NUMERICAL SURVIVAL RATE ESTIMATION FOR CAPTURE-RECAPTURE MODELS USING SAS PROC NLIN

Kenneth P. Burnham

USDA-Agricultural Research Service,
and Department of Statistics
Box 8203, NC State University
Raleigh, North Carolina 27695, USA

ABSTRACT   This paper discusses and illustrates recent developments in the modeling and analysis of capture-recapture data for open populations. Results are presented using a unified theory of cohort survival processes. The statistical models for survival rate estimation are represented as products of conditionally independent multinomials; this is achieved by always conditioning on the known number of releases initiating the cohorts. This approach facilitates extension of the Jolly-Seber model to control-treatment studies, or other contexts where comparison of population survival rates is of most interest. Numerical computation of maximum likelihood survival and capture rate estimators, under any model, is easily achieved using iteratively-reweighted nonlinear least squares in SAS PROC NLIN. In his classical paper, Jolly (1965) illustrated his methods using summary statistics on female black-kneed capsids (Blepharidopterus angulatus). Here I use the entire data set on males and females to illustrate survival rate estimation and flexible modeling of capture-recapture data.

KEY WORDS   capture-recapture, iteratively reweighted nonlinear least squares, Jolly-Seber model, SAS, survival rate

In the analysis of capture-recapture data, successful parameter estimation depends critically on using the correct model. However, we do not know the correct model; consequently, much of the effort in a "good" analysis of capture-recapture data should be directed at model selection. By "model" I primarily mean the structure of the expected recapture rates, given the releases, expressed as functions of survival and capture probabilities (Ø and p, respectively). Also relevant, but not considered here, is the structure of the the expected captures of unmarked animals as functions of Ø, p and abundance parameters (N, or B).

The Jolly-Seber model provides a good starting point for considering open population models (Jolly 1965, Seber 1965). The capture-recapture literature since 1965 is dominated by attempts to formulate alternative models, both more general than, and special cases of, the Jolly-Seber model, as well as explicit systems of models with associated formal model selection procedures (see, e.g., Robson 1969, Pollock 1975, Arnason and Baniuk 1980, Buckland 1980, Cormack 1981, Pollock 1981a, Sandland and Kirkwood 1981, Jolly 1982, Brownie and Robson 1983, Crosbie and Manly 1985, Brownie et al. 1986, Burnham et al. 1987 and Clobert et al. 1987). Moreover, it is clear that what is needed, and being sought, is a convenient, unified theory for capture-recapture which allows easy model formulation and analysis in the context of some standard statistical theory, such as the log-linear approach of Cormack (1981). Recently, Brownie and Robson (1983) pioneered the basis of another possible unified approach to modeling and analysis of marked-animal cohort survival processes (see also Brownie et al. 1985: 170-175; also note that Cormack 1964 is relevant here). Capture-recapture models can be formulated using multinomial and binomial models for first recaptures following the known releases. Standard maximum likelihood analysis is then used. Burnham et al. (1987) elaborate on this approach for the analysis of survival rates in capture-recature experiments.

Burnham et al. (1987) was in page proofs one year ago. Since then I have realized how to implement numerical maximum likelihood estimation of Ø and p in these models using iteratively reweighted nonlinear least squares. In particular, PROC NLIN in SAS (SAS 1985) can be used as a basis for easy, flexible estimation of the Ø and p parameters in capture-recapture models. The objectives of this paper are to (1) outline the underlying unified theory, and (2) illustrate flexible, numerical analysis of capture-recapture data (re Ø and p) with the full black-kneed capsids data of C. R. Muir (Muir 1957). Summary statistics for the female capsids were used as an example by Jolly (1965), and in fact it was this study by Muir which got George Jolly involved in capture-recapture (Jolly, personal communication).

Some General Capture-Recapture Theory

The notation of Burnham et al. (1987) is used here. Let k be the number of capture-release occasions, indexed as occasion i=1 to k. At each occasion there are releases; let h(i) stand for a capture history at release time (occasion) i. When the value of i is clear from the context, this notation is reduced to just "h." One format for the complete data set is the complete capture history array: a listing of every observed h(k) (i.e., capture history at the end of the study) and the corresponding number of animals X_h, with final capture history h.

When there are different groups of animals the notation is extended by a group index, $v=1, \ldots, V$. For example with males and females we have (h, X_{mh}, X_{fh}) (see, e.g., Table 1). A negative X_{vh} denotes that the animals with that history were lost on the last capture indicated in the string of 1's (captured) and 0's constituting "h." The complete capture history array is a convenient way to summarize the data. Models and analysis methods can be based on this data representation, for example the loglinear method of Cormack (1981) and the multinomial modeling of Crosbie and Manly (1985). There are, however, strong advantages to basing modeling and analysis on the data representation outlined below.

Let u_i be the number of unmarked animals captured at time i. Any of these animals released at time i have history $h(1)=1$ (for $i=1$), or $h(2)=01$ (for $i=2$), or $h(3)=001$ (for $i=3$), and so forth. Let R_{ih} be the number of animals released at time i that have capture history h ($\equiv h(i)$). Let m_{ijh} be the number of animals, of the R_{ih}, captured again at time j and not captured at times $i+1$ to $j-1$. Let r_{ih} be the sum $m_{i,i+1,h}+\cdots+m_{ikh} =$ the total number of animals, of the R_{ih}, that are ever recaptured again. Conversely, $R_{ih}-r_{ih}$ of this release subcohort are never captured again. The m_{ijh} represent mutually exclusive events; a reasonable model to assume is that these m_{ijh}, $j=i+1, \ldots, k$, are multinomial random variables given the known releases R_{ih}. The data from a capture-recapture study with only one group (category) of animals can be presented, without loss of information, as the statistics

$$u_i, \quad i=1, \ldots, k$$
$$R_{ih}, \quad m_{ijh}, \quad j=i+1, \ldots, k, \quad h(i) \in H(i), \quad i=1, \ldots, k-1.$$

Here $H(i)$ is just a symbol for the set of all observed capture histories in release cohort i. Each $h(i)$ defines a subcohort. If there are 2 or more groups we add the index $v=1, \ldots, V$, e.g., u_{vi}, R_{vih}, m_{vijh}. Numerous summary statistics can be defined from these basic variables, such as cohort totals R_i and m_{ij} obtained by summing over all subcohorts at time i, and the total recaptures on each occasion: $m_2=m_{12}$, $m_3=m_{13}+m_{23}$, and in general $m_j=m_{1j}+\cdots+m_{j-1,j}$, $j=3, \ldots, k$.

The probability distribution of a capture-recapture data set takes the conceptual form Pr{data} = Pr{first captures}*Pr{releases|captures}* Pr{recaptures|releases}. An explicit form for this Pr{data} can be given (Burnham 1987. A unified approach to animal release-resampling studies of survival, processes and population estimation. North Carolina State University, Institute of Statistics Mimeograph Series 1698, Raleigh, NC). It suffices here to say that Pr{first captures}=Pr{u_i, $i=1, \ldots, k$} and this term is the only one involving the population abundance and recruitment parameters (N and B in Jolly 1965). Unless we make very strong assumptions about the recruitment process, there is no information in this first term about the survival and capture parameters $\emptyset$ and p. The second term, Pr{releases|captures}, is equivalent to the distribution of losses on capture given the captures (if there are no such losses this second term vanishes). Except under extraordinary circumstances, all the information about $\emptyset$ and p is in the term, Pr{recaptures|releases}. For purposes of this paper, this term will now be expressed as 3 new terms, which will be formally labelled as terms 1, 2 and 3:

$$
\begin{aligned}
\text{Pr{recaptures|releases}} &= \text{Pr{JS\_MSS}*Pr{cohorts|JS\_MSS}*} \\
&\quad \text{Pr{subcohorts|cohorts}.} \\
&= (\text{term 1})*(\text{term 2})*(\text{term 3}). \quad (1)
\end{aligned}
$$

Here, JS_MSS denotes the component of the minimal sufficient statistic under the Jolly-Seber model which comes from the recaptures given releases. Cohorts refers to the m_{ij} data (given the R_i); subcohorts refers to the m_{ijh}.

The Jolly-Seber Model

The Jolly-Seber model assumes that the survival probability between capture times i and $i+1$ ($i=1, \ldots, k-1$), and the capture probability at time i are the same for all animals in the population at risk of capture. These parameters are $\emptyset_i$ and p_i respectively. For a full discussion of the assumptions in capture-recapture see Jolly (1965) or Pollock (1981b). The structure of the model, in terms of the m_{ij} given R_i is

$$E(m_{i,i+1}) = R_i(\emptyset_i p_{i+1}) \quad ,i=1, \ldots, k-1,$$
$$E(m_{i,j}) = R_i(\emptyset_i q_{i+1}) \cdots (\emptyset_{j-2} q_{j-1})(\emptyset_{j-1} p_j) \quad ,i=1, \ldots, k-1, j=i+1, \ldots, k,$$

where $q_i=1-p_i$. There are $2k-3$ estimable parameters: $\emptyset_i$, $i=1, \ldots, k-2$, p_i, $i=2, \ldots, k-1$, and the product $\emptyset_{k-1} p_k$.

Given the assumption that the parameters depend only upon time and that the recaptures (m) given releases (R) have independent multinomial distributions, the minimal sufficient statistic (JS_MSS) is the sums r_i, $i=1, \ldots, k-1$ and m_j, $j=2, \ldots, k-1$. This JS_MSS has dimension $2k-3$. For proofs and more details see Brownie and Robson (1983).

The distribution of JS_MSS can be expressed as the product of $2k-3$ conditional independent binomials. Before that distribution can be given we need some more notation. Define the totals T_i, $i=2, \ldots, k-1$ as $T_2=r_1$ and $T_j=T_{j-1} - m_{j-1} + r_{j-1}$, $j=3, \ldots, k-1$. Define two functions of the basic parameters: $\lambda_i=E(r_i|R_i)/R_i$, $i=1, \ldots, k-1$, and $\tau_j=E(m_j|T_j)/T_j$, $j=2, \ldots, k-1$. As functions of the parameters, λ is best expressed recursively:

$$\lambda_i=\emptyset_i(p_{i+1} + q_{i+1}\lambda_{i+1}) \quad ,i=1, \ldots, k-1,$$

using the convention that $\lambda_k=0$. Note that $\lambda_{k-1}=\emptyset_{k-1}p_k$. It is easy to compute these λ_i using backwards recursion. Finally,

$$\tau_j= \frac{p_j}{(p_j + q_j\lambda_j)} \quad ,j=2, \ldots, k-1.$$

Let $y|n \sim bin(n,\theta)$ denote that given n, y is a binomial random variable with sample size n and parameter θ. Then

$$r_i|R_i \sim bin(R_i,\lambda_i) \quad ,i=1, \ldots, k-1,$$
$$m_j|T_j \sim bin(T_j,\tau_j) \quad ,j=2, \ldots, k-1.$$

(2)

This defines the distribution of JS_MSS, hence term 1 in formula (1). Under the assumption that the data fit the Jolly-Seber model, terms 2 and 3 in (1) are products of multiple hypergeometric distributions and these terms provide the basis of a fully-efficient goodness of fit test to the Jolly-Seber model (Pollock et al. 1985, Burnham et al. 1987).

For inference about $\emptyset$ and p, assuming the Jolly-Seber model fits, all that is needed are these summary statistics and their distribution in (2) above. Note that any sub-model that is a special case of Jolly-Seber can be dealt with by starting from (2). Such sub-models are, e.g., the p_i and/or the $\emptyset_i$ constant (Jolly 1982, Brownie et al. 1986) or modeling the survival rates as functions of any time-varying variable (Clobert et al. 1987).

An immediate generalization is to allow V data sets, each one fitting the Jolly-Seber model with parameters $\emptyset_{vi}$ and p_{vi}, v=1, ..., V. In particular one might have data on adult males and females, or control and treatment population groups. Burnham et al. (1987) give an extensive treatment of this situation: testing goodness of fit, testing equality of the parameters over groups and parameter estimation for a class of models having closed-form solutions. Also, Burnham et al. (1987) discuss using program SURVIV (White 1983) to fit special cases of the Jolly-Seber model using numerical methods. Properly implemented, the SAS PROC NLIN approach below, and use of program SURVIV will give the same results. The advantages of this PROC NLIN approach is that SAS is a readily available package on a wide variety of machines and it is fairly easy to use, especially for anyone who has already used SAS.

Analysis Using SAS

Given the distribution in (2) we can in principle impose constraints on the parameters (then estimators are no longer closed-form) and compute the maximum likelihood estimates (MLE's). The problem is how to do this easily. Using PROC NLIN (nonlinear least squares) in SAS (1985) provides one solution to the computing problem.

For many statistical distributions, including binomials, the MLE's can oe found using iteratively reweighted nonlinear least squares (see e.g., Jennrich and Moore 1975, Green 1984). The binomial case will be outlined here. Let $y_i \sim \text{bin}(n_i, \pi_i(\underline{\theta}))$, independently for i=1, ..., g. Let $\underline{\theta}$ be an f-dimensional parameter vector with $f \leq g$, and let the π_i be known, well-behaved functions. Define weights w_i by

$$w_i = \frac{n_i}{\pi_i(\underline{\theta})(1-\pi_i(\underline{\theta}))} \quad , \quad i=1, ..., g;$$

note that w_i is the reciprocal of the variance of the ratio y_i/n_i. The MLE can found as the value of $\underline{\theta}$ which minimizes the weighted residual sum of squares,

$$\sum_{i=1}^{g} w_i((y_i/n_i) - \pi_i(\underline{\theta}))^2 \quad ,$$

in terms of $\underline{\theta}$, provided that the weights, as functions of the parameters, are recomputed at each step in the numerical optimization based on the current value of $\underline{\theta}$. PROC NLIN in SAS has the capability to do this iterative reweighting. PROC NLIN "thinks" it is solving the regression problem wherein the model is

$$y_i/n_i = \pi_i(\underline{\theta}) + \sigma\epsilon_i \quad , i=1, ..., g,$$

with the variance of ϵ_i being $1/w_i$ and σ (or σ^2) being an unknown parameter
to be estimated from the residual mean square. Subject to the Jolly-Seber
assumptions, theoretically $\sigma^2=1$. The SIGSQ=1 option is used to force $\sigma^2=1$
rather than estimate it. This results in PROC NLIN computing theoretical
variances and covariances which are identical to the appropriate estimated
likelihood theory variances.

Version 5 of SAS, and SAS for personal computers is even more powerful.
In these versions one can declare an alternative loss function to be mini-
mized (denoted in PROC NLIN as _LOSS_). Thus the negative of the log-
likelihood function can be used directly as the minimization criterion,
and SAS then prints out this (negative) log-likelihood value. It is still
necessary to declare the "regression" version of the model and to use the
SIGSQ=1 option to force SAS to set $\sigma^2=1$.

One more powerful advantage of PROC NLIN is that bounds can be put
on the parameters. In particular, one can bound a rate parameter to be
between 0 and 1. This is useful even for the fully-parameterized Jolly-
Seber model because the closed-form MLE's of the $\emptyset_i$ can exceed 1 in
which case there are benefits of constraining them (Buckland 1980).

Examples Using the Black-kneed Capsids Data

The capsids data.   In Aug. through Sept. 1955 Mr. C. R. Muir conducted a
capture-recapture study of adult black-kneed capsids in an apple orchard
at the East Malling Agricultural Research Station (Muir 1957). This insect
was important as a predator of orchard insect pests; this capture-recapture
study was just one part of a research program in the orchard. There were k=13
capture occasions. It often took 2 days to complete an "occasion;" treating
the midpoint of 2 days as the time instant of the occasion, capture occasions
averaged 3.5 days apart. The times between occasions varied from 3 to 4 days.
Capsids were individually marked and their sex determined. To analyze the
data Mr. Muir asked the station statistician, Mr. G. M. Jolly, for help.
At that time Mr. Jolly had only a passing acquaintance with capture-recapture.
Out of this request for help, came the now classic 1965 paper of Jolly
which uses summary statistics from the female capsids data as an example.

In the course of work on Burnham et al. (1987) I asked Mr. Jolly for the
full capsids data. He was able to secure for me a copy of the original
field records. The data were decoded and entered to a data file as capture
histories. Table 1 shows a few example records. For just those records,
Table 2 shows the data reformated as subcohort releases and recaptures. The
full data set consists of 252 distinct capture histories at occasion 13.

Using program RELEASE, which implements theory in Burnham et al. (1987),
I tested the fit of the Jolly-Seber model. Chi-square goodness of fit tests
are computed based on terms 2 and 3 in formula (1); these are TESTS 2 and
3 in RELEASE. The Jolly-Seber model is not rejected by either test for
either sex (the overall chi-square goodness of fit is 130.2 on 130 df).
Burnham et al. (1987) also give a test of the equality of the male and
female survival and capture rates (TEST 1; it is based on term 1 in (1)).
That test produces a chi-square of 259.3 on 23 df (P=0). It is clear that
the male and female capsids had different parameters, but separately they
fit a time-specific model.

Table 3 gives the usual summary statistics and the statistics needed
for the distribution of JS_MSS. The data for the females differ some from
what Jolly (1965) used. Upon seeing this, I carefully checked the raw data
and the data entry. There were no errors. The discrepancy cannot be resolved.
However, the parameter estimates under the Jolly-Seber model here and in
Jolly (1965) are very similar.

The Jolly-Seber model in PROC NLIN. Figure 1 shows the necessary code to implement the Jolly-Seber model numerically in PROC NLIN of SAS using the JS_MSS (as per the representation of formulae (2)) as input. The numerical solution converged properly. Figure 2 shows the key output: the sum of loss line (2368.882) is the negative log-likelihood function at its unconstrained minimum; parameter estimates, and their standard errors, agree with results using explicit formulae for these MLE's. Note that p_{12} is set to 1, so S12 is really S12*P13 ($=\emptyset_{12}p_{12}$).

Starting values (lines 350-380 in Fig. 1) must be supplied. The BOUNDS statement was inactivated in this run; a BOUNDS statement is optional. Statements 440-490 compute the current values of λ and τ. Lines 550-560 set up the negative log-likelihood function. When weights are used to get a weighted residual sum of squares for the model, SAS then computes a weighted loss function, $\Sigma(\_LOSS\_*\_WEIGHT\_)$; summation is over the observations in the data set. Thus, it is necessary in line 560 to define _LOSS_ to SAS as the log-likelihood divided by the weight. Also, it is advisable to add code to guard against the case of $r_i=0$ or $m_j=0$ which would otherwise result in an infinite weight and a zero argument to a log function. However, in Fig. 1, I want to show the bare minimum code required.

A critical point not so far mentioned is that there is in PROC NLIN a solution method that does not require analytical partial derivatives: method DUD (Ralston and Jennrich 1978). If analytical partial derivatives (of the model structure) were needed with respect to the parameters, this analysis method would not be practical. This is because it can be very difficult, and time consuming, to get the correct derivatives and to code them. With this approach using SAS one can try different models in the course of a few minutes.

The unconstrained MLE of $\emptyset_2$ is 1.045 (se=0.111). Re-running the job with the bounds statement active (just remove the leading "*" on line 390, Fig. 1) gives, as expected, S2=1.0. Also affected are S1, S3, P2 and P3, but only slightly, e.g. S3 becomes 0.895, compared to 0.874 without the constraint.

Restricted Models for the Female Capsids. To fit a model with a constant capture probability requires only a few changes to the code in Fig. 1. Replace P2 through P12 in the PARAMETERS statement with the single PCON, for the constant capture probability, p. Similarly change the bounds statement, if active bounds are desired, to reference only PCON as 0<PCON<1. Next, add lines 434-436 as shown in Fig. 3 to set each P(I)=PCON (in this SAS code, PI and P(I) are the same variable). Finally, remove the P13=1.0 in line 440. Under the constant p_i model $\emptyset_{12-1}$ and p_{12} are not confounded so the identifiability constraint P13=1.0 is not needed. The treatment of P1 remains arbitrary here although logically PCON would be assumed to also estimate p_1. (P1 shows up in the SAS code only because I have an array for the p_i which is indexed from 1 to 13).

In the analysis of variance table in Fig. 3 the weighted residual SS is asymptotically distributed as a central chi-square (10 df) if the true model is Jolly-Seber with constant p_i. Thus, a quick test of this hypothesis is provided by noting that 8.8 is not at all significantly large for a chi-square on 10 df. In my opinion a better test is the likelihood ratio test; here that test statistic is 2(2373.27-2368.88)=8.78 (10 df). The reduced model fits quite well compared to the full Jolly-Seber model

Let t_i be the time interval from occasion i to i+1. To get a model
with time-constant survival rate set $\emptyset_i=\emptyset**t_i$. Here $\emptyset$ will be a daily
survival rate. Figure 4 shows the changes made to the code in Figure 1 to
get a fully-reduced model, i.e., constant p and $\emptyset$. The t_i for the capsids
study are implicit in Fig. 4. From Fig. 4 this constant p and $\emptyset$ model does
not fit very well compared to the saturated Jolly-Seber model (e.g., a
weighted residual SS of 61.9 on 21 df, whereas the 1% critical level for
a central chi-square on 21 df is 38.9). The likelihood ratio test of this
2 parameter model vs. the 13 parameter model of Fig. 3 is 2(2399.6-2373.9)=
51.4 (11 df). The model with a constant p and varying daily survival rate
is acceptable; a constant p and $\emptyset$ model does not fit the data.

The Male Capsids Data. From Table 3 we see that the male capsids data
are sparse for cohorts 10-12. The lack of recaptures for cohorts 11 and 12
($r_{11}=r_{12}=0.0$) renders p_{11}, p_{12} and $\emptyset_{11}$ nonestimable (division by
zero occurs in an attempt to estimate them). However, there are still 23
estimable parameters in the saturated Jolly-Seber model; we may take them
to be $\emptyset_1$, ..., $\emptyset_{10}$, λ_{11}, λ_{12}, p_2, ..., p_{10}, τ_{11} and τ_{12}.
The SAS code to numerically fit this model is in Fig. 5. Note the
parameterization as S1, ..., S10, LAM11, LAM12, P2, ..., P10, TAU11 and
TAU12. The negative log-likelihood and estimated parameters are in Fig. 6.
Sub-models can be obtained by adding the necessary code and altering the
PARAMETERS and BOUNDS statements in Fig. 5. Alternatively, the SAS code
in Fig. 1 can be used as the starting point for suitably constrained models
because, e.g., if p is constant then $\emptyset_1$ through $\emptyset_{12}$ are all mathe-
matically estimable. For this latter model I get PCON=0.158 (se=0.0168) and
the log-likelihood = -898.552. The likelihood ratio test of the p-constant
model vs the p-varying (saturated) model gives a chi-square = 15.7 (10 df,
P=0.108).
    If there is a reason to suspect a trend in capture rates p_1 to p_9
a logit-linear model can be used. To model the log-odds of the p_i as
linear on occasion (this is reasonable only if occasions are equally
spaced in time) replace P1 to P9 in the PARAMETERS statement by A=-1 B=0.
Then add to Fig. 5 the lines

```
00451  DO J=2 TO 9;
00452     TEMP=A+(B*(J-2));
00453     P{J}=1/(1+EXP(-TEMP));
00454  END;
```

Convergence occurs easily with this model; log-likelihood = -891.311, 16
parameters. This further illustrates the ease of modeling using PROC NLIN.

Joint analysis of female and male data. The flexibility and power of
this analysis method really becomes evident when exploring parsimonious
models for 2, or more, related data sets. One such model will be given by
way of illustration for the male and female data jointly. From separate
analyses, having constant capture rates over time (within sex) seems
tenable. It is also clear that the survival rates should be allowed to
vary by sex and time. However, perhaps female and male survival rates are
proportional. The simplest such model is to have $\emptyset_{mi}=\theta*\emptyset_{fi}$. Figure 7
shows SAS code to fit this model to the male and female data (wherein θ
is denoted SRMDF for "Survival Rate of Male Divided by Female"). The model
is set-up by lines 620 through 720 in Fig. 5. Figure 8 shows output from
this model fitting.

The fit of this joint model is marginal. From the Jolly-Seber models separate by sex, we get the log-likelihood of -3259.57 (=-2368.88-890.69) for fitting a saturated joint model. From Fig. 7 the log-likelihood for the "proportionality" model is -3283.79. The likelihood ratio statistic is 48.44 on 31 df (P=0.024).

Discussion

Complete flexibility is now possible in fitting models to capture-recapture data. The theory which makes this possible is the modeling of the data using conditional (on release) multinomial distributions, thereby getting a simple, explicit and standard type of likelihood for the survival and capture parameters. The practical, comprehensive analysis of capture-recapture data is made possible by the sophisticated computing power now available, both in terms of hardware and software. In particular, excellent derivative-free optimization algorithms are the key to the software power, especially when packaged in a product like SAS. Implementing MLE via iteratively reweighted nonlinear least squares is just a device for using SAS to get the MLE's; it would be more convenient to have an equally flexible package directly computing likelihood inference.

The examples only touch the surface of what is possible. As suggested by Clobert et al. (1987) we could reparameterize on logits. For example, define, to SAS, the "survival" parameters as, say, LS and then set $S(I)=1/(1+EXP(-LS(I)))$ before computing ERATIO. Such logits (LS) may be more normally distributed than the $\hat{p}_i$ and $\hat{\emptyset}_i$. Good confidence intervals remain a concern in capture-recapture (Manly 1984). If the direct logit transform does not solve the problem (I do not think it will), compute likelihood intervals. With current software it takes some effort to get the likelihood intervals, but it can be done and seems to be one of the few justifiable methods when parameters are bounded and the MLE is on a boundary.

The theory and methods here extend easily to capture-recapture models more general than Jolly-Seber and to other cohort survival processes such as bird banding models (Brownie et al. 1985). A first extension of Jolly-Seber is to allow a one-period ahead effect on either p and/or Ø after a release. The model of Sandland and Kirkwood (1981) is such a generalization wherein after a known release at time i the capture probability is $p_{a,i+1}$ at time i+1 and after that (given no capture at time i+1) it is $p_{b,j}$, $j>i+1$. The Sandland and Kirkwood model can be implemented in SAS. Another generalization is to allow age-specific effects if there are identifiable age classes. Maximum likelihood estimation with SAS can be done for the models of Pollock 1981a. The control-treatment release-recapture models in Burnham et al. (1987) can be implemented in SAS. In these examples, the advantage of using the approach given here would be the ability to explore more restricted models, which do not have closed-form MLE's.

An integrated model and analysis that includes Ø, p and population size, N (or recruitment, B) is possible. The theory exists because the distribution of the u_i is known. However, that distribution of initial captures involves binomials where the sample size is itself a parameter. So far it has not been possible to put that part of the problem into PROC NLIN. This entails no loss of statistical efficiency unless one wants to place restrictions (models) on the recruitment process, such as in Crosbie and Manly (1985).

The method outlined here is the last step in a thorough analysis of
a capture-recapture data set. First one should use programs like POPAN
(Arnason and Baniuk 1980), RELEASE (Burnham et al. 1987), and JOLLY or
JOLLYAGE (Brownie et al. 1986, Pollock et al. in prep) to do extensive
goodness of fit testing and comparisons between the standard models
fit by these programs. Once the search for a parsimonius and meaningful
model has focused on a class of models (such as to the time-specific
Jolly-Seber model), then intensive analysis of customized models can
begin. Then PROC NLIN in SAS, as illustrated here, can be very useful.

Acknowledgement

References cited

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

Brownie, C., D. R. Anderson, K. P. Burnham, and D. S. Robson. 1985.
    Statistical inference from band recovery data - a handbook, 2nd
    edition. U.S. Fish and Wildlife Service Resource Publication 156.

Brownie, C., J. E. Hines, and J. D. Nichols. 1986. Constant parameter
    capture-recapture models. Biometrics 42:561-574.

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

Buckland, S. T. 1980. A modified analysis of the Jolly-Seber capture-
    recapture model. Biometrics 36:419-435.

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

Clobert, J., J. D. Lebreton, and D. Allaine. 1987. A general approach to
    survival rate estimation by recaptures or resightings of marked birds.
    Ardea 75:113-142.

Crosbie, S. F., and B. F. J. Manly. 1985. A new approach for parsimonious
    modeling of capture-recapture studies. Biometrics 41:385-398.

Jennrich, R. I., and R. H. Moore. 1975. Maximum likelihood estimation by
    means of nonlinear least squares. Pages 57-65 in American Statistical
    Association 1975 Proceedings of the Statistical Computing Section.
    American Statistical Association, Washington, D. C.

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

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

Cormack, R. M. 1981. Loglinear models for capture-recapture experiments on open populations. Pages 217-235 in R. W. Hiorns and D. Cooke (eds.). The mathematical theory of the dynamics of biological populations. Academic Press, London.

Green, P. J. 1984. Iteratively reweighted least squares maximum likelihood estimation and some robust and resistant alternatives. Journal of the Royal Statistical Society, Series B 46:149-192.

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

Manly, B. F. J. 1984. Obtaining confidence limits on parameters of the Jolly-Seber model for capture-recapture data. Biometrics 40:749-758.

Muir, R. C. 1957. On the application of the capture-recapture method to an orchard population of Blepharidopterus angulatatus (Fall.) (Hemiptera-Heteroptera, Miridae). East Malling Annual Report 1957:140-147.

Pollock, K. H. 1975. A K-sample tag-recapture model allowing for unequal survival and catchability. Biometrika 62:577-583.

Pollock, K. H. 1981a. Capture-recapture models allowing for age-dependent survival and capture rates. Biometrics 37:521-529.

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

Pollock, K. H., J. E. Hines, and J. D. Nichols. 1985. Goodness-of-fit tests for open capture-recapture models. Biometrics 41:399-410.

Pollock, K. H., J. D. Nichols, C. Brownie, and J. E. Hines. (in prep). Statistical inference for capture-recapture experiments. Wildlife Monographs, The Wildlife Society, Washington, D. C.

Ralston, M. L., and R. I. Jennrich. 1978. Dud, a derivative-free algorithm for nonlinear least squares. Technometrics 20(1):7-14.

Robson, D. S. 1969. Mark-recapture methods of population estimation. Pages 129-140 in N. L. Johnson and H. Smith, Jr. (eds.). New developments in survey sampling. Wiley Interscience, New York.

SAS Institute Inc. 1985. SAS user's guide: statistics, version 5 edition. SAS Institute Inc., Cary, NC 27511.

Sandland, R. L., and P. Kirkwood. 1981. Estimation of survival in marked populations with possibly dependent sighting probabilities. Biometrika 68:531-541.

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

White, G. C. 1983. Numerical estimation of survival rates from band-recovery and biotelemetry data. The Journal of Wildlife Management 47:716-728.

Table 1. Example capture histories for the female (X_{fh}) and male (X_{mh}) black-kneed capsids data of Muir (1957); the entire data set, as complete capture histories, has 252 distinct histories.

Capture History, h	X_{fh}	X_{mh}
0000000000010	36	14
0000000001100	4	1
0000000101000	7	1
0000001000000	74	58
0000100110000	2	0
0000111001100	0	-1
0100110000000	0	2
1000000000000	30	97
1001000000000	3	5

Table 2. The full m-array representation of female captures from the nine example capture histories data in Table 1; capture history is conditional on release occasion, i; R_{fih} denotes number released; m_{fijh} denotes (first) recaptures, at occasion j, after release at occasion i.

Capture History, h(i)	i	R_{fih}	Recaptures, m_{fijh}, at occasion j											
			j=2	3	4	5	6	7	8	9	10	11	12	13
1	1	33	0	0	3	0	0	0	0	0	0	0	0	0
1001	4	3	.	.	.	0	0	0	0	0	0	0	0	0
00001	5	2	.	.	.	.	0	0	2	0	0	0	0	0
0000001	7	74	.	.	.	.	.	.	0	0	0	0	0	0
00000001	8	7	.	.	.	.	.	.	.	0	7	0	0	0
00001001	8	2	.	.	.	.	.	.	.	2	0	0	0	0
000010011	9	2	.	.	.	.	.	.	.	.	0	0	0	0
0000000001	10	4	.	.	.	.	.	.	.	.	.	4	0	0
0000000101	10	7	.	.	.	.	.	.	.	.	.	0	0	0
00000000011	11	4	.	.	.	.	.	.	.	.	.	.	0	0
000000000001	12	36	.	.	.	.	.	.	.	.	.	.	.	0

Table 3. Summary statistics as an (reduced) m-array for female and male black-kneed capsids from the study of Muir (1957); under the Jolly-Seber model (Jolly 1965, Seber 1965) the r_i and m_j are minimal sufficient statistics, i=2, ... , 13, j=2, ... , 12.

i	R_{fi}	m_{fij}												r_{fi}
		j=2	3	4	5	6	7	8	9	10	11	12	13	
1	54	10	3	5	2	2	1	0	0	0	1	0	0	24
2	143		36	18	8	4	7	4	2	0	2	1	1	83
3	166			33	14	8	5	0	4	1	3	3	0	71
4	202				30	20	10	3	2	2	1	1	2	71
5	185					32	22	9	7	2	0	1	3	76
6	196						52	17	10	4	4	2	3	92
7	230							42	27	17	8	6	2	102
8	164								49	21	11	4	10	95
9	160									33	16	11	9	69
10	122										28	15	12	55
11	117											26	18	44
12	118												35	35
	m_{fj}	10	39	56	54	66	97	75	101	80	74	70	95	
	T_{fj}	24	97	129	144	166	192	197	217	185	160	130	95	

i	R_{mi}	m_{mij}												r_{mi}
		j=2	3	4	5	6	7	8	9	10	11	12	13	
1	134	19	10	6	1	0	0	0	0	1	0	0	0	37
2	156		24	8	9	3	1	0	1	1	0	1	0	48
3	173			24	2	5	4	0	0	0	0	0	0	35
4	190				23	6	1	4	0	0	1	0	0	35
5	171					15	7	2	3	2	0	0	0	29
6	140						17	5	2	0	0	0	1	25
7	93							4	1	3	1	0	0	9
8	84								6	2	1	0	0	9
9	48									3	2	1	0	6
10	24										3	1	1	5
11	24											0	0	0
12	17												0	0
	m_{mj}	19	34	38	35	29	30	15	13	12	8	3	2	
	T_{mj}	37	66	67	64	58	54	33	27	20	13	5	2	

```
00010   DATA;
00050       INPUT K INDEX$ I SIZE COUNT;
00060       RATIO=COUNT/SIZE;
00070   CARDS;
00080   13 R  1   54 24
  ...     .  .  .    .  .    (data for r_i given R_i, i=2, ..., 11 not shown)
00190   13 R 12 118 35
00200   13 T  2  24 10
  ...     .  .  .    .  .    (data for m_j given T_j, j=3, ..., 11 not shown)
00300   13 T 12 130 70
00320   PROC NLIN METHOD=DUD SIGSQ=1.;
00330   ARRAY S{12} S1-S12;
00340   ARRAY P{13} P1-P13;
00350   PARAMETERS S1=.8   S2=.8   S3=.8    S4 =.8   S5 =.8   S6 =.8
00360                 S7=.8   S8=.8   S9=.8    S10=.8   S11=.8   S12=.8
00370                 P2=.25 P3=.25 P4=.25   P5 =.25 P6 =.25 P7 =.25
00380                 P8=.25 P9=.25 P10=.25 P11=.25 P12=.25;
00390   * BOUNDS   0<S1<1, 0<S2<1, 0<S3 <1, 0<S4 <1, 0<S5 <1, 0<S6 <1,
00400                 0<S7<1, 0<S8<1, 0<S9 <1, 0<S10<1, 0<S11<1, 0<S12<1,
00410                 0<P2<1, 0<P3<1, 0<P4 <1, 0<P5 <1, 0<P6 <1, 0<P7 <1,
00420                 0<P8<1, 0<P9<1, 0<P10<1, 0<P11<1, 0<P12<1;
00430
00440   P1=0;  P13=1.0;  LAMBDA=0.0;
00450   DO J=K-1 TO I BY -1;
00460       LAMBDA=S{J}*(P{J+1}+((1-P{J+1})*LAMBDA));
00470   END;
00480   IF INDEX='R'  THEN ERATIO=LAMBDA;
00490       ELSE ERATIO=P{I}/(P{I}+((1.-P{I})*LAMBDA));
00500
00510   MODEL RATIO = ERATIO;
00520
00530   _WEIGHT_=SIZE/(ERATIO*(1.-ERATIO));
00540
00550   TEMP=(COUNT*LOG(ERATIO))+((SIZE-COUNT)*LOG(1-ERATIO));
00560   _LOSS_=-TEMP/_WEIGHT_;
```

Fig. 1. The minimum SAS code needed to compute the maximum likelihood
estimates of Ø and p for the Jolly-Seber Model; the input data used (not
all lines of it are shown) are the minimal sufficient statistics for the
female black-kneed capsids (Muir 1957, Jolly 1965) under the Jolly-Seber
model; line numbers and blank lines were added to aid in presentation and
discussion of the code.

NON-LINEAR LEAST SQUARES SUMMARY STATISTICS, DEPENDENT VARIABLE: RATIO

SOURCE	DF	WEIGHTED SS	WEIGHTED MS
REGRESSION	23	2841.4297359	123.5404233
RESIDUAL	0	0.0000000	0.0000000
UNCORRECTED TOTAL	23	2841.4297359	
(CORRECTED TOTAL)	22	63.4524645	
SUM OF LOSS		2368.8824582	(= - log-likelihood)

PARAMETER	ESTIMATE	ASYMPTOTIC STD. ERROR
S1	0.631860776	0.10628284876
S2	1.044788956	0.11099508876
S3	0.874287849	0.10806684315
S4	0.666549961	0.08178264958
S5	0.690566732	0.07490871725
S6	0.760837669	0.07271115787
S7	0.642954617	0.05809526318
S8	0.987653795	0.09616624802
S9	0.729416467	0.08765730567
S10	0.852843232	0.11839846444
S11	0.787677281	0.13115340975
S12	0.296610169	0.04204845639
P2	0.293079096	0.08705161056
P3	0.223360490	0.03915847541
P4	0.212370473	0.03374631217
P5	0.197745013	0.03067532532
P6	0.236522281	0.03175587944
P7	0.311680947	0.03479260099
P8	0.262595364	0.03120009381
P9	0.272983664	0.03250080282
P10	0.255665304	0.03405035730
P11	0.244481160	0.03662964003
P12	0.257082895	0.04316148561

NOTE: STANDARD ERRORS COMPUTED USING SIGSQ = 1

Fig. 2. Key output from the SAS code in Fig. 1; the iterative procedure to find the estimates converged; the estimates and their standard errors are identical to results produced with the closed-form formulae for the maximum likelihood results under the Jolly-Seber model.

```
  ...
  00350   PARAMETERS  S1=.8   S2=.8   S3=.8    S4 =.8   S5 =.8   S6 =.8
  00360               S7=.8   S8=.8   S9=.8    S10=.8   S11=.8   S12=.8
>00370               PCON=.2;
>00380
  00390   * BOUNDS   0<S1<1, 0<S2<1, 0<S3 <1, 0<S4 <1, 0<S5 <1, 0<S6 <1,
  00400               0<S7<1, 0<S8<1, 0<S9 <1, 0<S10<1, 0<S11<1, 0<S12<1,
>00410               0<PCON<1;
>00420
  00430
>00434   DO J=2 TO 13;
>00435       P{J}=PCON;
>00436   END;                              '
  00440   P1=0;            LAMBDA=0.0;
  ...
```

SOURCE	DF	WEIGHTED SS	WEIGHTED MS
REGRESSION	13	2828.1104845	217.5469603
RESIDUAL	10	8.7620265	0.8762027
UNCORRECTED TOTAL	23	2836.8725110	
(CORRECTED TOTAL)	22	63.2371756	
SUM OF LOSS		2373.2690762	(= - log-likelihood)

PARAMETER	ESTIMATE	ASYMPTOTIC STD. ERROR
S1	0.650706501	0.10328666951
S2	0.996329491	0.08522365006
S3	0.830993926	0.07626432419
S4	0.637195561	0.05892305338
S5	0.718919377	0.06085674848
S6	0.847713571	0.06603788121
S7	0.610036053	0.04828335586
S8	1.018955533	0.07688920410
S9	0.709766800	0.06446824831
S10	0.834852186	0.08137611661
S11	0.809092992	0.08753249145
S12	1.167308533	0.13152248360
PCON	0.250936690	0.01086588653

Fig. 3. Altered and added code (lines indicated by ">") to change from the Jolly-Seber model code in Fig. 1 to a model with constant capture rate PCON; also shown is a key part of the SAS output from running this model (parameter estimates were not bounded).

```
>00350   PARAMETERS SDAY=.8 PCON=.2;
>00390   * BOUNDS    0<SDAY<1, 0<PCON<1;
>00431   S1=SDAY**3.5;  S2=SDAY**3.0;  S3=SDAY**4.0;  S4=SDAY**3.0;
>00432   S5=SDAY**4.0;  S6=SDAY**3.0;  S7=SDAY**3.5;  S8=SDAY**3.5;
>00433   S9=SDAY**3.5; S10=SDAY**3.0; S11=SDAY**4.0; S12=SDAY**3.0;
>00434   DO J=2 TO 13;
>00435      P(J)=PCON;
>00436   END;
>00440   P1=0;           LAMBDA=0.0;
```

SOURCE	DF	WEIGHTED SS	WEIGHTED MS
REGRESSION	2	2779.5777020	1389.7888510
RESIDUAL	21	61.8774339	2.9465445
UNCORRECTED TOTAL	23	2841.4551359	
(CORRECTED TOTAL)	22	65.6963314	
SUM OF LOSS		2399.5911136	(= - log-likelihood)

PARAMETER	ESTIMATE	ASYMPTOTIC STD. ERROR
SDAY	0.9313858542	0.00349320369
PCON	0.2574917154	0.01109920572

Fig. 4. To get a fully reduced "Jolly-Seber model," i.e. constant p (PCON)
and constant daily survival rate (SDAY), replace lines 350-440 in Fig. 1
with the SAS code shown here; also shown here are key parts of the SAS
output for this fully reduced model.

```
00010  DATA;
00050     INPUT K INDEX$ I SIZE COUNT;
00060     RATIO=COUNT/SIZE;
00080  CARDS;
00090  13 R  1   134  37
...    . . .   .    .
00310  13 T 12    5   3
00330  PROC NLIN METHOD=DUD SIGSQ=1.;
00340  ARRAY S(12) S1-S12;
00350  ARRAY P(13) P1-P13;
00360  PARAMETERS   S1=.5  S2=.5  S3=.5   S4=.5  S5=.5  S6=.5  S7=.5
00370               S8=.5  S9=.5 S10=.5  LAM11=.0  LAM12=.0
00380               P2=.2  P3=.2  P4=.2  P5=.2  P6=.2  P7=.2
00390               P8=.2  P9=.2 P10=.2  TAU11=.6  TAU12=.6;
00400  BOUNDS       0<S1<1, 0<S2<1, 0<S3<1, 0<S4<1, 0<S5 <1,
00410               0<S6<1, 0<S7<1, 0<S8<1, 0<S9<1, 0<S10<1,
00420               0<P2<1, 0<P3<1, 0<P4<1, 0<P5<1, 0<P6<1, 0<P7<1,
00430               0<P8<1, 0<P9<1, 0<P10<1, 0<TAU11<1, 0<TAU12<1;
00440
00450  P1=0; P11=1.0; P12=1.0; P13=1.0; S11=0.0; S12=0.0;
00460  LAMBDA=0.0;
00470  DO J=K-1 TO I BY -1;
00480     LAMBDA=S(J)*(P(J+1)+((1-P(J+1))*LAMBDA));
00490  END;
00500  IF INDEX='R' THEN DO;
00510     ERATIO=LAMBDA;
00520     IF I=11 THEN ERATIO=LAM11;
00530     IF I=12 THEN ERATIO=LAM12;
00540     END;
00550  ELSE DO;
00560     ERATIO=P(I)/(P(I)+((1-P(I))*LAMBDA));
00570     IF I=11 THEN ERATIO=TAU11;
00580     IF I=12 THEN ERATIO=TAU12;
00590     END;
00600
00610  MODEL RATIO = ERATIO;
00620
00630  IF ERATIO>0 AND ERATIO<1 THEN
00640       _WEIGHT_=SIZE/(ERATIO*(1.-ERATIO));
00650       ELSE _WEIGHT_=1;
00660
00670  IF ERATIO>0 THEN AA=LOG(ERATIO);    ELSE AA=0;
00680  IF ERATIO<1 THEN BB=LOG(1.-ERATIO); ELSE BB=0;
00690  TEMP=(COUNT*AA)+((SIZE-COUNT)*BB);
00700  _LOSS_=-TEMP/_WEIGHT_;
```

Fig. 5. SAS code used to get the maximum likelihood estimates of $\emptyset$ and p
for the Jolly-Seber model with the male capsids data; the input data
are the minimal sufficient statistics (JS_MSS).

SOURCE	DF	WEIGHTED SS	WEIGHTED MS
REGRESSION	23	816.38777618	35.49512070
RESIDUAL	0	0.00000000	0.00000000
UNCORRECTED TOTAL	23	816.38777618	
(CORRECTED TOTAL)	22	218.42040425	
SUM OF LOSS		890.69397659	(= - log-likelihood)

PARAMETER	ESTIMATE	ASYMPTOTIC STD. ERROR
S1	0.5783582091	0.10908135019
S2	0.8959040961	0.17193372200
S3	0.5901130189	0.12704913011
S4	0.5929276329	0.13521889098
S5	0.5596491265	0.13785027199
S6	0.9193121581	0.32861992568
S7	0.5366569128	0.24061628362
S8	0.4960317468	0.24170144061
S9	0.3150000003	0.16217537204
S10	0.2083333333	0.08289816935
LAM11	0.0000000000	0.00000000111
LAM12	0.0000000000	0.00000000733
P2	0.2451612902	0.06481033300
P3	0.1769253642	0.04209747325
P4	0.1944444443	0.04544179148
P5	0.1699029118	0.04272007267
P6	0.1515151490	0.04103969646
P7	0.1079136717	0.04032621559
P8	0.0819672042	0.03550586683
P9	0.1039999996	0.05055056786
P10	0.2380952376	0.10988046879
TAU11	0.6153846154	0.13493200297
TAU12	0.6000000000	0.21908902300

Fig. 6. Key SAS output from fitting the saturated Jolly-Seber model to the male capsids data using the SAS code in Fig. 5.

```
00010  DATA;
00050     INPUT SEX$  K INDEX$ I SIZE COUNT;
00060     RATIO=COUNT/SIZE;
00070  CARDS;
00080  F 13 R 1    54  24
 ...   .  . . .     .   .
00520  M 13 T 12   5   3
00530  PROC NLIN METHOD=DUD SIGSQ=1.;
00540  ARRAY S(12) S1-S12;
00550  ARRAY P(13) P1-P13;
00560  PARAMETERS S1=.8 S2=.8 S3=.8  S4=.8  S5=.8  S6=.8
00570             S7=.8 S8=.8 S9=.8 S10=.8 S11=.8 S12=.8
00580             PMALE=.15 PFEMALE=.25 SRMDF=.75;
00590  BOUNDS     0<S1, 0<S2, 0<S3, 0<S4, 0<S5, 0<S6, 0<S7, 0<S8, 0<S9,
00600             0<S10, 0<S11, 0<S12, 0<PMALE, 0<PFEMALE, 0<SRMDF;
00610
00620  IF SEX='M' THEN PC=PMALE;   ELSE PC=PFEMALE;
00630  DO J=1 TO 13;
00640     P(J)=PC;
00650  END;
00660  IF SEX='M' THEN FAC=SRMDF;   ELSE FAC=1.;
00670  LAMBDA=0.0;
00680  DO J=K-1 TO I BY -1;
00690     LAMBDA=(S(J)*FAC)*(P(J+1)+((1-P(J+1))*LAMBDA));
00700  END;
00710  IF INDEX='R' THEN ERATIO=LAMBDA;
00720     ELSE ERATIO=P(I)/(P(I)+((1.-P(I))*LAMBDA));
00730
00740  MODEL RATIO = ERATIO;
00750
00760  IF ERATIO>0 AND ERATIO<1 THEN
00770     _WEIGHT_=SIZE/(ERATIO*(1.-ERATIO));
00780     ELSE _WEIGHT_=1.;
00790
00800  IF ERATIO>0 THEN AA=LOG(ERATIO);    ELSE AA=0.0;
00810  IF ERATIO<1 THEN BB=LOG(1.-ERATIO); ELSE BB=0.0;
00820  TEMP=(COUNT*AA)+((SIZE-COUNT)*BB);
00830  _LOSS_=-TEMP/_WEIGHT_;
00840
00850  OUTPUT OUT=TWO P=ERATIO R=RESIDS;
00860
00870  DATA; SET TWO;        /* COMPUTE SQUARED, NORMALIZED RESIDUALS  */
00880     CHISQ=RATIO - ERATIO;
00890     CHISQ=CHISQ*CHISQ*SIZE/(ERATIO*(1-ERATIO));
00900  PROC PRINT;
```

Fig. 7. SAS code used to get the maximum likelihood estimates of Ø and p
    under the proportionality model for the joint analysis of the female
    and male capsids data; input data are the separate minimal sufficient
    statistics (JS_MSS) for females and males.

SOURCE	DF	WEIGHTED SS	WEIGHTED MS
REGRESSION	15	3630.4138029	242.0275869
RESIDUAL	31	40.8056670	1.3163118
UNCORRECTED TOTAL	46	3671.2194699	
(CORRECTED TOTAL)	45	502.4760777	
SUM OF LOSS		3283.7880001	(= - log-likelihood)

PARAMETER	ESTIMATE	ASYMPTOTIC STD. ERROR
S1	0.765409878	0.08313774439
S2	1.030986631	0.07375562610
S3	0.832016498	0.06484330243
S4	0.664863549	0.05216473620
S5	0.712054069	0.05354383095
S6	0.856763266	0.06125679050
S7	0.605919798	0.04548736195
S8	0.971788533	0.07106438045
S9	0.711756066	0.06160583043
S10	0.852055508	0.08010304218
S11	0.782686338	0.08296160196
S12	1.132987942	0.12626904575
PMALE	0.166778450	0.01567718261
PFEMALE	0.249997448	0.01084282574
SRMDF	0.747500908	0.02896779703

Fig. 8. Key SAS output from fitting the proportionality model to the female and male capsids data; unconstrained maximum likelihood estimates are shown for S1-S12, the survival rates of the female capsids, SRMDF, the ratio of male to female survival rates (assumed here to be constant over time), and the constant capture probabilities PMALE and PFEMALE.