

# CHAPTER 8

## OPEN MODELS

### Basic Concepts

Capture-recapture estimation techniques are usually classified as belonging to one of two general categories—those appropriate for (demographically) closed populations and those appropriate for (demographically) “open” populations (*Arnason and Baniuk 1980; Seber 1981*). An open population is defined as one in which the processes of birth, death, and migration are allowed to operate. Just as there is more than one model for the closed-population case, there is a variety of open models for capture-recapture studies. However, the essential elements of open models are the types of parameters they involve. In the now classic Jolly-Seber model (see *Seber 1973:196-232*), the parameters are population size, survival rate, recruitment, and capture probability. In closed models, by way of comparison, one assumes that during the capture study, the animals have a survival rate of one and that the recruitment into the population is zero.

Field aspects of capture-recapture studies are not necessarily different when open models are used, except that the time period may be long relative to the population dynamics of the target species. For example, instead of trapping once a day for 6 days, one may trap only once a month or even once a year for studies on most mammals, birds, and fish. In many insect capture studies, open models are necessary even when capture occasions are on consecutive days.

The population size of open models varies with each capture occasion; thus, the one population size parameter  $N$  of the closed model becomes population sizes  $N_1, \dots, N_t$  for the  $t$  capture occasions of the open model. The survival rate parameters are  $S_1, \dots, S_{t-1}$ , where  $S_j$  represents an individual's survival rate, in the population exposed to trapping, between occasions  $j$  and  $j + 1$ . The survival rate applies to those animals alive in the trapped population just after the  $j^{\text{th}}$  trapping occasion. The recruitment parameters are  $B_1, \dots, B_{t-1}$ . Here,  $B_j$  is the total influx of new individuals into the population being trapped, between trapping occasions  $j$  and  $j + 1$ ; as such,  $B_j$  is not a rate. Finally, there are the capture probabilities  $p_1, \dots, p_t$ ; all the  $N_j$  animals in the population on the  $j^{\text{th}}$  capture occasion are assumed to have capture probability  $p_j$ . The Jolly-Seber model thus allows time variation in capture probabilities, but no variation due to behavioral response or heterogeneity. It is thus the open version of Model  $M_t$ .

Interrelationships exist among these parameters—primarily,  $N_{j+1} = N_j S_j + B_j$ , which simply indicates that the population size at occasion  $j + 1$  is the number of survivors ( $N_j S_j$ ) from the  $j^{\text{th}}$  occasion plus the recruits ( $B_j$ ) during this time period. Also, some parameters cannot be estimated for the first or last occasions—for example,  $N_1$ ,  $N_t$ , and  $S_{t-1}$ . Although both of these factors reduce the real (as opposed to apparent) number of parameters, open models do have more parameters than their closed counterparts. And the more realistic the model, the more parameters it has.

An additional parameter, totally derived from the survival rates and data, is  $M_j$ , the number of marked animals still alive just before the  $j^{\text{th}}$  capture occasion. This parameter is the key to estimation in the open models, but it is of no intrinsic biological interest.

Because the processes of population dynamics are always operating at least to some degree in natural populations, one might ask why so much attention has been given to capture-recapture data analysis methods based on models closed to these processes. The answer is familiar to statisticians and is becoming familiar to biologists as they gain experience in the construction and manipulation of mathematical models: as the model for a capture-recapture experiment is made increasingly realistic by the introduction of more parameters, such as time- and age-specific survival rates and birth rates, each individual parameter is estimated with less and less precision. Ultimately, the model has more parameters than data and becomes so general that no parameter estimates can be produced from it.

*Cormack (1979)* makes the valuable analogy that the data collected from an experiment contain only a fixed amount of information that can be used to estimate parameters of interest. As the number of parameters increases, this information is spread thinner and thinner among the parameters, and finally results in very little information about any individual parameter. Thus, because estimation techniques for closed populations are generally based on far fewer parameters than those for open populations the techniques can provide more precise estimates of population size  $N$  if the closure assumption is valid.

Of course, there are situations where open models must be used, but there are also situations where closed models should be used. When closed models are used, population size  $N$  is practically constant during the study and only one estimate of  $N$  is necessary. Use of an open model gives several  $(t - 2)$  separate estimates, which then must be combined into a single estimate. Before discussing the advantages, differences, and tradeoffs of closed versus open models, we present a brief synopsis of the state of the art of analytical techniques for capture data from open populations.

## State of the Art

Like the closed-population models discussed in Chapter 3, open models can vary widely in their generality, and as a result many methods and models have been introduced into the literature. (Note that open removal models are impossible.) Although there were earlier efforts to deal with the processes of birth, death, and migration in capture studies (*Fisher and Ford 1947; N. T. J. Bailey 1951; Leslie and Chitty 1951; Leslie 1952; Leslie et al. 1953*), the appropriate beginning point for our purposes is the appearance of the Jolly-Seber model, published independently by *Jolly (1965)* and *Seber (1965)*. The basic model allows birth, death, immigration, and permanent emigration to occur during the experiment.

The following points concerning the Jolly-Seber method must be made clear.

1. The estimated "survival" rate is really survival in the population at risk of capture. An animal fails to survive if it dies or emigrates. If  $D_j$  is the death rate during the time between the  $j^{\text{th}}$  and  $j + 1^{\text{th}}$  trapping occasions and  $E_j$  is the emigration rate during this time, the survival rate parameter is the product  $S_j = (1 - D_j)(1 - E_j)$ . If the emigration rate is large,  $\hat{S}_j$  will severely underestimate the true survival rate of the species under study. The true survival rate ( $= 1 - D_j$ ) can be estimated if one validly assumes that the emigration rate ( $E_j$ ) is zero.

Similarly, birth and immigration cannot be estimated separately. The influx or dilution parameter  $B_j$  is the sum of the number of recruits, due to reproduction in the population under study, and the number of immigrants, which are not produced by the population under study. In practice, a separate estimate of the number of immigrants is possible if these individuals are much older (bigger) than the locally generated recruits. However, the Jolly-Seber model applies to only one age class, which is typically adults, and no age effects on parameters are allowed. Thus, "recruitment" means entry into the adult segment of the population, and distinguishing whether recruits are from reproduction of the population being trapped or whether they are immigrants may be impossible. \*Figure 8.1 shows the fully open model and special cases that allow only influx or only outflux.

2. Parameters involving influx, outflux, and catchability are allowed to vary among sampling occasions, but not among animals. For example, survival rate may change from  $S_j$  between the  $j^{\text{th}}$  and the  $(j + 1)^{\text{th}}$  occasions to  $S_{j+1}$  between the  $(j + 1)^{\text{th}}$  and the  $(j + 2)^{\text{th}}$  occasions, but during each of those time periods, all members of the population are assumed to have the same probability of survival. Similarly, on the  $j^{\text{th}}$  sampling occasion, all  $N_j$  animals then in the population are assumed to have the same probability of capture  $p_j$ . (This capture probability structure corresponds to Model  $M_t$  of Chapter 3.)

3. As *Cormack (1979)* points out, the estimates of survival, birth, and probability of capture are optimal only if the parameters really do differ between sampling occasions. Thus, for example, if the biologist wants to assume that the survival rate remains constant between all sampling periods, the optimal estimate of this single survival rate parameter requires a substantially more complex estimation method.

The mathematical notation found in the publications by *Jolly (1965)*, *Pollock (1975)*, *Robson (1969)*, and *Seber (1965)* is somewhat intimidating. This being the case, it is worthwhile to note that the general problem is solved if the biologist can estimate two quantities for each sampling occasion  $j$ : the number of marked animals  $M_j$  alive in the population and the capture probability  $p_j$ . The problem is a little more complicated in the extensions proposed by *Robson (1969)* and *Pollock (1975)*, because at each sampling occasion  $M_j$  and  $p_j$  may differ over some subclasses of



George M. Jolly

George Jolly's interest in capture-recapture studies began while he was at the East Malling (Fruit) Research Station. There he encountered an entomologist using capture-recapture methods on a local population of an orchard predator. He began by studying the research papers by J. N. Darroch and by P. H. Leslie and his colleagues, and the work resulted in the well-known, general model for open populations published in 1965. That, of course, was before the computer era, when mathematically involved solutions had not yet acquired the respectability they enjoy today.

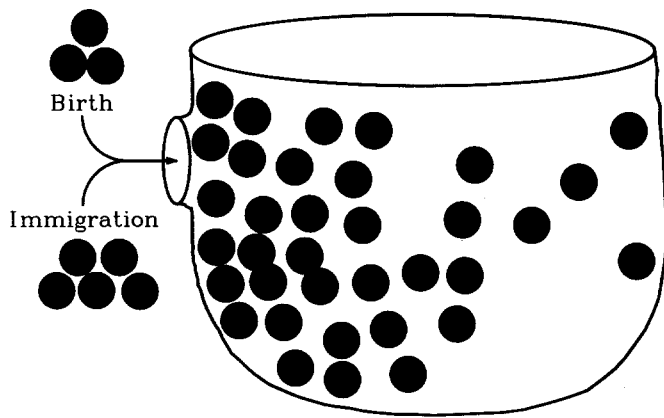
Jolly completed a mathematics degree at Aberdeen, Scotland, and studied statistics at Edinburgh University with a view to agricultural research. He then worked for 5 years at Rothamsted Experimental Station under Frank Yates. He has been with the Agricultural Research Council for many years.

His recent work has been in developing estimation methods allowing survival to remain constant over the study period. He believes that the full potential of capture-recapture methods is difficult to assess until further models, more complex but more realistic biologically, have been developed and studied. Although many estimators must be found numerically, Jolly feels it is helpful when formulas can be expressed in easily interpretable form. He believes it is unfortunate that some biologists cling to the out-dated and inefficient techniques still appearing in publications. (Photograph taken in mid-1960s.)

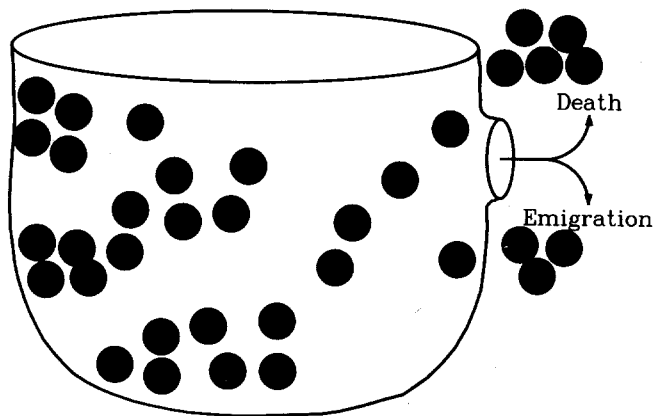
the population, such as those defined by previous capture history. Still, the essentials of the estimation problem for open models are related to the estimation of these two types of parameter sets: the number of surviving marked animals and the capture probabilities.

A subtle, but serious, problem arises with capture studies if simple correlation-regression analyses are applied to the estimated parameters in an attempt to infer anything about the underlying processes of population dynamics. The application is invalid primarily because both the estimators  $\hat{N}_2, \dots, \hat{N}_{t-1}$  and the estimators  $\hat{S}_1, \hat{S}_2, \dots, \hat{S}_{t-2}$  have very strong sampling correlations among themselves. The meaning of the sampling correlations is hard to make clear without presenting technical material beyond the intended level of this primer. However, the basic idea is that because all of the parameter estimates are computed from one common set of data, spurious relations—that is, relations that have nothing to do with the true dynamics of the population—are forced to appear among parameter estimates. As a simple analogy, if we have a set of data,  $x_1, \dots, x_n$ , define a new variable,  $y_i = 1/x_i$ , and then look at the correlation (relation) of  $x$  to  $y$ , we will find a strong, negative correlation. It is wrong to interpret this relation as implying anything about a physical or biological process underlying the “variables”  $x$  and  $y$ . This mistaken approach to investigating population processes is, unfortunately, encountered often in the literature pertaining to ecology. See, for example, *Eberhardt (1970)*; *Anderson and Burnham (1976:13-15, 31, 34-39)*; *Brownie et al. (1978:177-179)*.

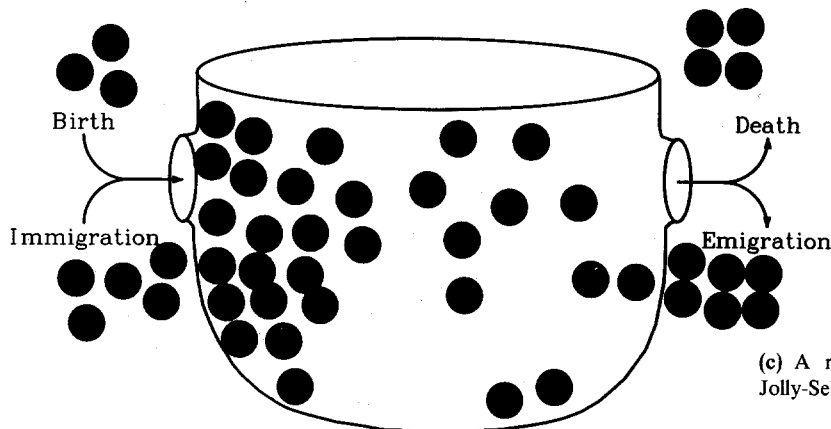
The interrelations of *estimated* parameters are spurious as regards any existing relations between the *true* parameters. As a result, we cannot validly explore population dynamics, such as density dependence of (true) survival rates, by the use of simple regression techniques that depend on estimates of survival rates and population sizes from capture-recapture studies. This is an important limitation of the use of



(a) A representation of the Jolly-Seber model with birth and immigration, but no losses: survival rate is assumed to be 1 and emigration rate zero.  $N = 35$  in this example.



(b) A representation of the Jolly-Seber model with death and emigration, but no gains: birth and immigration rates are assumed to be zero.  $N = 27$  in this example.



(c) A representation of the (demographically) fully open Jolly-Seber model.  $N = 31$  in this example.

\*Fig. 8.1. Open models allow the assumption of demographic closure to be relaxed. The concept of geographic closure is still required (something analogous to the sides of the glass container is needed), because without it, the population size parameter  $N$  has no well-defined meaning. Jolly (1965) and others have developed models for the special case represented by (a) and (b). However, birth and immigration are confounded in (a) as are death and emigration in (b). The rates of the two processes shown in each figure cannot be estimated separately. For example, the estimate of the "influx" parameter (a) is biologically meaningful only if one process, say, immigration, is zero. If immigration is zero, this parameter can be interpreted as an estimate of reproductive recruitment. When this influx parameter is estimated, it frequently is called a dilution rate. (c) represents the usual fully open population model for which several elegant methods of analysis have been developed (Jolly 1965; Seber 1965; Robson 1969; Pollock 1975). In the open models,  $\hat{N}_j$  is an estimator for the size of the population at the  $j^{\text{th}}$  sampling occasion. Of course,  $N_j$  changes through time as animals come and go.

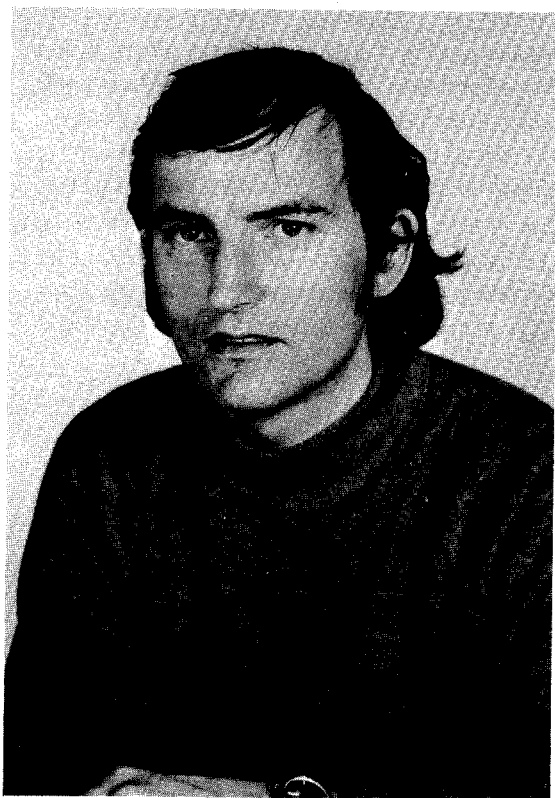
capture studies. As yet, the proper use of capture-recapture data to explore population processes has not been developed.

Clearly, the Jolly-Seber method can be too general (survival rates, in fact, may not vary over time), too specific (no age effects are allowed for), or not appropriate, depending upon the particular experimental situation. A more restrictive model in which survival rate is constant over time has been considered by *Jolly (1979)*, who provided ML equations for the parameters of interest. Direct solutions have not yet been developed. *Crosbie (1979)* considered a "modified" Jolly-Seber model, which accounts for the animals that enter and exit the population between two successive sampling occasions and thus have no chance of being marked. He then developed a sequence of models formed by making different assumptions, generally more restrictive than Jolly-Seber assumptions, concerning capture probabilities, survival rates, and birth distributions, and created an elaborate computer program that attempts to choose the most parsimonious model for the data at hand. In concept, this approach is very similar to that used by *Otis et al. (1978)* in their treatment of closed models. In addition, there are published models allowing different classes of animals in the population to have different probabilities of capture and survival. *Robson (1969)* has expanded the basic model so that the first capture is allowed to affect the animal's survival rate for either one or two subsequent time intervals. *Cormack (1972)* has investigated the case in which the first capture affects both the animal's probability of capture and its survival for the subsequent time period. The most general capture-recapture model for an open population is that formulated by *Pollock (1975)*. In his model, animals with different capture histories may have different survival and capture probabilities in subsequent sampling occasions, but he actually derives estimates only for the two cases in which the first capture affects survivability for  $\ell$  subsequent time periods and catchability for either  $\ell$  or  $\ell - 1$  subsequent periods. The fact that estimators for other forms of heterogeneous probabilities are theoretically possible, but computationally very difficult explains why Pollock's models, as well as those of Robson and Cormack, have not been put to much practical use. We hope that the availability of versatile, easy-to-use computer packages, like the one developed by *Arnason and Baniuk (1978, 1980)*, will facilitate practical implementation of these methods.

## Discussion

Now that we have developed a feel for what the open-model capture-recapture techniques have to offer, we can discuss when and how the two general classes (open and closed) of techniques compete with each other. First, however, we must deal with the common misconception that open-population models avoid the concept of geographic closure. Although the assumption of geographic closure is often approximate at best, both closed and open models suffer if it is violated. To illustrate this point, consider an experiment in which a grid of live traps has been placed in a large planting of sugar cane for the purpose of estimating the size of the cotton rat population using the field. Clearly, a closed-population model used for this experiment involves the parameter  $N$ , the absolute number of rats in the "population." A biologist may feel uncomfortable with this concept, because he knows that rats can come and go across irrigation ditches surrounding the cane field. Consider, however, that any open-population model, for example, the general Jolly-Seber model, postulates the existence of a population of  $N_j$  individuals present on the  $j^{\text{th}}$  sampling occasion. When a model contains a parameter representing the number of individuals in the population at a specific time, the idea of geographic closure is involved, because when an experimenter speaks in terms of a population of absolute size he implies the existence of a geographic area in which the population resides. Thus, the concept and assumption of geographic closure are required of both open and closed models of populations.

We have assumed that the objective of the capture-recapture experiment is to estimate population size. Often, this objective is only secondary, and the main purpose of the experiment is estimation of birth rates or death rates, or both (*Robson 1963; Cormack 1964*). In such instances, of course, closed-population models are not useful, and the investigator must use methods that allow these processes to occur.



Bryan F. J. Manly

Bryan Manly has made several contributions to the theory for open-population models. He was the first to develop estimators for age-dependent populations. He conducted simulation studies to explore the small-sample properties of the Jolly-Seber model. He has published some 15 papers on capture-recapture methods—many of them in an entomological setting.

Manly took a B.Sc. degree in mathematics from the City University in London and worked as an industrial statistician before going to the University of Salford. There, M. J. Parr introduced him to ecological problems and to some of the statistical questions arising in the analysis of capture-recapture data. They shared ideas and began computer simulation studies to see how theoretical results worked in practice. In 1970 Manly accepted a Lectureship in Statistics at the University of Papua and New Guinea. Since 1973 he has been in the Biometrics Unit at the University of Otago, New Zealand.

Manly's recent work in capture-recapture studies has been concerned with accounting for animal movement and parsimonious modeling of capture data. He believes these and other related problems deserve a great deal more work. (Recent photograph.)

Return now to the case where estimation of numbers is of primary concern and, therefore, where a choice must be made between the two general classes of estimation techniques. We can reduce the basic dilemma to one question that the investigator must ask himself: Which class of models best approximates the experimental situation—open models, which allow the processes of birth, death, and migration to operate, or closed models, which allow individuals to possess varying probabilities of capture, depending upon which of three sources of variation are present? This question must be answered because no practical estimation techniques that allow both types of assumptions are available. There is a good reason for this, one that we have alluded to before: capture-recapture models can be generalized only to a certain degree before they lose their ability to provide real information about the parameters of interest. Thus, the problem of choosing between the two classes of models may be with us indefinitely. Unfortunately, statistical studies of the models involved do not permit the conclusion that one or the other class of model will be robust to failure of the constraining assumptions each requires. For example, both *Carothers (1973a, b)* and *Gilbert (1973)* have documented the fact that Jolly-Seber estimates can exhibit significant negative bias when individuals show certain types of unequal catchability. Although the models of *Otis et al. (1978)* have not been examined for robustness to the presence of birth, death, and migration, it is reasonable to expect that the presence of these processes could effect significant biases in the population estimators.

Thus, we see that the experimenter must bring all his biological knowledge of a particular experimental situation to bear on the selection of proper estimation techniques. According to *Cormack (1968:456)*, "In all cases every iota of information, both biological and statistical, must be gathered to check and countercheck the unavoidable assumptions." Although each situation will be different at least to some degree, a few general principles may be of help. First, closed-population techniques are more likely to be appropriate for short-term studies. In this context, short-term is taken as relative to the mean life span of the species. For example, daily trapping of an area for a week might be appropriate for *Microtus* spp.,

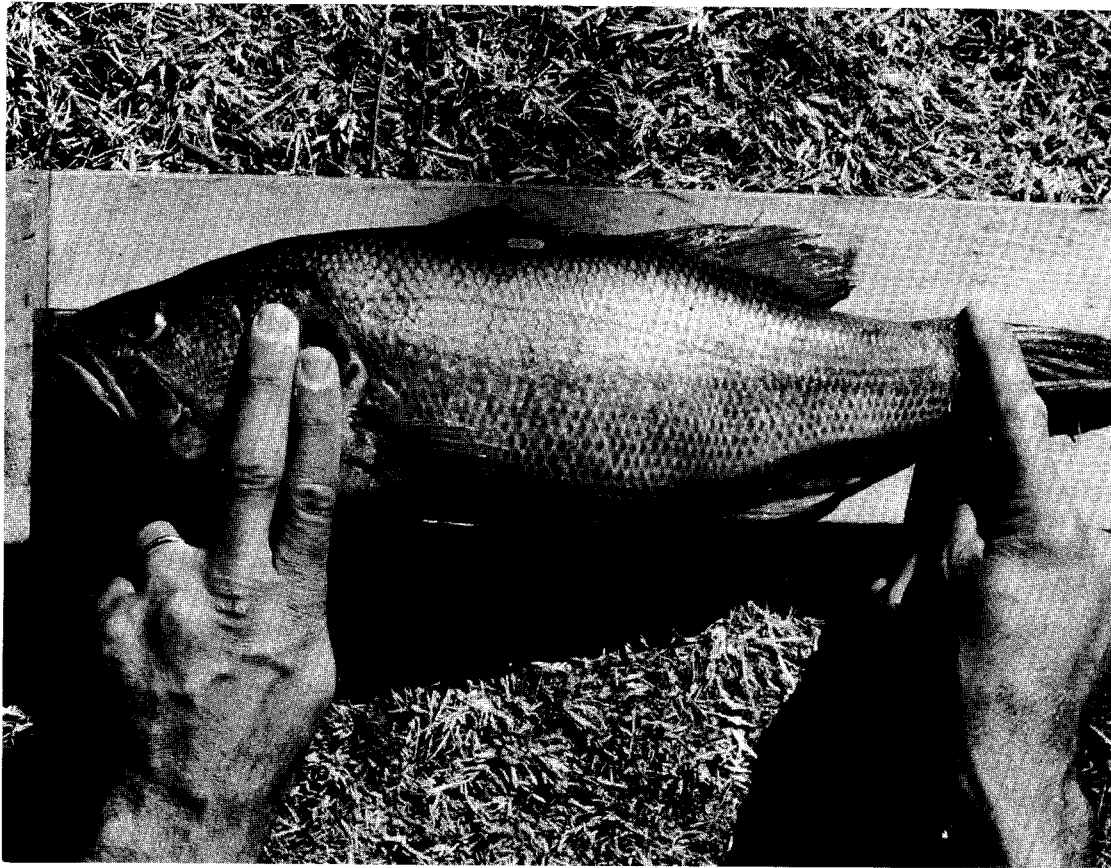
whereas weekly trapping for a month or two might be appropriate for *Lepus* spp. As the length of the trapping period increases, the probability increases that the processes of birth, death, and migration will affect the population significantly and, therefore, open models may become more appropriate. Further, there is the common sense notion that the size of the area trapped must be relatively large compared to the home range of the species under study if closed models that do not allow migration are to be used (see Chapter 5). For example, we would not assume that the effects of migration were negligible in an experiment in which coyotes were being trapped in a 1.6-km<sup>2</sup> grid of traps. Finally, we mention that closed-population models are likely to be most effective in producing good estimates if the trapping is not done during a period when young of the year appear in the population or during a time of significant dispersal.

### Summary

1. Open models allow the demographic closure assumption to be relaxed. The population size is allowed to vary among the sampling occasions, as  $N_1, N_2, \dots, N_j$ , because animals may enter the population, through birth or immigration, or leave the population, through death or emigration.
2. The most commonly used approach is the Jolly-Seber model, which is an extension of Model  $M_t$ . It allows estimates of population size, probability of capture, loss rate (deaths and emigration), and entry rate (births and immigration) for each sampling occasion  $j$  except, primarily, the first and last sampling occasions.
3. Heterogeneity has not been modeled for the open case, and, conceptually, no open removal model is possible. Behavior has been modeled for the open case only to a limited extent.
4. Geographic closure is a critical assumption for estimating population sizes  $N_j$ , but it can be ignored if only estimates of loss rates are of interest, (because these estimates are based only on the marked animals).
5. In most instances, open and closed models do not compete with each other: only one or the other is appropriate for satisfying the objectives of a given study.
6. Computer programs have been developed to allow a comprehensive analysis of open-population data (Arnason and Baniuk 1980).

### Questions and Exercises

1. A biologist is to study lake trout in Lake Superior and considers an open capture-recapture model in his work.
  - a. What problems might he face?
  - b. What parameters might be estimated?
  - c. What alternative methods should he consider?
2. You are interested in estimating the monthly survival rate of fish in a 1-ha pond. Would an open model be appropriate? Why?
3. Are the samples required for the open models relatively larger than those required for comparable closed models?
4. Using the Jolly-Seber model, a researcher obtains the following data.



A largemouth bass (*Micropterus salmoides*) has been tagged with a disk-dangler tag under the dorsal fin. (Photograph courtesy of Paul R. Turner.)

j	$\hat{N}_j(\hat{se}(\hat{N}_j))$	$\hat{S}_j(\hat{se}(\hat{S}_j))$	$\hat{B}_j(\hat{se}(\hat{B}_j))$
1	---	0.50 (0.09)	---
2	389 (42)	0.40 (0.06)	114 (70)
3	270 (51)	0.69 (0.08)	236 (106)
4	422 (59)	0.24 (0.05)	206 (72)
5	307 (67)	0.65 (0.04)	288 (106)
6	488 (96)	0.23 (0.07)	199 (90)
7	311 (60)	0.56 (0.15)	106 (130)
8	280 (100)	---	---
9	---	---	---

- Plot  $\hat{N}_j$  vs  $\hat{S}_j$ . Do these estimates provide support for the hypothesis that survival is low when the population is high (that the survival rate is density-dependent)?
  - What is the value of  $cv(\hat{N}_4)$ ?
  - What is the 95% confidence interval on  $S_5$ ?
  - Are the estimates of  $B_j$  related to the estimates of population size?
5. If a survival rate is assumed to be constant during a capture-recapture study, can this one survival rate be estimated with a Jolly-Seber model?