}$	5	22	16	21	6	5	6	7

unmarked animals in all remaining samples (i.e., samples $j + 1$ through k). Thus, in cases where numbers of unmarked animals are disproportionately high in samples $j + 1$ through k, which occurs when there is immigration or recruitment into the population between occasions j and $j + 1$, $\chi^2_{NR,j}$ will tend to reject. In Table 4 we see that, as expected, $\chi^2_{NR,j}$ rejected at close to nominal rates for permanent and temporary emigration. The slightly elevated rejection rates for temporary emigration are caused by unmarked individuals returning to the population. For permanent and temporary immigration, $\chi^2_{NR,j}$ rejected at somewhat higher rates, with fixed intervals being detected more strongly than random intervals. Interestingly, for permanent immigration with a fixed arrival of $j = 3$, $\chi^2_{NR,2}$ rejected while the remaining $\chi^2_{NR,j}$ rejected at nominal rates. This suggests that the interval during which immigration occurs might in some cases be discernable from examination of the $\chi^2_{NR,j}$.

Subcomponent $\chi^2_{NM,j}$ tests the null hypothesis that the proportion of recaptures for the jth cohort does not differ from the proportion of recaptures after the jth occasion for cohorts released prior to the jth occasion. Thus, in cases where emigration or mortality occurs between occasions $j - 1$ and j, the proportion of recaptures for the jth cohort would on average be disproportionately high and $\chi^2_{NM,j}$ will tend to reject. Indeed, in Table 4 we see that this is what happens. For permanent emigration with a fixed departure interval $\chi^2_{NM,3}$ detects the pulse of emigrants with relatively high power. Similarly, the pulse of temporary emigrants with a fixed departure interval is detected, albeit with somewhat lower power. For emigration with random departure intervals, the power of $\chi^2_{NM,j}$ is relatively low. For permanent and temporary immigration we find, as expected, that $\chi^2_{NM,j}$ tends to reject at nominal rates. The exception is for temporary immigration with fixed intervals, where $\chi^2_{NM,j}$ rejects at greater than nominal rates in response to individuals leaving the population after immigrating.

6. Discussion

The closure test we present is very versatile, and provides researchers with a new tool for testing demographic closure in the presence of time-specific variation in capture probabilities. The test is developed in a framework unifying open- and closed-population models, and uses partitions of the residual distribution for model M_t that, in the spirit of Pollock *et al.* (1974), can be used to select an appropriate model or submodel for estimation. While we present only two partitions of this residual distribution, one resulting in the no-recruitment model and one resulting in the no-mortality model (Pollock *et al.*, 1990), other less appealing partitions are possible (e.g., constraining recruitment or mortality over only some of the capture occasions to give "partial recruitment" or "partial mortality" models).

Our simulation results demonstrate the test statistic χ^2_c is sensitive to permanent and temporary emigration and immigration. For fixed intervals, which we used to model a "pulse" of migrants, the relative power of χ^2_c for the forms of migration investigated was: permanent emigration > temporary immigration > permanent immigration > temporary emigration. For random intervals, which we used to model a "continuous flow" of migrants, the relative power was: permanent emigration > permanent immigration > temporary immigration > temporary emigration. Thus, in general the test was most sensitive to permanent emigration and least sensitive to temporary emigration.

For populations consisting of 100 individuals, the overall power of χ^2_c to detect closure violation was moderate to low. Even in cases where 60% of the population migrated, only permanent emigration and temporary immigration were detected with reasonable power. For populations of 300 individuals power improved somewhat, but for temporary emigration and permanent and temporary immigration power was still somewhat low even when as many as 100 individuals migrated. Thus, χ^2_c will be of greatest value to researchers when the number of migrating individuals is relatively large and there is permanent emigration.

Of some concern to ecologists is how robust the estimator for model M_t is when closure is violated, but is not detected by the closure test—for example, temporary emigration where the power of the test is low. If the estimator is robust to such violations, then our inability to detect them may be of no consequence. However, if bias is high, then our inability to detect closure violations, such as temporary emigration, are more significant. To our knowledge, studies evaluating the robustness of the closed-population models to violation of the closure assumption have not been conducted, and we did not do so in this study. Rather, our focus was on the performance and power of the proposed closure test under a wide array of biologically realistic forms of closure violation. Nonetheless, we think studies evaluating estimator robustness under closure violation are of value and merit future research.

A useful feature of the test we present is that it can be decomposed into informative components. Examination of these components can lead to greater understanding of the nature of the closure violation. To illustrate, we found that for permanent emigration χ^2_{tNR}, χ^2_{NM}, and χ^2_{tNM} tend to reject, while χ^2_{NR} does not. In contrast, for permanent immigration, χ^2_{NR} and χ^2_{tNM} tend to reject while χ^2_{tNR} and χ^2_{NM} do not. Finally, for temporary emigration and temporary immigration, all the component statistics reject to a greater or lesser degree. Thus, in principle, it is possible to distinguish between closure violation due to permanent emigration, permanent immigration, and temporary emigration or immigration using the pattern of rejection of the component statistics. In a similar manner, examination of the subcomponent statistics of χ^2_{NR} and χ^2_{NM}, in conjunction with an understanding of the specific hypotheses being tested, can provide insights into the

intervals during which migration is occurring. This is particularly true in the case of permanent emigration or permanent immigration, where migration occurs as a "pulse".

The closure test presented in this paper was developed under a null model allowing for time-specific variation in capture probabilities in the absence of behavioral variation or heterogeneity in capture probabilities. Indeed, in the presence of behavioral variation or heterogeneity in capture probabilities and no migration, the test rejects at rates greater than the nominal error rates (Stanley and Burnham, 1999). Thus, the test is not intended to replace the closure test of Otis *et al.* (1978) in which the null model allows for heterogeneity in capture probabilities but is sensitive to time or behavioral variation in capture probabilities. We see this test as a tool to be used in conjunction with the Otis *et al.* (1978) test, to better detect closure violations in a capture-recapture data set. Like the Otis *et al.* (1978) test, our overall test, χ^2_c, rejects in the presence of behavioral variation in capture probabilities (Stanley and Burnham, 1999). However, examination of the component statistics reveals that, in the presence of behavioral variation in capture probabilities and no migration, χ^2_{NR} and χ^2_{tNM} reject but χ^2_{tNR} and χ^2_{NM} do not. Thus, certain forms of closure violation, such as permanent emigration and temporary immigration where χ^2_{tNR} and χ^2_{NM} tend to reject, may be detectable even in the presence of behavioral variation in capture probabilities. For this reason, the four component statistics of χ^2_c will need to be reported in any implementation of this test.

References

Burnham, K.P. (1997) Distributional results for special cases of the Jolly-Seber model. *Communications in Statistics*, **26**, 1395–409.

Burnham, K.P., Anderson, D.R., White, G.C., Brownie, C., and Pollock, K.H. (1987) Design and analysis methods for fish survival experiments based on release-recapture. *American Fisheries Society Monograph*. **5**.

Darroch, J.N. (1958) The multiple-recapture census: I. Estimation of a closed population. *Biometrika*, **45**, 343–59.

Jolly, G.M. (1965) Explicit estimates from capture-recapture data with both death and immigration-stochastic model. *Biometrika*, **52**, 225–47.

Jolly, G.M. (1982) Mark-recapture models with parameters constant in time. *Biometrics*, **38**, 301–21.

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

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

Pollock, K.H., Solomon. D.L., and Robson, D.S. (1974) Tests for mortality and recruitment in a K-sample tag-recapture experiment. *Biometrics*, **30**, 77–87.

Pollock, K.H., Nichols, J.D., Brownie, C., and Hines. J.E. (1990) Statistical inference for capture-recapture experiments. *Wildlife Monographs*. **107**.

Robson, D.S. and Flick, W.A. (1965) A non-parametric statistical method for culling recruits from a mark-recapture experiment. *Biometrics*, **21**, 936–47.

Robson, D.S. and Regier, H.A. (1968) Estimation of population number and mortality rates. In *Methods for Assessment of Fish Production in Fresh Waters*, IBP Handbook No. 3, W.E. Ricker (ed), Blackwell Scientific Publications: Oxford. pp. 124–58.

Seber, G.A.F. (1965) A note on the multiple recapture census. *Biometrika*, **52**, 249–59.

Seber, G.A.F. (1982) *The estimation of animal abundance and related parameters*, Second edition. Macmillan. New York, New York, USA.

Seber, G.A.F. (1992) A review of estimating animal abundance II. *International Statistical Reviews*, **60**, 129–66.

Stanley, T.R. and Burnham, K.P. (1999). A goodness-of-fit test for capture-recapture model M_t under closure. *Biometrics*, **55**, 127–38.

White, G.C., Anderson, D.R., Burnham, K.P., and Otis, D.L. (1982) *Capture-recapture and removal methods for sampling closed populations*. Los Alamos National Laboratory, LA-8787-NERP, Los Alamos, New Mexico, USA.

Biographical sketches

Thomas R. Stanley has a B.S. in Conservation, an M.S. in Zoology, and a Ph.D. in Wildlife Biology from Colorado State University (1997). He is currently a Wildlife Research Biologist in the Vertebrate Ecology Section of the USGS Midcontinent Ecological Science Center in Fort Collins, Colorado. His research interests are in the areas of vertebrate population biology, biostatistics, and wildlife-habitat relations. He has published papers on model selection for closed-population capture-recapture studies, the effects of environmental contaminants on avian survival and reproduction, and ecosystem management.

Kenneth P. Burnham has a B.S. in biology (1966) and an M.S. (1969) and Ph.D. (1972) in statistics. After having worked in Oregon, Alaska, Maryland and North Carolina, Ken is now part of the Colorado Cooperative Fish and Wildlife Research Unit at Colorado State University. Ken's professional interests are in design of wildlife research, statistical methods for capture-recapture and distance sampling, and data analysis in general for estimation of biological abundance and demographic parameters. A more recent research emphasis has been theory and application of information theoretic model selection (i.e., "AIC"); he is a coauthor of a recent book on this subject. Ken is a Fellow of the American Statistical Association.